JRC SCIENTIFIC AND POLICY REPORTS

An analysis of national research systems (I): A Composite Indicator for Scientific and Technological Research Excellence

Sjoerd Hardeman, Vincent Van Roy, Daniel Vertesy,
with contributions from Michaela Saisana

November 2013

Report EUR 26093 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Sjoerd Hardeman
Vincent Van Roy
Daniel Vertesy

Address: Joint Research Centre, Via Enrico Fermi 2749, TP 361, 21027 Ispra (VA), Italy

E-mails:
sjoerd.hardeman@jrc.ec.europa.eu
vincent.van-roy@jrc.ec.europa.eu
daniel.vertesy@jrc.ec.europa.eu

Tel.: +39 0332 78 3976
Fax: +39 0332 78 5733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.

JRC83723

EUR 26093 EN

ISBN 978-92-79-32527-4

ISSN 1831-9424

doi: 10.2788/95887

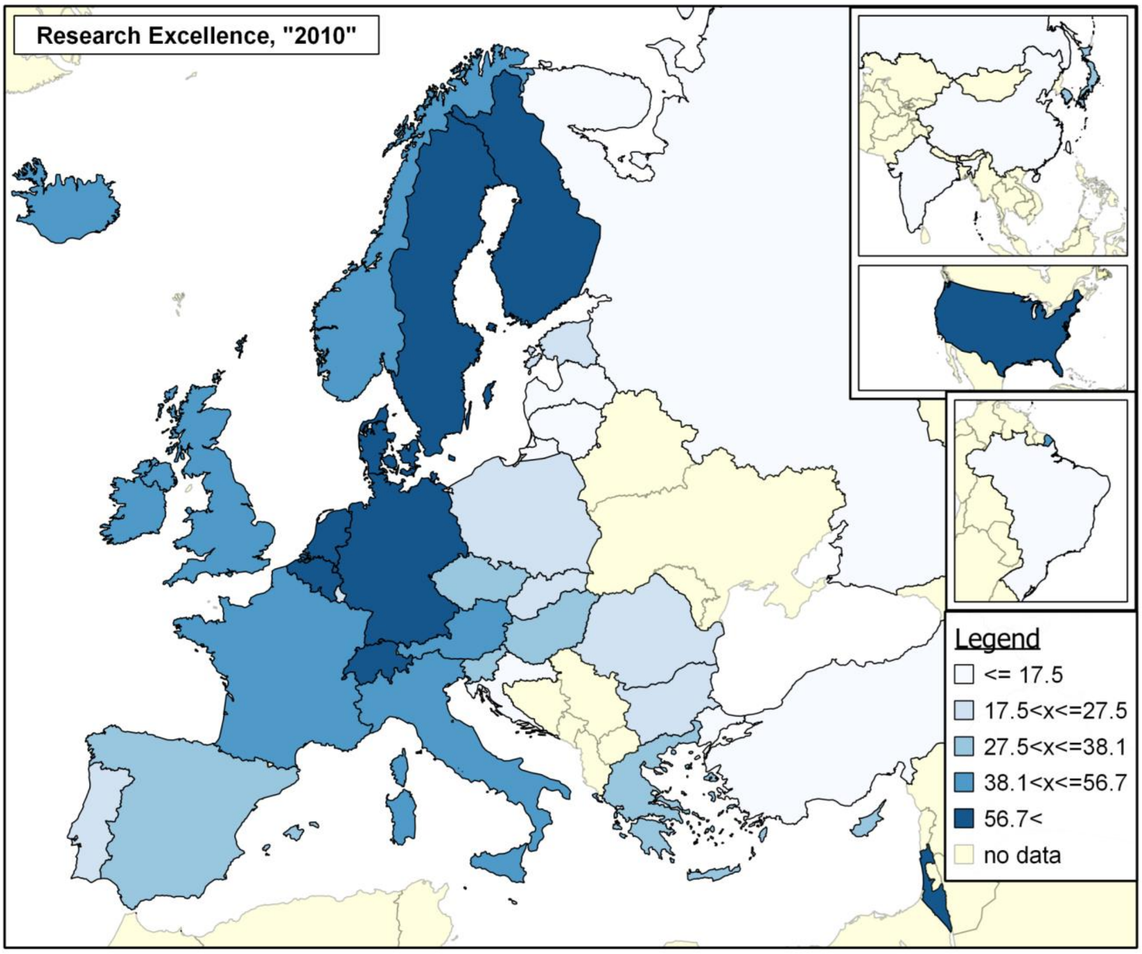
Luxembourg: Publications Office of the European Union, 2013

Printed in Luxembourg

An analysis of national research systems (I): a composite indicator measuring scientific and technological research excellence

Sjoerd Hardeman, Vincent Van Roy, Daniel Vertesy
with contributions from Michaela Saisana

November 2013



Executive summary

Monitoring Research Excellence in and beyond the European Research Area

It is widely acknowledged that **many European countries are outperformed by countries like the United States when it comes to technological and scientific research**. To remedy this situation, the European Commission aims at stimulating research excellence by increasing competition among researchers at a European level. The results reported in this report follow from a project initiated by the Directorate-General for Research and Innovation (DG RTD) of the European Commission within the context of developing composite indicators for the Innovation Union (Composites_4_IU) project. The main objective of the overall project is to develop indicators that are capable of measuring and monitoring patterns and trends in research across countries. This report assesses the performance of countries in terms of their record in producing state-of-the-art scientific and technological research outcomes; that is, research excellence.

This report proposes a novel way to **conceptualise and measure research excellence** at the country level using a composite indicator approach. So far, few studies measure scientific and technological research excellence at the country level whilst taking into account the multidimensional nature of research excellence. Following the OECD Oslo Manual, we define **research as creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications**. Akin to the idea of national innovation systems, a national research system is made up of the actors within a country that jointly produce research outcomes. In our conceptual framework, **national research systems contain four core elements**: components (the operating parts of the system), relationships (interactions), attributes (motivations and goals), and outcomes (the creation of excellent knowledge). **Scientific and technological research excellence is defined as the top-end quality outcome of systematically performed creative work undertaken to increase the stock of knowledge and new applications**. Having evaluated the quality profile of a large set of potential variables, we focus on four variables to measure the top-quality output of scientific and technological research activities at the national level:

1) a field-normalised number of **highly cited publications** of a country as measured by the top 10 % most cited publications (in all disciplines) per total number of publications (HICIT);
2) the number of **high quality patent applications** of a country as measured by the number of patents filed under the Patent Cooperation Treaty (PCT) per million inhabitants (PCTPAT);
3) the number of **world class universities and research institutes** in a country as measured by the number of organisations of a country in the top 250 universities and 50 research institutes divided by gross expenditures in R & D of a country per (TOPINST); and
4) the number of **high prestige research grants** received by a country as measured by the total value of European Research Council grants received divided by public R & D expenditures of a country (ERC).

The field-normalised number of highly cited publications of a country and the number of high quality patents of a country represent new knowledge attributable to a country that is inscribed in texts and artifacts, the number of world class universities and public research institutes in a country and the number of high prestige research grants received by a country are proxies for monitoring new knowledge that is embodied in the human capital of that country. In what follows we present the key messages that emerge from the JRC study which compares 40 countries (and the EU) at two points in time: based on the most recent data available for this type of analysis (labelled '2010') and five years prior ('2005'). The 40 countries are 33 ERA countries (the EU 27 members plus Croatia, Turkey, Switzerland, Iceland, Norway and Israel) and seven benchmark countries (Brazil, Russia, India, China, South Korea, Japan and the United States).

Research excellence within the European Research Area and beyond

There is no ideal country in terms of their performance in producing excellent research; instead, there is space for improvement in all ERA countries. First, in '2010' the share of the field-normalised number of **highly cited publications** in total publications is highest in Switzerland, Denmark, the Netherlands, Iceland and Sweden. On the other end, Latvia, Russia, Croatia, Poland and Bulgaria rank the lowest. Second, the ratio of the number of **world class universities and public research institutes** over GERD is outstanding in Switzerland, and also high in the Netherlands, Denmark and Israel, whilst there are 13 countries (10 are EU members) that do not have a single university or institution in the global top 250. Third, the number of **high quality patents** per million inhabitants is very high in Sweden, Switzerland, Finland and Israel (over 20), whilst on the other end, Cyprus, Bulgaria, Poland, Romania, Lithuania, Turkey and Greece have less than 2.5 patents per million population. Finally, the ratio of **high prestige research grants** to public R & D is highest in Switzerland, Israel, the Netherlands, Sweden, the United Kingdom, Denmark and Belgium, whilst Lithuania, Luxembourg, Latvia, Malta, Slovakia, Croatia and Turkey are among the lowest performers (Figure E1).

Figure E1 Normalised variable and composite scores of Research Excellence, '2010'

Country		Highly Cited Publications per Total Publications	Top Universities & Public Research Org's per GERD	PCT Patents per population	ERC Grants per public R&D	*Overall Score*
Austria	AT	75	22	51	79	*50.5*
Belgium	BE	86	46	38	86	*59.9*
Bulgaria	BG	23	23	11	66	*24.7*
Cyprus	CY	59	10	11	93	*27.8*
Czech Republic	CZ	31	29	15	60	*29.9*
Germany	DE	69	44	69	73	*62.8*
Denmark	DK	97	63	73	82	*77.7*
Estonia	EE	55	10	16	50	*25.9*
Greece	GR	57	27	13	79	*35.3*
Spain	ES	58	23	18	75	*36.6*
Finland	FI	73	28	90	85	*62.9*
France	FR	63	30	37	78	*48.2*
Hungary	HU	39	20	17	82	*31.9*
Ireland	IE	71	10	36	83	*38.1*
Italy	IT	62	29	24	78	*43.1*
Lithuania	LT	33	10	11	10	*13.9*
Luxembourg	LU	53	10	29	10	*19.8*
Latvia	LV	14	10	12	10	*11.5*
Malta	MT	64	10	15	10	*17.5*
Netherlands	NL	93	69	66	91	*78.9*
Poland	PL	21	13	11	59	*20.5*
Portugal	PT	59	10	13	65	*26.5*
Romania	RO	26	10	10	38	*17.8*
Sweden	SE	78	52	100	88	*77.2*
Slovenia	SI	48	10	25	48	*27.5*
Slovakia	SK	29	28	12	10	*17.7*
United Kingdom	UK	78	38	37	90	*56.1*
EU-27	EU27	58	31	36	80	*47.9*
Croatia	HR	17	10	13	10	*12.2*
Turkey	TR	32	10	11	10	*13.8*
Switzerland	CH	100	98	93	100	*97.6*
Iceland	IS	88	10	33	78	*38.8*
Norway	NO	73	30	46	71	*51.8*
Israel	IL	70	62	83	99	*77.1*
Brazil	BR	26	12	11	n.a.	*14.6*
Russia	RU	14	11	11	n.a.	*12.1*
India	IN	30	10	10	n.a.	*14.7*
China	CN	32	11	11	n.a.	*15.7*
Rep. of Korea	KR	43	22	49	n.a.	*35.7*
Japan	JP	38	23	64	n.a.	*38.1*
United States	US	79	44	52	n.a.	*56.7*

NB: Bar lengths indicate country scores for the four research excellence indicators (highly cited publications, Top 250 Universities, PCT patent applications and ERC grants received). The minimum score for each indicator is 10, the maximum is 100. Non-ERA countries are not assessed on the ERC grants indicator due to home region bias, thus marked by 'n.a.' Median scores within each aspect and the overall median are shown beneath the bars.

Comparing the European Research Area with the rest of the world

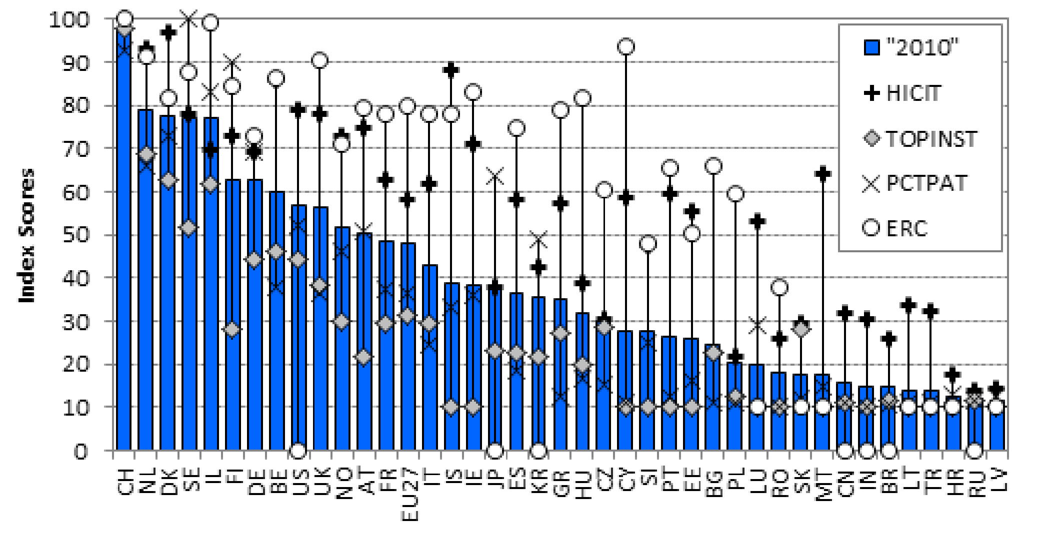
Research excellence scores in '2010' are high in Switzerland, the Netherlands, Denmark, Sweden and Israel (Figure E2). Countries that are ranked in the middle involve both big countries (like Germany, Japan, France, and the United States) and Mediterranean countries (like Spain, Greece, and Italy). Lower ranked countries are both emerging economies (like Brazil, India, China and Russia) and

Eastern European countries (like Latvia, Lithuania and Croatia). The figure also helps to decompose research excellence: even on the top, each country has their own 'competitive' strength from among the various indicators.

If one considers field-normalised highly cited publications as a proxy for scientific research excellence and high-quality patents as a proxy for technological research excellence, then **it is evident that some countries perform especially well in technological research excellence and less so in scientific research excellence, whilst for other countries it is the other way around**. However, distinguishing scientific from technological research excellence is rather difficult given that on the one hand a fair share of the outcome of technological research is not patented and is evaluated by the market in a different manner from the peer-review based evaluation of scientific research excellence and, on the other hand, a fair share of scientific research excellence does not end up in publications but as artifacts.

Finally, lower ranked countries seem to perform better in terms of scientific research excellence than in technological research excellence. Including data on ERC grants provides to a great extent a similar picture on research excellence for the ERA countries, yet including this indicator to the aggregate, makes a notable difference for Iceland, Cyprus, and Hungary; countries which succeeded in attracting relatively high amounts of ERC grants in '2010'. What we conclude is that, **although almost a quarter of the ERA countries outperform the United States, the United States outperforms the European Union in general**.

Figure E2 Research excellence composite indicator and individual variable scores, '2010'



Research Excellence and research funding

Correlations show that countries scoring low on research excellence are also those that spend less on research and development (R & D) as a share of GDP and countries scoring high on excellence spend a large share of their GDP on R & D (Figure E3). There appears to be a 'critical mass' of overall R & D expenditure which is needed to achieve excellence, corresponding to gross R & D expenditures above 1.5 % of GDP. Yet higher spending does not automatically mean excellent results: a few of the higher

R & D spenders such as Japan, South Korea and Iceland have lower than expected excellence scores, which may either signal inefficiencies in the research system or the development of future centres of excellence. In any case, the relationship between funding and excellence might well go the other way around; that is, those countries that produce excellent research spend relatively much on R & D *because* they are excellent.

Figure E3 Research excellence versus R & D Intensity by type ('2008')

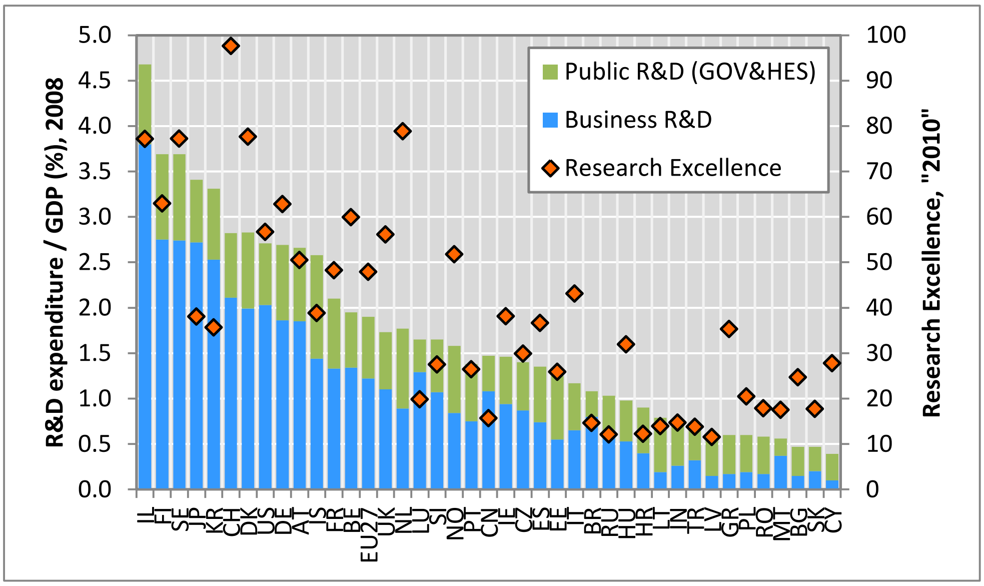
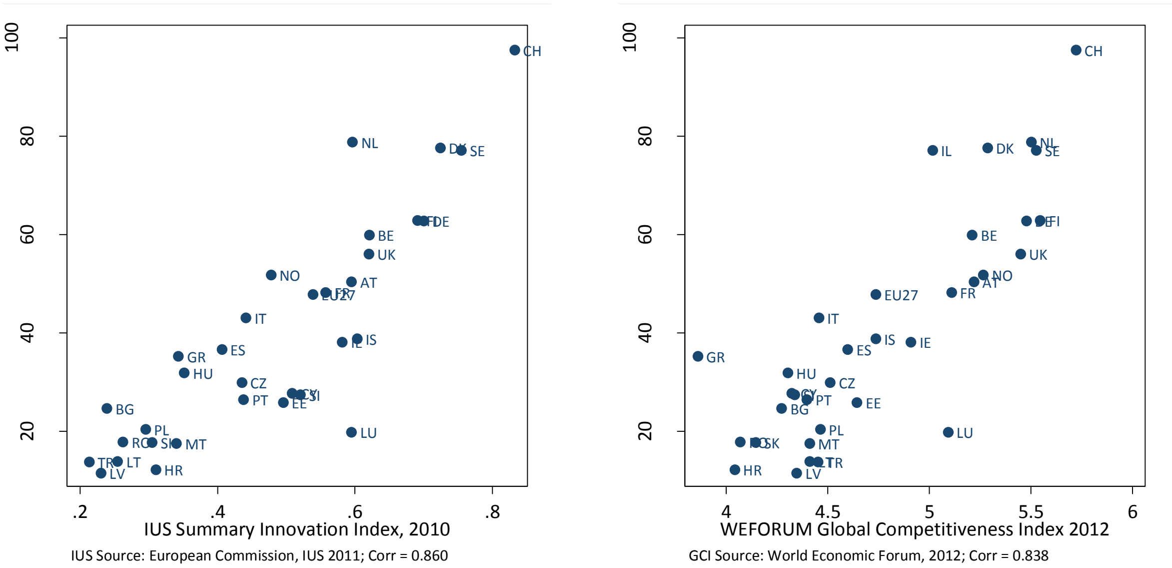


Figure E4 Research Excellence ('2010') versus the Summary Innovation Indicator and the Global Competitiveness Index



Research Excellence versus innovation and competitiveness

Measured against the most commonly used composite indicators of innovation and competitiveness, research excellence correlates strongly and positively (Figure E4). This is partly due to the fact that conditions for innovation and competitiveness overlap with science-based innovation. Differences nevertheless exist, since innovation and competitiveness cover dimensions not included in research excellence and vice versa (i.e., not all innovation is science and technology based; and competitiveness of countries does not just depend on the availability of excellent science and technology). Overall, the analyses show that that **research excellence is akin to but still a somewhat different phenomenon than innovation and competitiveness**.

Recommendations: focus on people and balance science with technology

In conclusion, the proposed composite indicator measuring research excellence fills a gap in measuring research excellence at the country level and has an added value on top of other country-level performance indicators in the field of science and technology assessment. Measured against the most commonly used composite indicators of innovation and competitiveness it is shown that the proposed composite indicator on research excellence are akin to but also different from measurements dealing with the phenomena of innovation and competitiveness.

From the analysis and results presented in this report we make two recommendations. One recommendation holds that relatively poorly performing countries need to focus more on establishing research excellence embodied by people. This recommendation follows from the observation that some countries' relative under-performance is mainly due to their relatively low scores for the number of world class universities and research institutes and the number of received ERC grants (both variables measuring (scientific) research excellence embodied in people) and not so much for the number of highly cited publications and the number of high-quality patents (both indicators of research excellence embodied in texts and artifacts).

The second recommendation holds that countries should reconsider their strategic orientation to better balance their performance in scientific research excellence with their performance in technological research excellence. For most countries there is room for improvement in overall research excellence by either improving their scientific or technological research excellence. Few countries perform equally well in both scientific and technological research excellence. That is, scientific and technological research excellence do not necessarily go hand in hand. Here, we do not favour one form of research excellence over the other. What holds then is that, either way, there is room for improving research excellence in most if not all countries.

Contents

Executive summary i
Contents 7
1. Introduction 9
1.1. Background of the project 9
1.2. Contribution of this report 10
1.3. Outline of the report 11
2. Background: three accounts on research 14
2.1. Research in contemporary society – a descriptive account 14
2.2. The policy discourse on research – a normative account 15
2.3. The economics of research – an explanatory account 17
3. Conceptual framework: from national research systems to research excellence 21
3.1. A characterisation of national research systems 21
3.2. A definition of research excellence 24
4. Data 27
4.1. General measurement issues 27
4.2. Measuring research excellence 27
4.3. Specification of variables measuring research excellence 30
4.4. Alternative variables considered but not included in the analyses 38
5. Methodology 44
5.1. Descriptive statistics 44
5.2. Multivariate analysis 46
5.3. Normalisation, weighting and aggregation 49
6. Results 53
6.1. Country scores and rankings in their performance on research excellence 53
6.2. Robustness assessment of the research excellence index 56
6.3. The link between research excellence and related phenomena 62
7. Conclusion 76
7.1. Summary 76
7.2. Discussion 78
7.3. Recommendations 80
References 82
Acknowledgements 88

1. Introduction

1.1. Background of the project

With the introduction of the Europe 2020 strategy and its Innovation Union flagship initiative, the European Commission has made a shift in orientation from fostering 'research in Europe' towards fostering 'European research' (Nedeva and Stampfer, 2012). Recognising that coordinating national research efforts on a case-by-case basis is practically unfeasible, attention has shifted towards the construction of a pan-European research system called the European Research Area (ERA). Accompanying this, it is widely acknowledged that many European countries are outperformed by countries like the United States when it comes to both technological and scientific research (Pavitt, 2000, Dosi et al., 2006). To remedy this situation, the European Commission aims at stimulating research excellence by increasing competition among researchers at a European level; for example, by establishing a central research funding agency, the European Research Council (ERC). Meanwhile, the current economic crisis has increased budgetary pressures across the board. Hence, allocating scarce resources to research has become an issue to be dealt with in the context of growth promoting policies. Overall then, it is unlikely that the economic crisis has no impact on research at all (Filippetti and Archibugi, 2011).

While some take investments in research as a necessary condition to foster welfare growth (Gruss, 2012), others discuss the kind of institutional and organisational arrangements that are needed to make research most productive (Marty, 2012). This project follows the latter strand of thought and investigates these issues for research at the country level. The results reported follow from a project initiated by the Directorate-General for Research and Innovation of the European Commission (DG RTD) within the context of developing indicators for the Innovation Union. The main objective of the overall project is to develop indicators that are capable of measuring and monitoring patterns and trends in research across countries. As such, the focus is on measuring three dimensions to research. One is about the interactions that take place between research actors within and across Europe. The main aim here is to track patterns of mobile researchers, R & D investment flows, and collaborative research endeavours across and beyond EU Member States. Another dimension is about research interactions that take place between different kind of actors, such as universities, industry and government actors. Again the main aim is to track patterns of mobile researchers, R & D investment flows, and collaborative research endeavours along these institutional lines. Finally, a third dimension is about the impact that research activities have in terms of the outcomes produced and the ease with which inputs to research are transformed into research outputs. While follow-up reports address the first two dimensions, this report addresses the latter dimension.

Mostly, in particular, this report assesses the performance of countries in terms of their record in producing state-of-the-art scientific and technological outcomes; that is, research excellence. As such, this report builds on previous work and experience of the European Commission related to the measurement of research excellence. The Report of the Expert Group on the Measurement of Innovation (Barré et al., 2011) proposed a very broad set of indicators of research excellence that encompassed indicators of research funding, collaboration and specialisation alongside output-oriented indicator. A follow-up feasibility study on the development of a composite indicator of research excellence by the JRC (Vértesy and Tarantola, 2012) showed statistically that the indicators proposed by the Expert Group cover more than a single underlying phenomenon. For these reasons, this study dedicated substantial attention to a theoretical discussion of why excellence receives policy attention, how to define it and how to measure it.

1.2. Contribution of this report

The main contribution of this report lies in a proposal for conceptualising and measuring research excellence at the country level using a composite indicator approach.

The added value of measuring research excellence at the country level using a composite indicator approach lies (i) in the current lack of such indicators at the *country level*, (ii) in the nature of *research excellence as a complex and distinct phenomenon*, (iii) in the potential of composite indicators to *summarise such complex phenomena*, and (iv) in *the relation between research excellence as measured by a composite and other measured phenomena* (such as competitiveness and innovation) with which research excellence is often associated. Though indicators measuring research excellence at the organisational level abound (Saisana et al., 2011), they are virtually absent at the level of countries. Yet, given that research is still largely a national endeavour (Hoekman et al., 2009, Hoekman et al., 2010, Chessa et al., 2013) and provided that policies at the national level still play a large role therein (Nauwelaers and Wintjes, 2008), an indicator measuring research excellence at the country level is highly warranted. This is not to say that measuring research excellence at the country level is easy. On the contrary, given the complex and multidimensional nature of research excellence, measuring this phenomenon requires advanced methodological approaches. Here, a composite indicator approach is a suitable methodology to summarise different aspects of the multi-dimensional phenomenon that research excellence is in an efficient, parsimonious, and tractable way.

The country-level composite indicator on research excellence that is proposed and discussed throughout this report has an added value on top of other country-level performance indicators in the field of science and technology assessment. Measured against the most commonly used composite indicators of innovation and competitiveness it is shown that research excellence resembles these two phenomena to a great extent. This can partly be explained by the inter-linkages between scientific and technological research excellence on the one hand and competitiveness and innovation on the other. Differences nevertheless exist, since innovation and competitiveness cover dimensions not included in research excellence and vice versa. For one thing, not all innovation is necessarily science and technology based. Also, competitiveness of nations does not just depend on the availability of excellent science and technology. The analyses show that research excellence is akin to but also different from measurements dealing with the phenomena of innovation and competitiveness. As such, the proposed composite indicator on research excellence is related to but distinct from other (composite) indicators available in the field of science and technology assessment (the Global Innovation Index of INSEAD-WIPO (Dutta, 2012), the Summary Innovation Index of the Innovation Union (Commission, 2011), and the Global Competitiveness Index of the World Economic Forum (Sala-i-Martín et al., 2012)).

1.3. Outline of the report

The report proceeds as follows. Section 2 discusses the overall theoretical and policy background of the study. Three accounts are emphasised. First, a descriptive account that draws primarily on sociological perspectives on contemporary society and the place of research herein. Second, a normative account on the relation between research and the economy as it is apparent from the current policy discourse on excellence, efficiency, and competitiveness. Third, an explanatory account that starts from the old economics of research with its focus on the linear model of innovation and moves to the new economics of research with its emphasis on research as an evolving complex system. Taken together, these three accounts substantiate the validity of an investigation into national research systems and the crucial role played by research excellence therein.

Section 3 proposes a conceptual framework to assess research excellence in the context of national research systems. First, the main building blocks of the concept of national research systems is discussed; then, the main pillars of research excellence as revolving on, the one hand technological research excellence and on the other hand scientific research excellence are addressed.

Section 4 discusses the data used (and not used) in our proposal for measuring research excellence using a composite indicator approach. First some general measurement issues are discussed. Questions like 'what does measuring means?' and 'what does measuring do?' are briefly addressed. Second, we discuss these issues specifically in the context of measuring research excellence. That is, the data requirements for measuring research excellence are being set in this section. Third, following the data requirements, the rationale for selecting some variables and not others is discussed.

Section 5 discusses in depth the methodology applied to construct a composite indicator on research excellence in depth. The methodology follows four steps before turning to the main results in the next section. The first step is about data coverage and the computation of missing data. The second step deals with issues of scale and discusses means of normalisation to overcome this issue. The third step deals with multivariate analysis. Here, the results from cluster analysis and principal component analysis are presented. The final fourth step addresses ways of weighting and aggregating the different variables included in the analysis. In all, section 5 will discuss these four steps in turn.

Section 6 turns to the main findings of the report. First, the scores and rankings of countries' performance on research excellence are discussed. Second, a sensitivity analysis on these scores and rankings is presented. Sensitivity analysis is an important tool in assessing how sensitive a composite indicator is to the particular methodological choices made throughout its construction. As it turns out, the proposed composite indicator is partially sensitive to the methodologies used. That is, the choice of denominating variables and aggregation have an impact on the final scores and rankings. Third, the relation between research excellence as measured by the proposed composite and other measured phenomena with which research excellence is often associated is assessed. Finally, section 7 concludes with a summary of the main approach taken and findings of this study, a discussion of the main results of this report, and recommendations for policy and further research.

2. Background: three accounts on research

2.1. Research in contemporary society — a descriptive account

Many concepts are currently in use to describe contemporary society. While some speak of a post-industrial society (Bell, 1976), others speak of an era of post-Fordism (Amin, 1994). Still others speak of the knowledge economy (Drucker, 1969), the second industrial divide (Piore and Sabel, 1984), post-modernity (Harvey, 1989), reflexive or radicalised modernity (Giddens, 1990), second modernity (Beck, 1992), the post-normal age (Funtowicz and Ravetz, 1993), the information society (Webster, 2006), the network society (Castells, 1996, van Dijk, 2005), liquid modernity (Bauman, 2000), Mode 2 society (Nowotny et al., 2001), hyper modernity (Lipovetsky et al., 2005), and the knowledge society (David and Foray, 2002), amongst others. Notwithstanding important differences among these various characterisations of contemporary society, they share an emphasis on at least three developments that allegedly take place in and characterise contemporary society.

Intensified use of intangible assets in the production and consumption of goods. One development concerns an intensification in the use of intangible assets in the production (and consumption) of products and services. Production no longer depends primarily on natural resources (e.g. fossil energy) and physical capital (i.e. machineries), rather than on intangible inputs such as knowledge and information (Powell and Snellman, 2004). Accompanying an intensification in the use of knowledge and information comes an accelerated pace of technological change. While new technologies are continuously being introduced, older technologies become equally rapidly obsolete. In other words, intangible aspects play an ever more important role in the creation, accumulation, and depreciation of economic (David and Foray, 2002). However, not only do new technologies become rapidly obsolete by virtue of being replaced by newer ones; so do new technologies confront society with new problems that demand new solutions (Beck, 1992, Funtowicz and Ravetz, 1993, Nowotny et al., 2001). As such, every new solution to existing problems introduces new uncertainties setting in motion a self-reinforcing process of innovation (Nowotny et al., 2001). Hence contemporary society can be characterised by a continuous demand for and supply of products and services that primarily build on intangible assets such as knowledge and information.

Globalisation and localisation. Most descriptions of contemporary society take issue with the changing relationships among people across time and space. Changing spatial-temporal relationships among people have often been addressed with reference to the notion of globalisation. Globalisation can be defined as the compression of time and space that renders the circulation of people, goods, and ideas increasingly a worldwide affair (Harvey, 1989, Hoekman, 2012). Accordingly, some have argued that in a globalised world there is a ‘death of distance’ (Cairncross, 1997) and, hence, the world has become ‘flat’ (Friedman, 2005): people interact with each other without being restricted by physical distance. At the same time, however, economic activities are extremely spatially concentrated (Feldman, 1999). While some countries are extremely rich, others are extremely poor; and while some regions experience strong economic performance, others experience an economic downturn; while the population of some cities is rapidly growing, the population of other areas is experiencing tremendous negative growth. In addition, and although the number of global interactions is increasing, they do not come at the expense of less local interaction (McCann, 2008). In all then, contemporary society is characterised by both processes of globalisation and processes of localisation. As such, the national dimension to economic production, though firmly placed within a global context, becomes more – and not less – important.

Hybridisation of societal domains. A third development concerns a hybridisation of societal domains. At the same time as commercialisation plays an ever more important role in academia, scientisation becomes an ever more dominant feature of the market place (Moore et al., 2011). The commercialisation of academia can be observed with reference to both an increased emphasis on the accountability of science to the general public and an increase in the demand for societal relevant scientific research outcomes that are immediately applicable to society. While the former can be illustrated by an increased emphasis on university rankings and science evaluation in general (Whitley and Gläser, 2007, Fealing, 2011, Hicks, 2012), the latter becomes apparent in the increased use of such concepts like the entrepreneurial university (Etzkowitz, 2003) and universities' 'third mission' (Laredo, 2007). Scientisation as a dominant feature of the market place refers to the interpenetration of science into the market domain (Drori, 2003) and can be observed by the increase in science-based industries (Powell and Snellman, 2004) and an increase of highly educated people working for the private sector (Gibbons et al., 1994).

The implications of a hybridisation of societal domains are at least twofold. First, a particular set of activities is no longer attributable to a particular set of actors only. For example, scientific research is no longer confined to the domain of universities alone. In other words, different kinds of actors perform the same kind of function in society (Leydesdorff, 2006). Second, a particular set of norms and rules is no longer attributable to a particular set of actors only. For example, the norms of open science (Merton and Storer, 1979) are very much in line with those operating within open source software development (Raymond, 2008, Lerner and Tirole, 2002). Under such 'institutional isomorphism' (DiMaggio and Powell, 1983), different kind of actors adhere to the same kind of values and perform similar kind of activities (Stark, 2011). As such, and in line with our earlier observation about the intensified use of intangible assets in the economy, research is no longer strictly confined to particular functional and institutional domains of society only (Leydesdorff, 2006).

Summarising, three developments can be identified that characterise contemporary society. One is that production is characterised by an intensification in the use of intangible factors such as knowledge, information and skills. Second, contemporary society is characterised by changing relations among people across time and space. While on the one hand society is globalising, activities become ever more geographically concentrated on the other hand. A third and final development concerns a hybridisation of societal domains. With this we mean that the boundaries between science, markets, and politics have become ever more blurred; suggesting that each societal domain influences others nowadays. From these three developments, an interest in the spatial and institutional arrangements that bring about knowledge through research becomes self-evident. While the first development points at the importance of knowledge (and hence research) in contemporary society, the latter two developments point at the changing spatial-temporal and institutional arrangements characterising it. Together, these three developments thus point at the changing nature of research as being performed 'everywhere.'

2.2. The policy discourse on research — a normative account

The focus on efficiency and excellence in research follows from a general rise of accounting practices that have recently been accompanied by calls for austerity on the one hand and the emergence of a policy discourse that focusses on steering competitiveness and outperforming others on the other.

The rise of accounting practices has been described in terms of the emergence of an audit society (Power, 1994, Power, 1997). The emergence of the audit society refers to 'the spread of a distinct mentality of administrative control' (Power, 1994, p. 3) in which there is an increasing demand for accountability and transparency and increasing supply of models of quality assurance and organisational control.

Research policy in an age of professionalisation and accountability. Following the tenets of the audit society, research can be characterised by a period of professionalisation and accountability (Elzinga, 2012). According to Nowotny (2006, pp. 1-2) 'from the 70s onwards … the budget cuts from the government initially triggered by a situation of economic stringency, were never to return to normal, but became a new normality themselves. … It became deeply enmeshed in a culture of searching utilitarian objectives, driven by norms of efficiency and accountability.' After a period of legitimation during the Cold War in which the ideological clashes between East and West provided an important rationale for research policy, the following period of professionalisation and accountability demands from researchers to do more with less (Elzinga, 2012). Within research's period of professionalisation and accountability an interest in efficiency in research follows from a concern with trying to get the best out of the research system at the least cost. One may argue about the desirability of this period of professionalisation and accountability in research. At the very least, however, one cannot deny that current concerns about financial austerity has also entered the domain of research and hence deserves considerable attention.

Research policy towards steering competitiveness. In parallel with the period of professionalisation and accountability, policy-makers have sought to steer the competitiveness of their administration (Slaughter and Rhoades, 1996, Bristow, 2005, Elzinga, 2012). Competitiveness can be defined as the conditions or amenities available in a country or region that increase its living standards (Boschma, 2004, Kitson et al., 2004). Some interpret the competiveness of a region or country in comparison to the performance of other regions and countries (Thurow, 1992). Derived from the notion of competition, the performance of a competitive country or region can then be characterised as positive by virtue of being better than other countries or regions. Herein, policy is legitimated for otherwise the European Union runs the risk of lacking behind the United States and Japan or becomes overtaken by countries like China and Brazil. Others, however, adhere to a dynamic interpretation of competitiveness and focus on the ability of countries and regions to continuously upgrade their economies (Boschma, 2004, Kitson et al., 2004). Herein, the main concern resides not so much in a country's or region's performance compared to other countries or regions. Rather, it focusses on the extent to which a country's or region's current conditions and amenities contribute to raising its future living standards (irrespective of others).

Notwithstanding these important differences in interpretation, most accounts of competitiveness attribute an important role to research. However, policy-makers' interest does not just reside in any kind of research but particularly in excellent research. Admittedly, the focus on excellent research follows most directly from an interpretation of competitiveness as 'keeping up with the Joneses': outperforming others in terms of raising standards of living then means outperforming others in terms of making excellent scientific discoveries and technological inventions. More in general then, the focus on constructing competitive economies thus translates into a focus on establishing excellence in research (Power and Malmberg, 2008). As such, especially in the context of an audit society, a focus on excellence follows naturally from embracing the aim of getting the best out of the research system at the least cost.

2.3. The economics of research — an explanatory account

Ever since the seminal works of Abramovitz (1956) and Solow (1957) it has been widely recognised in economics that beyond capital accumulation, technological change is a (if not the) major source of welfare growth. With research as the hallmark of activities that brings about technological change, many scholars became interested in the nature and role of research in both science (Nelson, 1959, Stephan, 1996, Stephan, 2012, Dasgupta and David, 1994) and innovation (Nelson, 1962, Nelson and Winter, 1982, Dosi, 1988, Jaffe et al., 2002). Once we recognise that research contributes to welfare growth, a major follow-up question holds how many resources should be spend on research activities.

From the 'old' economics of research … Answering this question is not straightforward given that 'if we allocate a given quantity of resources to science, this implies that we are not allocating these resources to other activities and, hence, that we are depriving ourselves of a flow of future benefits that we could have obtained had we directed these resources elsewhere' (Nelson, 1959, p. 297). In other words, if we are to assess the optimal amount of resources to be allocated to research, we have to take into account the uses of all other activities to which resources could be allocated. Overall then, ascertaining 'what research is really worth' is a difficult if not an impossible task to accomplish (Macilwain, 2010). What has been assessed, however, is productivity of research. Rather than assessing the optimal allocation of resources to research, some economists are thus interested in the returns from research. Using a production function approach some studies attempt to determine the average rate of return on investments in research (see Salter and Martin (2001) and Hall et al. (2010) for reviews). Differences in methodology and measurement across these studies apart, the main finding of most of them hold that there are considerable returns on investment in research. Notwithstanding this important finding, estimating the average returns from research using a production function approach has met with fierce criticism (Bonaccorsi and Daraio, 2005).

… Towards a new economics of research. One such criticism takes issue with the particular way of describing the relation between research and welfare growth in terms of a linear model of innovation (Kline and Rosenberg, 1986, Godin, 2006). In a linear model of innovation a distinction is made between basic research, applied research, and the development of research towards commercial ends. As such, the transformation of research into welfare follows three steps. In the first step, basic research takes place within the public realm of science. Given the public nature of scientific research, the outcomes of this research are 'freely' available to society at large (Arrow, 1962, Foray, 2004). In a second step then, economic agents use the publicly available outcomes from science to develop new technologies. Finally, in a third step these technologies diffuse on the market to become widely used in society. A main drawback of the linear model of innovation (especially when it is addressed while using a production function approach) is that it obscures the processes that underlie the transformation of research into welfare growth. In other words, the linear model of innovation treats the transformation of research into welfare as a 'black box' (Rosenberg, 1982).

Alternative conceptual approaches have been proposed to address research, such as the national innovation systems approach (Lundvall, 1988, Nelson, 1993, Edquist, 1997), the regional innovation systems approach (Cooke et al., 1997, Cooke, 2001), post-academic science (Ziman, 1994), mode 2 knowledge production (Gibbons et al., 1994), Pasteur's quadrant (Stokes, 1997), and the triple helix of university-industry-government interactions (Etzkowitz and Leydesdorff, 2000, Leydesdorff and Etzkowitz, 1996), and post-normal science (Funtowicz and Ravetz, 1990, Funtowicz and Ravetz, 1993). Whatever the differences among these approaches (for an overview of this literature see Hessels and van Lente (2008)), most of them take issue with some of the features of research that are not covered by the linear model of innovation. First, instead of conceiving of research as a sequential process, these approaches stress that research involves many feedbacks between basic and applied research, science and technology, invention and innovation. Although in the very long run one might be able to identify distinct scientific discoveries that form the basis of subsequent technological breakthroughs (Balconi et al., 2010), in the short to medium run the relation between science and technology is much more diffuse. Research involves considerable and fundamental uncertainties (Knight, 1921) and as such t is hard – if not impossible – to determine ex ante whether and if so what exactly comes out of research ex post.

Second, though related, instead of taking research as an orderly process, the alternative approaches stress that research is a complex process. The complexity revolving research has at least two aspects. One is that there are many different types of agents involved in research. These types of agents range from venture capitalists and public funding agencies to public research organisations (such as many universities) and private firms (such as pharmaceutical companies). Another aspect of the complexity involved in research concerns its interactive nature. The various types of agents involved in research compete but also collaborate with each other in the production of research outcomes. It follows from the different types of agents involved in research that the interactions are also variegated. As such, venture capitalists and public funding agencies often co-finance research projects; academic researchers compete for funding but also collaborate throughout research projects among each other and with industrial researchers (Hicks and Katz, 1996, Godin and Gingras, 2000). In all then, research can be characterised in terms of a system of interactions that take place along various dimensions; involving many different kind of actors (Boschma, 2005, Hessels and van Lente, 2008, Frenken et al., 2009, Hardeman, 2012).

3. Conceptual framework: from national research systems to research excellence

Akin to the idea of national innovation systems (Freeman, 1987, Lundvall, 1988, Lundvall, 2010, Nelson, 1993, Edquist, 1997), a national research system is made up of the actors within a country that jointly (i.e. in interaction with each other) produce research outcomes. It follows from its resemblence, however, that the challenges confronting a systems approach to innovation (Edquist, 2005), equally apply to assessing research at the country level using a systems approach. One is that a systems perspective on research in a country risks to fall into a trap of conceptual unclarity (Markusen, 1999). For example, the concept of institutions is often poorly defined in work on national innovation systems (Edquist, 2005). Hence, in order to avoid conceptual confusion, we need to make sure that the conceptual building blocks of our perspective on national research systems are clearly spelled out. Second, the innovation systems literature has been criticised for lacking theoretical content (Nelson and Nelson, 2002). Part of the lack of theoretical content in system approaches stems from the emphasis on the complexity therein. Given that aspects of research systems are linked with each other in dynamic, nonlinear, and complex ways; it is hard – and perhaps even undesirable – to derive general rules and laws from them. Another part stems from the lack of clearly specified system outcomes (Carlsson et al., 2002, Edquist, 2005). Notwithstanding the complexities involved in the transformation of research inputs into research outputs then, if we are to get to a clearer understanding of systems performance, at least we need to have a clear idea on what is meant by systems performance in the first place (Carlsson et al., 2002, Katz, 2006).

In all, the challenge thus resides in on the one hand making the national research system (and research excellence therein) analytically tractable, whilst accounting for the nonlinearity and complexity involved in research on the other (Carlsson et al., 2002, Lundvall, 2007, Castellacci and Archibugi, 2008, Castellacci and Natera, 2011). We will take up this challenge along two lines. One is that we use a composite indicator approach to measure and monitor patterns and trends in research across countries. The detail of this approach will be discussed more in-depth throughout the following sections (especially section 5), but we stress to say here that a composite indicator approach is especially suited to summarise different aspects of multi-dimensional phenomena in an efficient, parsimonious, and tractable way. The other line follows from the starting point of any composite indicator approach; that is, developing a sound conceptual framework in which the phenomenon to be measured is clearly defined. In what follows in this section, we will address the conceptual framework.

3.1. A characterisation of national research systems

Following the OECD (2002, p. 30) we define research (including experimental development) as 'creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.' As such, we attribute a number of characteristics to research.

One is that research is about a particular kind of activity, namely, creative work that is undertaken in a systematic way. As argued by Godin (2001) there are at least three interpretations of research as systematic. One focuses on the idea of research as an activity that follows inductive, logical steps. In other words, research starts with particular observations and ends with general rules and laws.

Another, though similar interpretation of research stresses that research follows the scientific method. Here, an important characteristic of research as systematic is that it produces outcomes that are reproducible and measureable. Contrary to the first two (epistemological) interpretations of research as systematic, a final interpretation of systematic focuses more on the institutional aspects of research as systematic. It holds that research is of an enduring, programmatic, organised nature. Here, we do not favour one interpretation of research as systematic over another. In our understanding, the production of (new) knowledge follows from research once these activities are non-serendipitous; the non-serendipitous nature of research involves both that resources are devoted to it as well as some kind of structure in search.

Another characteristic of research is that it has a particular goal orientation, namely, increasing the stock of knowledge. As such, research is primarily about producing new knowledge rather than using existing knowledge. This also implies the exclusion of education activities as these are primarily concerned with the dissemination of existing knowledge stock. A final characteristic of research is that its goal orientation is expressed in various types of outcomes (as diverse and diffuse as knowledge about man, culture, and society) with equally different kind of uses. Thus defined, research is a particular kind of activity that in principle can be performed within various domains (going from the sciences to markets to also possibly including the state, the media, and the arts).

As other systems, national research systems contain three core elements (Carlsson et al., 2002, Edquist, 2005): components, relationships, and attributes. First, components are about 'the operating parts of a system' (Carlsson et al., 2002, p. 234). In other words, the people doing research, the organisations providing the environments for doing research, the instruments that are needed to perform research, and the institutions (i.e. norms, rules, and policies) operating in a country that facilitate doing research. In what follows we refer to the components of national research systems as the research capabilities of a country (Van Looy et al., 2006, Cimoli et al., 2009).

Research assets. Along the lines of Castellacci and Archibugi (2008) and Castellacci and Natera (2011), we distinguish among two main dimensions of a country's research capabilities. First, research assets of a country refer to the set of research agents available in a country. Research assets can be further divided into physical (machines, instruments, and laboratories), human (skilled labour) and intellectual assets (knowledge and ideas). Countries that do not reach a certain threshold level of research assets available are less likely to contribute to or catch up with the technological frontier (Perez and Soete, 1988).

Structural capabilities. Another type of research capabilities concerns the structural capabilities of a country. These involve the sectorial and disciplinary composition of a country, as well as its institutional and geographical make-up. Given that the evolution of a country's economic and scientific activities follows a path dependent process (Neffke et al., 2011, Heimeriks and Boschma, 2013), both the sectorial and disciplinary composition of a country determine the extent to which and in which specific research fields a country can perform. In addition, institutions shape the behaviour of research agents. In other words, institutions both enable and constrain the behaviour of research agents in directing their research activities into certain directions and not others.

Interactions. Relationships concern the connections among the components. Relationships among researchers, the organisations they work in, and the institutions that shape their behaviour, bind the research capabilities of country to make it an actual system. In other words, relationships are about the interactions among the components of a system. Hence, following Lundvall (1988) in his description of innovation as an interactive process, we refer to the relationships among the components of national research systems as research interactions.

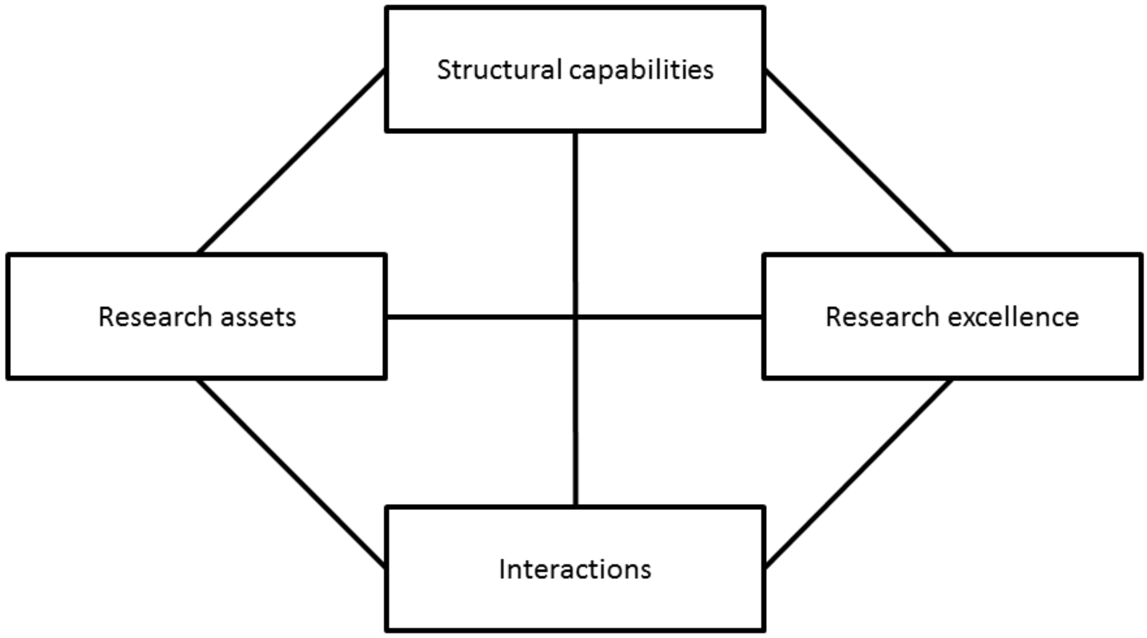
Dimensions to capabilities and interactions. Both the components and relationships that constitute a system have certain attributes or properties. In the context of national research systems, these attributes characterise the nature of the capabilities. For example, when we discuss the norms and rules that operate in a national research system or the organizations that perform research, we dealing with two completely different kinds of capabilities. Likewise, research interactions have different properties. While some interactions concern competitive pressures among researchers, others are about collaborative efforts (Carlsson et al., 2002). At a different level still, interactions in research can be about the transfer of knowledge or the sharing of research facilities. Note that these can be both of a collaborative and a competitive nature. Taken together, both research capabilities and research interactions have various – what we call – dimensions to them.

Following the different dimensions of proximity in innovation (Rallet, 1993, Boschma, 2005, Frenken et al., 2009), we distinguish among three such dimensions. The geographical dimension distinguishes interactions that take place within countries from interactions that take place between countries (i.e. intra-national versus inter-national interactions. The institutional dimension, as in a Triple Helix of university-industry-government relations (Etzkowitz and Leydesdorff, 2000, Leydesdorff and Etzkowitz, 1996), distinguishes between interactions that take place between agents of the same institutional type and those that cross institutional boundaries. The disciplinary or sectorial (i.e. cognitive) dimension distinguishes between interactions that take place within the same cognitive domain and those that take place across cognitive domains. Research interactions that cross sectoral and disciplinary boundaries are often said to render solutions to the grand challenges of our time more likely (Gibbons et al., 1994, Stokols et al., 2008).

Research excellence. Apart from the components, relationships, and attributes; national research systems have a particular goal or orientation (Carlsson et al., 2002). From the definition of research provided above, it follows that national research systems are oriented at the provision of new knowledge. Notwithstanding the difficulties in defining what is new (Witt, 2009), here we take new knowledge to refer to the outcomes of national research systems as excellent. That is, new knowledge is not about the obvious, the straightforward or the usual. Rather, new knowledge is about the remarkable, the original, the striking. In other words, and as first approximation, the prime objective of national research systems is to produce what we call research excellence.

Figure 3.1 pictures the conceptual building blocks of a national research system. Note that, apart from singling out the different components of national research systems, these are in turn interlinked with each other. These inter-linkages, however, should not be understood in causal terms going in one direction. The fact that there is a relation between the components of national research systems need not imply causality between them. Going from the literature on national innovation systems as complex evolving systems, these linkages are to be interpreted in terms of the influence different components of national research systems have on one another. As such, research excellence feeds back into structural capabilities, research assets, and research interactions just as the latter three building blocks of national research systems shape research excellence.

Figure 3.1 Conceptual building blocks of national research system



3.2. A definition of research excellence

Research excellence can be defined in at least two ways (Tijssen, 2003). One way is to refer to the intrinsic quality characteristic of research outputs. The other way is to refer to a comparative quality characteristic of research outputs; in other words, the quality of one particular research product as compared to other research products. Given that no widely accepted standard of intrinsic research quality exists, we are bound to think of research excellence in the latter sense only (Tijssen, 2003).

As such, there are three aspects to the excellence part of research excellence. One aspect concerns its orientation on the outputs of research rather than the inputs to research. A focus on outputs means that, when thinking of research excellence, we are not so much concerned with how many time or money is devoted to research in a country. Rather, what we are concerned with is the products that come out of research that is performed in a country. Another aspect concerns a focus on the quality rather than the quantity of research outputs. The fact that a country produces a lot of research outputs does not in itself imply that this country also produces a lot of good research. In other words, if we are to say something about research excellence we have to start reasoning from the quality aspects to research instead of its quantity aspects. A final aspect concerns an explicit focus on the top-end of the distribution of quality in research outputs. Instead of looking at average research quality, research excellence is about the best quality research outputs available. Research excellence, then, is about all those knowledge producing activities whose outputs are considered of high-end quality (Tijssen, 2003).

Given these concerns, a first issue we are confronted with in trying to map out the different dimensions to research excellence is that there is no agreed-upon ‘one-size-fits-all’ procedure for assigning quality labels to research outputs that stem from different kind of research activities. Remember from the OECD definition that research can in principle be performed by actors that operate within different domains of society; going from the sciences to markets to also possibly including the state, the media, and the arts. However, research performed in different domains of society gets evaluated by different kind of evaluation mechanisms (Stark, 2011). Here a distinction is often made between research that takes place within the realm of science and research that takes place within the realm of technology (Dasgupta and David, 1994).

While scientific research gets evaluated through a system of peer review, technological research gets evaluated on the market in terms of profits and prices. For other domains, such as the state and the arts, however, the mechanisms underlying the evaluation of its research activities are much less clear. At best their evaluation can be monetarised (using prices as in the evaluation of technological research), at worst there is no agreed-upon way of evaluating such research at all. In other words, in thinking of assigning quality standards to research outputs, it is not evident that an agreed-upon quality measure can be assigned to all kinds of research outputs. While for science and technology this might be done given their standardised quality assignment mechanisms of respectively peer review recognition and market prices, for other domains in which research takes place (most notably the state, the arts, and the media) such standardised quality assignment mechanisms are largely absent and hence measuring research excellence here is much more challenging (Bornmann, 2012, Godin and Doré, 2004). Anticipating on the absence of data on research excellence in societal domains other than science and technology, we will restrict our further discussion of research excellence to hold for the domains of science and technology only. Note then that strictly speaking we no longer address and hence measure research excellence at large rather than a restricted part of it; that is, scientific and technological research excellence.

Secondly, even if we have agreed-upon evaluation mechanisms to assign quality to research outputs, this in itself does not imply that such valuations are comparable across both time (i.e. across years) and space (i.e. across countries). In the case of science, although various kinds of research outputs get evaluated by peer review, this in itself does not make these valuations comparable across time, space, and socio-epistemic contexts. For example, a peer-review based quality judgment made in one discipline need not be on equal footing with another peer-review based judgment made in another discipline. In addition, for many of these valuations it is hard to distinguish 'ordinary' research outputs from excellent research outputs. To give an example, conference talks in science are often granted on the basis of peer review. Yet, no standardised records exist on the assignment of conference talks to scholarly works that make the quality of these talks comparable across both years and countries. And then again, even if such records would exist, the mere availability of such records would not ensure the possibility of discriminating ordinary talks from excellent talks.

In general then, the extraordinary use-value of research outcomes can be expressed by the diffusion of a research outcome in the economy and society at large. The more a research outcome diffuses, the more excellent it becomes. As any kind of knowledge, new knowledge can be embodied in people (e.g. tacit knowledge) or inscribed in texts or artifacts (e.g. codified knowledge) (Cowan and Foray, 1997, Foray, 2004, Collins, 2010). The diffusion of research outcomes can take shape along various dimensions (Godin and Doré, 2004). Not only do research outcomes have an impact on science and the economy; so do research outcomes have an impact on norms and values and the way work is organised outside these two domains. Consequently, the nature of excellent research outcomes (i.e. research excellence) is multi-faceted. Detailed measurement and data issues aside (to which we will turn in the next sections), research excellence is about those aspects of systematically performed creative activities that (i) are about the production of new knowledge, (ii) whose products are characterised by their high-end quality which can (iii) be expressed along various dimensions (i.e. scientific, technological, cultural, etc.) and embodied in both people and artifacts.

4. Data

4.1. General measurement issues

Measuring means using data to say something meaningful about a particular phenomenon of interest; in our case research excellence. As such, it is important that the data to be used is of a good quality. The better the data at hand, the more can be said about a particular phenomenon of interest. Following Griliches (1986, cited in Hall and Jaffe (2012, pp. 7-8)), we distinguish among three general types of quality issues: extent, reliability, and validity (see also: Nardo et al., 2005).

Extent. Extent refers to the scale and scope of the data. In other words, the coverage of the data in terms of years, countries (in our case), and the type of research activities and their outputs. In general it holds that the larger the coverage of the data, the better the data is.

Reliability. Reliability refers to the degree to which the data has been collected systematically. This issue amounts to whether differences exist in which data on a specific phenomenon have been collected in the same way across years and countries and are hence comparable across both in time and in space. Often, for example, definitions on measuring a particular phenomenon change over time. In order to make sure that apples are compared with apples and not oranges, we need to make sure that a phenomenon is measured consistently.

Validity. Validity, then, refers to the extent to which the data correspond to the phenomena that that we intend to measure. It is important to note here that all measures are only proxies of the underlying phenomena of interest (Hall and Jaffe, 2012). As a representation of real world phenomena, indicators always leave out some aspects while paying explicit attention to others. Hence, the issue is not whether indicators are imperfect representations of the phenomena we are interested in but more to what extent this is so (Hall and Jaffe, 2012).

Note that in choosing variables, one is often confronted with a trade-off among two or three of the data quality issues described above. For example, getting higher year and country coverage might come at a cost of these data being inconsistently collected across time and/or space. Apart from these trade-offs it is important to keep in mind that every measurement (at least of social phenomena) is imperfect. Measurements represent the state of affairs of a phenomenon of interest, but do not involve the particular phenomenon interest itself. In other words, all measurements are proxies of the complex realities which we are trying to capture. Needless to say, aiming at informing public policy, our intention is not to affect the performance of countries in terms of the variables that we propose. Rather, the intention is to affect the performance of countries in terms of the underlying phenomenon of interest ([1]).

4.2. Measuring research excellence

In the previous section 3, we defined excellence in research as the top-end quality outcome of creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise

([1]) In order to avoid confusion, in the remainder of this report we will speak of indicators when we discuss the proposed composite indicator(s) on research excellence and of variables when we discuss the underlying measures of the composite indicator(s) as much as possible.

new applications. Given the multi-faceted nature of research outcomes, some have argued that 'benchmarking and technological productivity needs to navigate through a difficult terrain between two equally unacceptable extremes: accepting that comparisons are impossible or coming up with nonsensical oversimplifications' (Barré, 2001, p. 259). Yet, recognising our inability to capture all dimensions of research excellence in our measurements (Barré, 2005, Godin and Doré, 2004, Edquist and Zabala, 2009, Bornmann, 2012) does not mean that we should refrain from trying to measure it altogether. Rather, it means that in measuring research excellence in particular ways and not others, we have to make clear which aspects of research excellence we are measuring and which not. In what follows then, we try to make clear our rationale for choosing particular variables for measuring research excellence. In so doing, we follow the threefold structure of extent, reliability, and validity outlined above.

Extent of variables measuring research excellence. Preferably we want to have data that covers the largest set of countries and years possible. At least, this set of countries should cover all 27 EU Member States ([2]). In addition, for purposes of comparing EU countries with its alleged competitors, we need the data to cover OECD countries that are not part of the EU27 and preferably also countries characterised as newly emerging economies like the so-called BRIC-countries (Brazil, Russia, India, and China). Concerning the time-span to be covered, preferably we would like to have recent data going far back in time. Also, the data has to be about the largest range of research outputs that stem from the largest range of research activities possible.

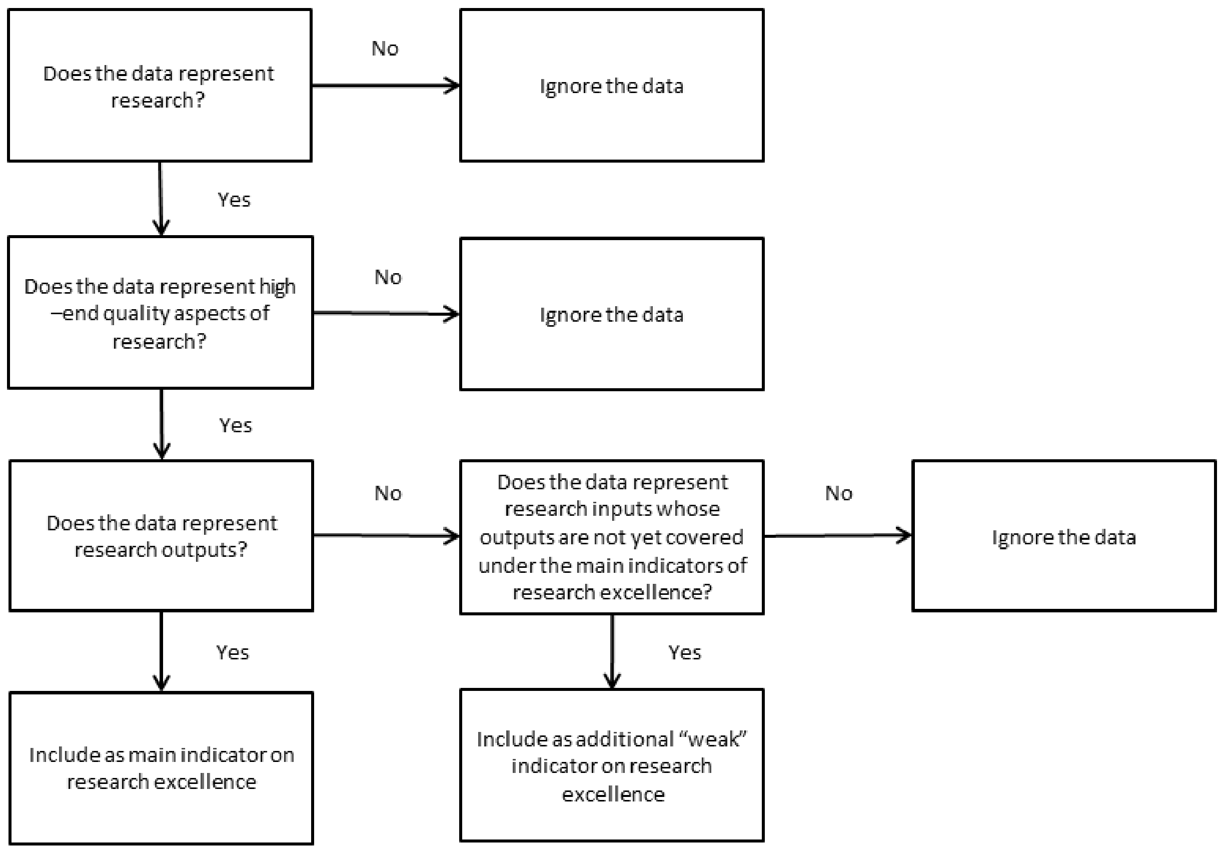
Reliability of variables measuring research excellence. Concerning the reliability of our data, what we need is data that measures research excellence consistently across both countries and over time. That is, the way Germany's research excellence gets measured has to be done in the same way as Italy's research excellence is measured. Likewise, Spain's research excellence in 2003 has to be measured in the same way as Spain's research excellence in 2008.

Validity of variables measuring research excellence. With respect to the validity of our data we need to make sure that we are comfortable with the extent to which the data correspond to our conceptualisation of research excellence. As such, the data that we use preferably reflects four characteristics. One is that it says something about creative, knowledge production activities; that is, research (regardless of the specific domain these activities are performed in). Another characteristic is that this 'something' characterises the output of such creative, knowledge producing activities. In addition, the characterisation of outputs from creative, knowledge production activities should be concerned with quality rather than the quantity of these outputs. As a final consideration, the outputs from creative, knowledge production activities should not characterise overall or average quality; instead, measuring research excellence is about measuring high-end quality research outputs. Overall then, concerning validity, we maintain a few absolute restrictions on the quality of our data. One is that they have to reflect on excellence as defined by high-end quality. Another is that we do not consider research input data which represent inputs to research for which we can measure the outputs. In other words, if we make use of input data these inputs need to cover aspects of research that are not already covered by the output data that are available to us.

([2]) We note that by the time of the publication of this study, the EU had 28 Member States. Nevertheless, during the time of data collection and analysis, the 27-member union was considered as the reference.

Figure 4.1 simplifies our approach into a number of steps. Suppose that we have access to all kinds of data. In the first step then, we throw away all those data that are not about research. In a second step we determine whether the data that represent research also reflect on high-end quality aspects of research. Those data that are not will again be further ignored. In a third step we assess whether the data that represent high-end quality aspects of research also do so on outputs. Those data that do will be included in our set of main variables on research excellence. Those data that do not represent outputs will not be immediately ignored. Rather, if these data cover dimensions of high-end quality in research that are not yet covered by the output data from the set of main variables on research excellence, we will still take them into account as additional variables on research excellence.

Figure 4.1 Flow chart on the selection of variables on research excellence



4.3. Specification of variables measuring research excellence

Following the scheme of Figure 4.1, we considered a set of potential variables to measure research excellence (see Table 4.1).

Table 4.1 List of variables on research excellence composite indicator

Code	Name	Definition	Sources
Variables included			
HICIT	Highly cited publications per total publications	Field-normalised count of the 10 % most highly cited publications divided by total publications with an author from that country (years reliable data available: 2000-2007)	Science Metrix (Scopus)
PCTPAT	PCT Patent applications per million inhabitants	Patent applications filed under PCT by inventors' country of residence (fractional counting) in all IPC classes per million inhabitants (years reliable data available: 2000-2008)	OECD
TOPINST	Top universities and public research institutes per total R & D expenditure	Number of top 250 world scientific universities and top 50 public research organisations in a country divided by total R & D expenditures (years available: 2003-2007, 2004-2008)	SciMago Institute Ranking (Scopus)
ERC	ERC grants received per public R & D expenditure	Value of ERC grants received by country of host organisation, equally spread over project duration divided by public R & D expenditures (years available: 2007-2011)	ERC, DG-RTD CORDIS
Variables not included			
TPF (alt. of PCT)	Triadic Patent Families per million inhabitants	Patent filings at the USPTO and applications at the EPO and JPO by inventors' country of residence in all IPC classes divided by million population (years available: 2000-2010, of which 2007-10 are nowcast)	OECD
OECD Composite (alt. of PCT)	OECD Composite Indicator on Patent Quality	A composite index of Patent Quality, built on a set of normalised indicators, such as backward and forward citations, family sise, number of claims, grant lag and patent generality. (available for time periods: 1990-2000 and 2000-2010, for 28 countries)	OECD (2012)
TOPUNIV1 Leiden	Top universities	Country aggregates of field-normalised share of top 10 % most cited publications output of universities ('pp10'). (Years not comparable; 2011/12 edition based on 2005-09 data)	Leiden CWTS (Web of Science data)
TOPUNIV2 Leiden	Top universities	Country aggregates of mean normalised citation scores of universities ('mncs'). (Years not comparable; 2011/12 edition based on 2005-09 data)	Leiden CWTS (Web of Science data)
TOPUNIV3 Shanghai	Top universities	Country aggregates of top research universities listed in the Shanghai Ranking top 500	Shanghai ARWU (Web of Science data)
HicitSci	Highly cited scientists	Country share in global top 250 most highly cited scientists, based on Index of Shanghai Ranking	Shanghai ARWU (Web of Science
BesFinGHRD	Contractual research	Business financing R & D performed by Government and Higher education divided by GDP	Eurostat, OECD

These variables characterise excellent national research systems in terms of high quality science and technology outputs produced both by public and private actors that are embodied in artifacts and people. Having assessed the quality profile of nine variables, we selected four: the number of highly cited publications published by a country as a share of total publications, the number of high-quality (i.e. PCT) patents by inventors from a country per million inhabitants, the number of world class universities and research institutes in a country as a fraction of total R & D expenditures, and the number of ERC grants received by a country as a fraction of public R & D expenditures. While highly cited publications and high-quality patents reflect new knowledge from respectively science and technology that is attributable to a country and inscribed in texts, the number of world class universities and research institutes and ERC grants are measures of new knowledge from science that is embodied in (groups of) people in a country. In the current section we discuss these four variables in turn and focus on the issues of extent, reliability, and validity therein.

4.3.1. Highly cited publications (HICIT)

The aim of this variable is to measure countries' top-quality scientific output through highly cited publications (Batty, 2003, Wichmann Matthiessen et al., 2002, Bornmann et al., 2011), with a cut-off point established to include the top 10 %. As such, we use a field-normalised count of the 10 % most highly cited publications as provided by Science Metrix which is based on data from Scopus Elsevier.

Extent. A main advantage of using highly cited publications to measure research excellence is that these data are widely available in standardised format across both time and space. As opposed to many other measures of excellent research outcomes (e.g. Nobel Prizes, Fields Medals, and prolific conference attendances), data on highly cited publications are widely available for measuring the research outputs of many countries across many years, and disciplines. Hence, data on highly cited publications are a likely candidate to conform to the requirement of having extent.([3])

More in detail, all publications in Scopus (17.5 million publications considered over the time period 2000-2011) are attributed to a document type, a subfield (or scientific specialties, such as anatomy,

([3]) Note that the threshold of highly citedness is arbitrarily set. While some consider only the top 1 % highly cited publications as representing excellent scientific research outputs, others take a broader view focusing on the top 10 %. Both measurements seem to correlate well with other (more ad hoc) measures of scientific excellence (see: TIJSSEN, R., VISSER, M. & VAN LEEUWEN, T. 2002. Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics,* 54**,** 381-397.). Although we would have preferred to experiment with different threshold levels, the data that is available to us only involves a 10 % threshold.
The number of publications by an entity (e.g., the world, a country, a NUTS2 region, an institution) in the 10 % most cited publications in the database is determined using the relative citation (RC) scores of publications computed using a 3-year citation window following the year of publication. Because some publications are tied based on their RC score, including all publications in the database that have a RC score equal or greater than the 10 % threshold often leads to the inclusion of slightly more than 10 % of the database. To ensure that the proportion of publications in the 10 % most cited publications in the database is exactly equal to 10 % of the database, publications tied at the threshold RC score are each given a fraction of the number of remaining places within the top 10 %. For example, if a database contains 100 publications (i.e., the top 10 % should contain 10 publications) and that the 9th, 10th, 11th and 12th publications all have the same RC score, they are each given a quarter of the remaining two places in the top 10 % (0.5 publications of the top 10 % each). An institution whose publications rank 2nd and 9th would therefore have 1.5 publications in the top 10 % using whole counting (at the level of addresses). Both full and fractional (here there can be fractions of fractions if, for example, the publication in 9th place in the top 10 % has been co-authored) counting of publications are used. The total number of citations for an aggregate (e.g., the world or a country) is obtained by totaling the number of citations of the papers that were assigned to this aggregate. The indicator 'highly cited publication' (HICIT) is then computed by taking the share of highly cited publications to total publications (full counting method) of a given country. In this way, both publications with co-authors in different locations, as well as with authors with multiple country affiliation are attributed to all countries listed in the affiliations.

evolutionary biology, and analytical chemistry) and field (such as chemistry, physics, and biology) by Science Metrix in a mutually exclusive journal-based definition. Each paper's citation count is divided by the average citation count of all publications of the corresponding document type (i.e., a review would be compared to other reviews, whereas an article would be compared to other articles) that were published the same year in the same subfield to obtain a Relative Citation count (RC). This renders data covering the years 2000-2007 (taking into consideration the publication year and a three year citation window) and 41 countries. Scopus Elsevier covers 15 000 scientific journals from all fields of science, including social science and humanities. Nevertheles, the latter fields are less covered by the data than the natural sciences.([4])

Reliability. In addition, the fact that these data are systematically collected also renders them reliable. That is to say, while other variables on scientific rewards (such as for example honorific awards) differ in interpretation from one country and year to another, the interpretation of highly citedness is relatively straightforward and remains constant over time and place. Nevertheless it should be noted that most bibliometric databases (including Scopus Elsevier) are biased in their representation of Anglo-American countries. As such, these data reflect on scientific research excellence of which the standard is set by this part of the world.

Validity. Scientific publications (and hence publication-based data) are an obvious candidate to base measurements of scientific research excellence on (Gilbert, 1978). Scientific publications are both considered an output to scientific research and involve processes of peer review. Provided that citations to publications are attributed according to the norms of science (Kaplan, 1965), the aggregate distribution of science citations provides an indication of the aggregate distribution of quality-based scientific reward (Moed, 2005). If science citation provides a picture of scientific research quality in general, then scientific research excellence is captured by those contributions that are cited most (Tijssen et al., 2002). As data on highly cited publications reflect on high-end quality characteristics of research outcomes, these data conform to our criterion of validity.

Nevertheless, some cautionary remarks need to be made about the validity of data on highly cited publications for measuring research excellence. Critiques on using publication citation data to measure research excellence are concerned with the use of publication data for measuring science in general and the use of citation data to measure quality (and hence excellence) in particular (MacRoberts and MacRoberts, 1996, Bornmann and Daniel, 2008, Edge, 1979). First, the use of publication data to measure research has been criticised in that not all knowledge that is brought about in scientific research eventually ends up in scientific publications. There are other outputs to research that also reflect on the knowledge produced therein. In fact, given the tacit nature of much of the knowledge brought about in scientific research, it is even impossible to make all knowledge explicit in scientific publications (Cowan and Foray, 1997, Foray, 2004, Collins, 2010). Publication data

([4]) Given that publication and citation practices differ across disciplines, differences in the extent to which highly cited publications are attributed to countries might arise just by virtue of different countries being specialised in different disciplines. This raises the question of how to delineate disciplines. While some simply use standard classifications to delineate disciplines, others very much question these standard classifications (see e.g. LEYDESDORFF, L. & RAFOLS, I. 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology,* 60**,** 348-362.). Our position holds that there is no and in fact cannot be one single best classification of disciplines and industries (on some more general problems involved in classifying see: BOWKER, G. C. & STAR, S. L. 1999. *Sorting Things Out: Classification and Its Consequences*, MIT Press.). Preferably then, preferably we would pay attention to whether different disciplinary and industrial classification systems render different outcomes to our analysis (e.g. in terms of rankings and explanations). As of now however, we only have access to data on highly cited publications that are normalised using the disciplinary classification of ScienceMetrix. For the attribution of 15000 journal sources, see the Ontology Report of Science Metrix: www.science-metrix.com/SM_Ontology_103.xls; Retreived: November 2012.

then only reflect on scientific research outputs in as far as these outputs can actually be made explicit therein and are not embodied in people or artifacts. In addition, although scientific publications probably contain some amount of scientific knowledge, this does not imply that all scientific publications contain an equal amount of scientific knowledge. While on the one hand the outcomes of some scientific research projects are presented across multiple scientific publications, on the other hand some scientific publications present results of multiple scientific research projects at once (Gilbert, 1978).

Second, given that science citation data are derived from scientific publication data, part of the critique on using publication data to measure scientific research outputs also hold for the use of science citation data. Hence, science citation data can only reflect on the quality of scientific research outputs in as far as these are made explicit in scientific publications. In addition, the extent to which science citation data actually reflect on the quality of scientific research is contested. On the one hand it is argued that, while scholars do not just cite for reasons of 'giving credit where credit is due' only, the underlying assumption for using science citation data as properly signaling the reward structure of science does not hold (Brooks, 1986, Bornmann and Daniel, 2008, MacRoberts and MacRoberts, 2010). Also, measuring the quality of scientific research on the basis of science citation data might lead to inadvertent consequences such as putting incentives on reaching high citation scores while losing track on producing high-quality scientific research outputs more directly (Weingart, 2005, Burrows, 2012).

Nevertheless, we believe that the number of highly cited publications produced in a country present a valid variable measuring some aspect of research excellence. First, the fact that data on highly cited publications only present a partial picture of (total) research excellence demands that we need to have additional variables that capture additional dimensions of research excellence. In a way then, this further legitimises the need for a composite indicator on research excellence. Second, given that on an aggregate (country) level science citation variables correlate highly with other variables measuring scientific reward, these can at least be considered as valid proxies of the quality of scientific research outputs (Cole and Cole, 1967, Tijssen et al., 2002, van Raan, 2006). As such, if not measuring pure scientific research quality, citation based data at least measure some kind of socially determined scientific research quality (Cole, 1989) or community-based impact (Martin and Irvine, 1983). Possible inadvertent consequences to the use of citation based data for purposes of research policy-making reside as much on the side of those constructing the indicators as it does on the side of policy-makers (van Raan, 2005). What is more, while the issue of inadvertent consequences is very much relevant at the micro level (especially in the context of allocating funding on the basis of citation based research evaluation), it is less relevant once you move to the macro level.

As for measuring research excellence then, we believe that publication citation data provide a viable basis for measuring research excellence at the country-level because (i) it is publication (and hence peer review) based rendering it a relatively direct measure of the quality of scientific research output, (ii) it correlates highly with other (less systematically collected) measures of scientific research excellence, and (iii) at the level of countries it is less likely to steer the research system into undesirable directions. In sum, any potential biases at the micro level are expected to compensate each other at the aggregate country level. In what follows we will discuss the specific characteristics of the data on highly cited publications that we use.

Conclusion. Following the steps from Figure 4.1, data on highly cited publications are (i) clearly about research, (ii) represent high-end quality aspects of research, and (iii) represent not just inputs to but outputs of research. Hence, we consider the number of highly cited publications a strong variable measuring research excellence.

4.3.2. High-quality patent applications (PCTPAT)

The aim of high-quality patent applications is to measure countries' top-quality technological output by looking at top-value patents. We use data from the OECD on patent applications filed under Patent Cooperation Treaty (PCT) by inventors' country of residence (fractional counting), divided by the population of a country in millions.

Extent. PCT patent applications data allows for a cross-country comparison, free from home country bias, as it refers to worldwide patent applications. Data is available for the years 2000-2010. Given that the variable fluctuates considerably across years (particularly for those countries with lower number of patent applications) we choose to use 3-year moving averages (both for patent application and population levels) with the 2010 time point including only the two most recent years. Given that PCT applications data is available for all countries of the world, we included all 41 countries considered in this study. As with the disciplinary coverage of publication data, PCT applications data cover all technological fields.

Reliability. Concerning the reliability of PCT patent applications as a proxy for patent value, some constraints should be highlighted in using this variable to compare patenting over time: prior to 2004, the applicant had to select the list of countries to designate and pay the fee accordingly; countries are generally well represented after 2000, as they gradually started using this procedure over the 1990-2000 transition period – with Japan and South Korea being the slowest adopters of the procedure. The main drawbacks of the PCT patent variable is that economic and legal constraints keep some countries from being able to use the PCT procedure (depends on an initial investment potential), and that PCT applications are options rather than actually filed technologies.

Validity. The definition of excellence in patenting (which is to be used as a proxy for technological excellence) is contingent upon an understanding of patent value. According to the OECD Patent Statistics Manual (Zuniga et al., 2009), patent value can refer both to the economic and the social value of a patent. The first concept refers to the revenue flows to its holder, the second to the patent's contribution to the stock of technology. Consequently, if research excellence is measured through patents, in line with the understanding of excellence as top of a statistical distribution, the most outstanding patents are distinguished by the very high revenue they generate, or their outstanding technological content. These two features may not necessarily coincide, as the revenue generating potential of a patent depends not only on the technological content of the invention, but also on whether the patent can be circumvented, which depends on the technological field and the market position of its owner. Yet, despite the bias, it would be problematic to dismiss the economic value from an understanding of technological research excellence, as the revenue generating potential is an important driver for research actors to patent new inventions. A common way to single out high-quality patents is to count those that have a broad geographical scope; that is, counting patents that were filed in patent offices across different countries (Putnam, 1996). If an applicant is ready to pay the additional costs of protecting the invention in many countries, it implies that the applicant expects that the patent will generate sufficiently high revenues. Inventions filed under the PCT are administered by the World Intellectual property organisation (WIPO), and the procedure designates all signatory states of the PCT.

Conclusion. Following the steps from Figure 4.1, data on high-quality patent applications (i) clearly are about research, (ii) represent high-end quality aspects of research, and (iii) represent not just inputs to but outputs of research. Hence, we consider the number of high-quality patent applications a strong variable measuring research excellence.

4.3.3. World class universities and research institutes (TOPINST)

The aim of the variable proposed here is to measure the presence of 'centres of excellence' that produce outstanding scientific research in a country. Such centres of excellence are universities and research institutes that attract outstanding researchers (senior and junior) from around the world, offer state-of-the art infrastructure and potentially (due to reputation based past excellence) attract a large share of public and private research funding, all of which create the virtuous circles or self-reinforcing mechanisms that sustain their excellence. Their favourable financial and scientific position allows such institutes to be strategic actors, at least partly independent from the general state of research climate of a country.

We chose an indicator that combines top universities based on the **Leiden Ranking** and Top public research organisations based on the **SciMago Institutions Ranking**. The Leiden Ranking, based only on bibliometric data, is the most suitable among rankings to measure research performance and excellence of universities. As the Leiden Ranking does not cover public research institutes which characterise the continental research systems, we opted for augmenting the university dataset with **SciMago** Institutions Ranking (using exclusively the ranking of research institutes). In this way the top of the distribution is selected from each source: the 250 global universities and 50 research institutions, all based on their research impact, resulting in a pool of 300 global centres of excellence.
Extent. The main issue with such ranking-based data is comparability over time. Global bibliometric data coverage has been increasing parallel to methodological improvements in the computation of institute rankings. It is therefore problematic to compare the results of rankings over time, also when aggregated to the country level. For instance, the 2011/12 edition of the Leiden Ranking (based on Web of Science data for period 2005-09) explicitly discourages the comparison of its results with those of previous editions, as it covers a larger pool of universities, uses a different methodology and excludes arts and humanities publications. This difference is apparent when summing up the number of universities within the global top 250 for a country: the number of universities for South Korea, Japan, Italy and Brazil drops sharply, while it increases dramatically for Norway and China ([5]). (For any future update of this indicator the use of the latest ranking may be considered, we would very much welcome the retrospective computations of the rankings according to the most recent methodology to make them comparable over time.) Results of the SciMago Institutions Ranking (IR) are similarly difficult to compare over time: we found that the availability of rankings based on the 'Q1' variable, (the Ratio of publications that an institution publishes in the scholarly journals that are ranked in the first quartile (25 %) in their categories as ordered by SciMago Journal Ranking, based on citations)

([5]) We find similar problem for earlier years: the 2007 edition of the Leiden Ranking covers only the top 100 universities of Europe, making global comparison impossible.

makes the 2010, 2011 and 2012 editions comparable, but this indicator is missing in previous editions. An even more appropriate indicator would be ranking based on the Excellence score (based on publications in the top 10 % most highly cited publications), but this is missing in the 2010 edition so we could not use it for historical comparison – but could nevertheless consider in the future. Table 4.2 gives an overview of the time coverage and scope of the various editions of the rankings we considered.

We note that SciMago ranks both public and private research institutes alongside universities and university hospitals, but remarkably only 3 private institutes were found among the global top 500. In addition, we see a potential headquarter bias in the attribution of research output of large companies with R & D locations around the world. We therefore do not include companies in this indicator.

Table 4.2 Comparison of time coverage and scope of rankings considered

Ranking	Edition	Source	Time Coverage	Scope
Leiden Ranking	2008	Web of Science	2003-2007	Universities with min. 700 publications
Leiden Ranking	2009	Web of Science	2004-2008	Universities with min. 700 publications
Scimago IR	2010	Scopus	2004-2008	2833 institutions with min. 100 publications in 2008, (covers 80 % of all publications)
Scimago IR	2011	Scopus	2005-2009	3042 institutions with min. 100 publications in 2009, (covers 80 % of all publications)
Scimago IR	2012	Scopus	2006-2010	3299 institutions with min. 100 publications in 2010, (covers 80 % of all publications)

NB: Leiden Ranking can be accessed at URL: [http://www.leidenranking.com/]; the SciMago Institutions Ranking can be accessed at URL: [http://www.scimagoir.com/]

Reliability. In line with the considerations on measuring excellence discussed above, world class universities and research institutes are identified by size- and field-normalised citation scores from bibliometric data, which makes TOPINST an indicator of high quality of research. Apart from the issue of coverage over time, another potential source of problems for this indicator could be the use of two different data sources. Our tests showed that this is not a significant bias. We noted that the SciMago IR also included not only research institutes but also universities, which were, however, different from those in the Leiden Ranking. After matching the top universities in the datasets of SciMago and Leiden referring to the year 2010, we found a relatively modest rank correlation (0.30) between the two. Given that the Leiden Ranking captures the top 10 % rather than the top 25 % of SciMago, the former was seen as more appropriate to obtain rank information on universities. At the same time, the correlation between the country aggregate scores computed on the Leiden Ranking only (covering universities) and the SciMago ranking only (covering universities and public research organisations) was found very strong (0.98), also when scale-normalised by GDP (0.89).

The threshold cut-off point for top universities could also affect reliability of this indicator. Tests indicated that in fact, cut-off point matters: the lower the threshold, the better European countries fair in terms of this indicator as more European universities are included. Thus, the US outperforms EU27 when the top 250 universities and research institutes are considered, but European countries along Brazil, South Korea and China would seemingly ‘forge ahead’ within the top 500. At the same time, it is remarkable that a few European countries (Cyprus, Estonia, Ireland, Lithuania, Luxembourg, Latvia, Malta, Portugal, Romania, Slovenia, Croatia, Turkey and Iceland) have no institutes within the top 250+50.

Validity. Centres of excellence are universities and research institutes that attract outstanding researchers (senior and junior) from around the world, offer state-of-the art infrastructure and potentially (due to reputation based past excellence) attract a large share of public and private research funding, all of which create the virtuous circles or self-reinforcing mechanisms that sustain their excellence. Their favourable financial and scientific position allows such institutes to be strategic actors, at least partly independent from the general state of research climate of a country. Overall, measuring world class universities and research institutes in a country reflects one form of scientific research excellence embodied in people and research groups. The variable is representative for different kinds of research systems, those relying on public research institutes and those on universities. For simplicity, the indicator measures the count of universities and research institutes within the selected sample. In this way, however, no difference is made between number one and number two hundred and fifty; we do not consider it as a problem as all of these institutes are well within the global top five percent universities and research institutes; thus all of them can be considered excellent. In the future, one could test taking into the actual scores of these institutes when computing an update of this indicator.

Conclusion. Following the steps from Figure 4.1, data on world class universities and research institutes are (i) clearly about research, (ii) represent high-end quality aspects of research, and (iii) represent not just inputs to but outputs of research. However, given that this variable is biased against countries with less but predominantly larger size institutes, we consider this a relatively weak but viable variable of research excellence.

4.3.4. ERC Grants received (ERC)

Excellent knowledge embedded in researchers and research teams can also be measured through research grants. The most prominent (high value and prestige) research grants, such as that of the European Research Council (ERC) or the National Science Foundation (NSF) of the United States are awarded based on demonstrated outstanding past performance of research teams on the one hand, and on expected outstanding performance on the other hand. Receiving such a grant can therefore be at the same time a proxy for recent excellence and 'excellence in the making'. This 'bridging property' between past and future results makes it a rather timely variable measuring research excellence. Of course, one has to take into consideration the uncertainty of research outcomes and the potential of failure of granted projects. Here, we use data on ERC grants as provided by the Directorate-General for Research and Innovation of the European Commission (DG RTD) from the Community Research and Development Information Service (CORDIS).

Extent. Time coverage of ERC grants data is rather limited. The ERC is a relatively recent institution and the first money flows to projects began in 2008. ERC supports both early and advanced career scholars with its Starting Grants and Advanced Grants instrument. While in the first year after ERC's launch only starting grants were awarded, in the second year only advanced grants, and it has been only since the third year (projects starting in 2010) that both instruments are available. Country totals for both grant value received and number of projects received in 2008 and 2009 correlate very highly. As for the coverage of countries, ERC grants are biased towards ERA countries. Therefore we decide that, whenever we include this variable, we do this for a separate composite indicator covering the ERA countries only.

Reliability. The main problem with using research council grants as a measure of excellence is international comparability. Both the ERC and the NSF, national research councils even more, are prone to the home country (or region) bias. The supranational characteristic makes the ERC a good candidate for comparison of European countries, but, even if participants from outside Europe can join research teams, it cannot be fairly applied for global comparison.

In order to avoid high annual fluctuations, we decided to spread multi-year projects equally for the years of their duration, starting at the year of awarding (i.e., a 3-year project of EUR 3 million granted in 2007 is accounted as EUR 1 million for each year between 2007 and 2009.) We do notice that the level of funding differs across disciplines; this effect however is found to average out at the aggregate level.

Validity. Another issue is the multitude of interpretation of the variable. It can be at the same time signal financial input to research, and thus characterise capabilities of research actors; signal outcomes of past research performance; and signal socio-political exposure given to selected researchers. While this property makes this variable more problematic for input-output analysis, we still consider it useful to include in a composite indicator as what is common in all potential interpretation is the element of being associated with excellence. In all, ERC grant are clearly about high-end quality aspects of research. However, given that ERC grants are not unambiguously about high-end research outputs, we consider it to be a weak variable measuring research excellence.

As an alternative, we considered testing not only the sum of grant value received by a country, but also the total number of projects. The two values correlate highly (when multi-year projects are both in funding and counts are spread over time), thus we found the additional information redundant.

Conclusion. Following the steps from Figure 4.1, data on ERC grants received are (i) clearly about research, (ii) represent high-end quality aspects of research, and (iii) represent not just inputs to but also outputs of research. However, given that this variable has low coverage in terms of countries and years and given that these data do not just reflect on research outputs but also on inputs, we consider this a relatively weak but viable variable measuring research excellence.

4.4. Alternative variables considered but not included in the analyses

Apart from considering the variables discussed above, we also considered to include other variables in the analysis. Some of these variables are closely related to the ones discussed above, others are ‘real’ alternatives. In what follows we will discuss them in turn.

4.4.1. Triadic patent family data as an alternative to PCT data

Triadic patent family (TPF) is defined as a set of patent applications filed at the European Patent Office (EPO) and the Japanese Patent Office (JPO), and granted by the US Patent and Trademark Office (USPTO), based on the assumption that the most valuable inventions deserve protection in the market of the global economic ‘triad’. The main advantage of PCT count-based data is their relative timeliness (applications are published 18 months after priority; see Table 4.3) and a global coverage. The triadic patent family variable, similarly to PCT applications, is relatively free from the home country bias. However, its main drawback is timeliness: due to the fact that unlike EPO and JPO, USPTO has not published data on applications prior to 2001, the data considers patent grants from the US. This, however, significantly hampers timeliness of the variable, as the average time lag between an application and granting is 35 months, which is further delayed by reporting and computation lags. A way to overcome this is applying the method of ‘nowcasting’ (see also: Dernis, 2007). This estimation technique is adequate for most OECD countries, but has shortcomings with respect to small patenting countries and fast-growing countries (i.e. China or India). In the same vein, the citation-based patent data are facing a severe limitation of timeliness, due to the application of a recommended 5 years of citations window (OECD, 2012).

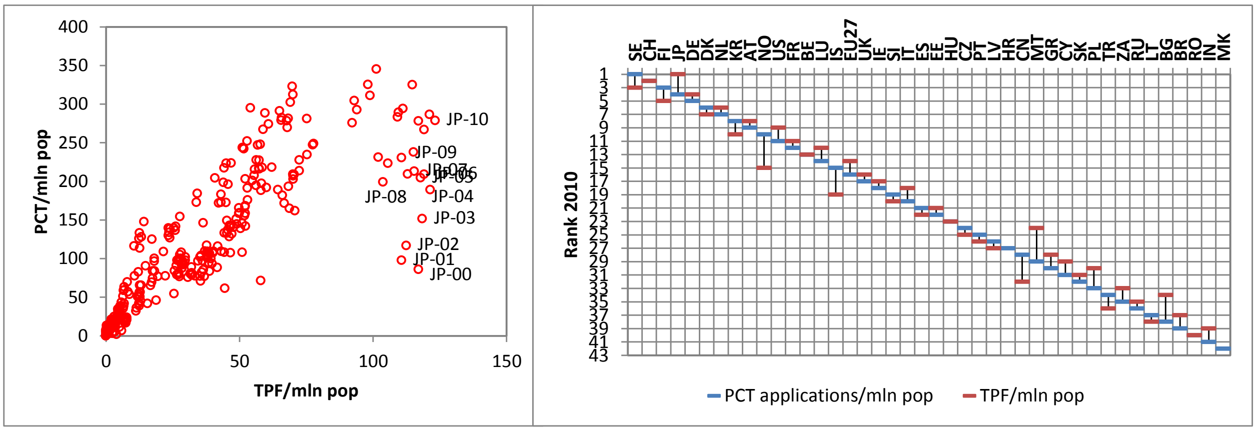
Comparing the differences in the quality profile of PCT patent applications and Triadic patent families (TPF, nowcast by OECD) we found that considering the 2000-10 period, TPF and PCT applications are highly correlated (0.90), thus it is sufficient to choose only one of them (see also Figure 4.2). Timeliness is a main consideration in favour of choosing PCT, as these are less dependent on nowcasting (see Table 4.3). A few countries rank differently for the two variables: most notably, Japan is underestimated if PCT applications are counted, in particular for the years before 2008 (left panel of Figure 4.2), showing all countries and all 11 years considered), while China, Scandinavian countries appear to be underestimated if TPFs are counted. We noted that TPF data was missing for Croatia. Overall, given that TPF data is poorly available for recent years, we choose to use PCT data.

Table 4.3 Timeliness of patent quality variables considered

Variable	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
PCTPAT	(x)	(x)	(x)	(x)	(x)	x	x	x	x	x	x	x	x	x	(x)	(x)
TPF	x	x	x	x	x	x	x	x	x	x	x	x	(x)	(x)	(x)	(x)
OECD Composite						x										x

NB: x = data available; (x) = data available, but weak for some countries; blank = data not available

Figure 4.2 Comparison of PCT applications and Triadic Patent families (TPF) (both per million inhabitants)



4.4.2. The 'Shanghai ranking' as an alternative to SciMago data

Considering that country aggregate positions may change when using other university rankings, we considered including the relevant variables of the ARWU, or Shanghai Ranking, seen as a robust measure of research performance (c.f. Saisana et al, 2012). We found the ARWU variables not to qualify for additional variables of research excellence in line with the considerations outlined in section 4 for the following reasons: (i) many of the research oriented variables of ARWU are biased towards natural-sciences, i.e. staff with Nobel prizes and Fields medals, publications in the journals Nature and Science and the number of highly cited scientists. In addition, (ii) the top world scientific prizes and medals variable covers only a very small fraction of scientists and similarly limited is the score based on the number of the top 250 most highly cited scientists. Furthermore, considering the most highly cited scientist scores, we see (iii) lack of comparability over time due to changes in the sample size (a scientist flagged to be part of the top 250 is counted in subsequent years in addition to the top 250, and the exact number is not published; moreover, the number of universities with scores for this variable above 0 varies hugely over time). For these reasons, we did not include information from the Shanghai ranking.

4.4.3. Excellence in contractual research

The variables considered so far have not covered excellent output of scientific and technological research activity that are not published or patented, such as research carried out on a contractual basis for private purposes. We decided not to include such a variable for conceptual reasons as well as due to measurement issues.

First, it can be argued that contractual research with a high private value (i.e., reflected by high amount of funds payed by a private commissioner of research) may be considered excellent similar to how private value qualifies PCT patents. However, unlike patents, there is a lack of a systematic peer review or broad quality control of contractual research deliverables. There is thus no guarantee that the more money spent on contractual research will deliver better quality outcomes. There is a trade-off: time devoted to contractual research is not devoted to research that is destined to test the scrutiny of peers. Second, similar to the considerations for ERC grants, it can be argued that the ability of a research system to attract private funding contributes to sustain excellence and in this way measures not only past results but future outcomes. However, while ERC grants (or other high-profile research international grants) have an explicit aim to produce results that can be published and used by a broad set of actors of a research system, the results of contractual research reaches a potentially narrow audience. Contractual research outcomes can indirectly lead to peer-reviewed research outcomes, if subsequently (and contracts permitting) researchers make the effort, but that requires additional time.

From a measurement point of view, it is very difficult to identify high-quality outcomes of contractual research. If contractual research is measured by R & D flow from business to public performers of research (universities or public research organisations), it is difficult to measure quality. Should high quality be proxied by high private value measured by the amount of project funding, there is no data available on the project level that would allow the identification of i.e. the top 10 % most costly projects (normalised by field). If excellent outcomes of contractual research are reflected in highly cited public-private co-publications, then such publications would be double-counted considering the HICIT variable already used.

Finally, in our view, in line with the framework presented in section 3, the excellent outcomes of scientific and technological research should be distinguished from knowledge diffusion or valorisation. It can be further investigated whether greater excellence is associated with more contractual research funds ([6]).

([6]) For instance, it remains an interesting research question whether there is a difference between this association in short and long terms. It is possible that in short term, researchers can valorise the outcome of research results accumulated in the past years, but in the long run private partners will select researchers or institutes with a better track record of highly cited publications or other indicators of excellence – unless the routines and common understandings developed in partnerships make it more difficult or costly to change partners.

5. Methodology

Whilst research excellence is a multidimensional, complex phenomenon, it is the strength of a composite indicator approach to make this multidimensionality and complexity analytically tractable. In fact, the added value of a well-constructed composite indicator lays in its ability to summarise different aspects of research excellence in a more efficient and parsimonious manner than is possible with a collection of relevant variables taken separately (Nardo et al., 2005). This composite indicator could then be used to (i) monitor trends in research excellence across countries and over time; (ii) assist with benchmarking and performance assessment; (iii) provide data-driven input to policy formulation and implementation; and (iv) enhance the use of research-related data systems for policy analysis and research. In this section we construct the composite indicator on research excellence following the procedure as proposed in the 'Handbook on Constructing Composite Indicators' from step 3 onwards (Nardo et al., 2005). ([7])

5.1. Descriptive statistics

In the previous section (section 4) we proposed to use four variables to measure countries' research excellence: highly cited publications, high-quality patents, world class universities and research institutes, and ERC Grants. The quality profile of the variables showed that not all four variables can be used to cover all countries for all years. The main problem with using European Research Council grants as a measure of excellence is international comparability. The supranational characteristic makes the ERC a good candidate for comparison of European countries, but, even if participants from outside Europe can join research teams, it cannot be fairly applied for a global comparison. Nevertheless we chose to construct only one composite indicator and flag those countries for which there is no data on ERC grants.

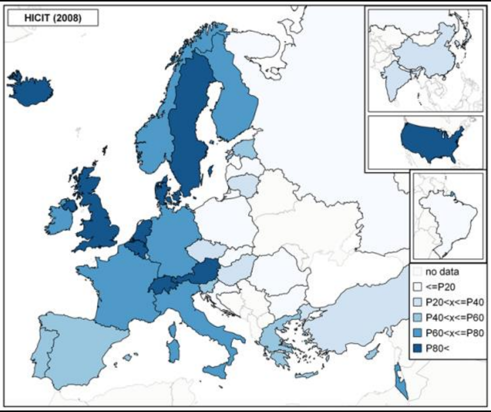
Table 5.1 Descriptive statistics of the four selected variables of research excellence

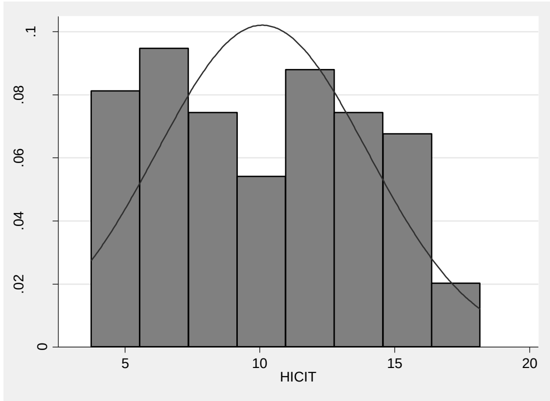
Indicators:	**HICIT** (Number of highly cited Publications per total publications)	**TOPINST** (No. of Top 250 universities and Top 50 research institutes per GERD)	**PCTPAT** (Number of PCT Patent Applications per mln. population)	**ERC** (Value of ERC project grants, per public R & D expenditure)
Years considered:	**2003 & 2007**	**2008 & 2010**	**2004 & 2008**	**2008 & 2010**
Nr. of Observations	82	82	82	68
Missing[a] (%)	0 %	0 %	0 %	17 %
Min	3.8	0.0	0.8	0.0
Max	18.2	927.0	348.9	14.3
Mean	10.1	166.1	82.4	2.1
Standard Deviation	3.9	207.4	92.8	3.0
Skewness	0.1	1.7	1.1	2.3
Kurtosis	-1.1	2.9	0.2	6.0

([7]) The first two steps are about developing a conceptual framework (step 1) and selecting appropriate variables (step 2) and have been discussed in respectively sections 3 and 4 of this report.

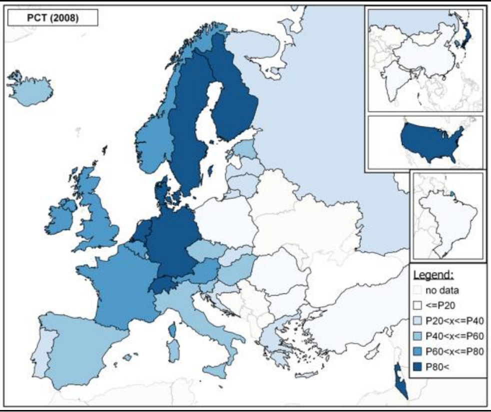
Figure 5.1 Map and distribution of country scores for the four variables of research excellence

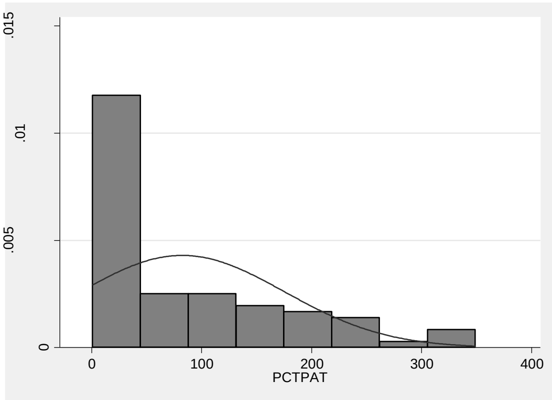
HICIT ('2010')



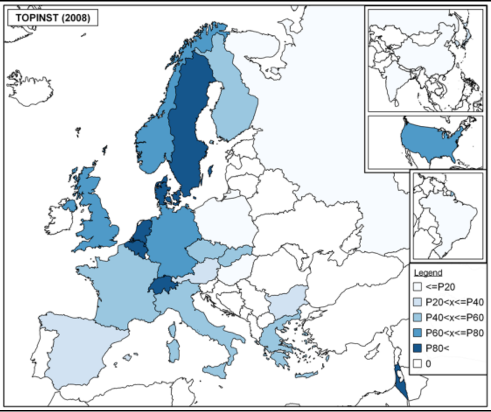


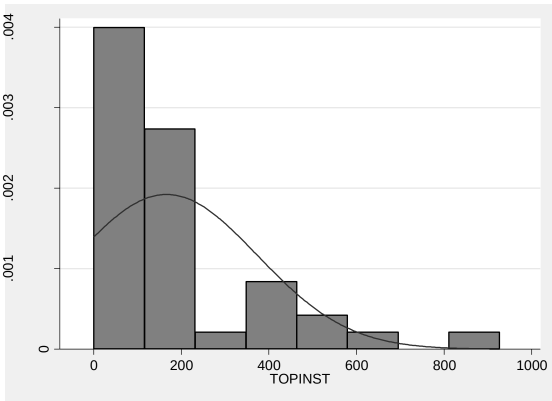
PCTPAT ('2010')



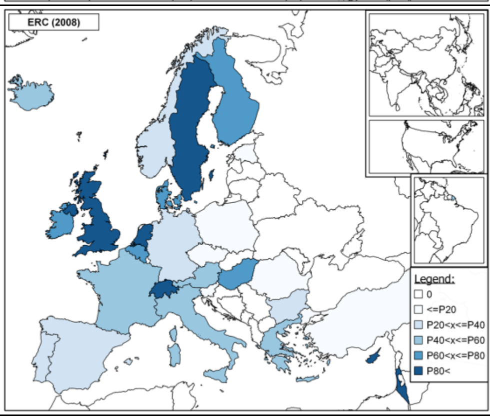


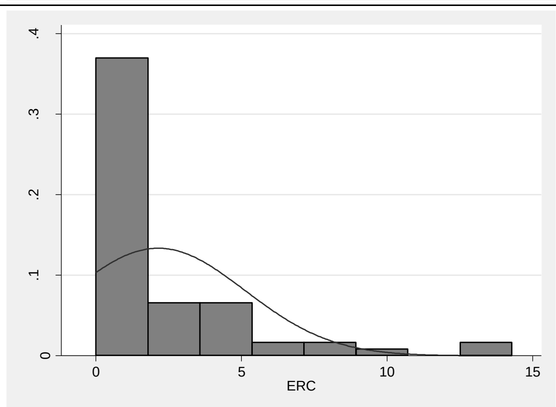
TOPINST ('2010')





ERC ('2010')





NB: Colour tones for maps represent to 20-percentiles, darker tones indicate higher values.

Table 5.1 presents the descriptive statistics of the four variables on research excellence for the sample of 40 countries and EU27 in the years '2005' and '2010" ([8]). Except for the data on ERC grants, all variables have a complete coverage. As noted above, ERC grants are only available for ERA-countries and hence render 17 % missing observations. Note further that all variables are extremely skewed in their distribution. Given their skewed distributions, we need to normalise all 4 variables on research excellence. Scale-normalisation is necessary given the issue of scale driving the amount of research excellence produced by countries. In general: bigger countries produce more research excellence. Hence, we choose to divide highly cited publications, high-quality patents, top institutes and ERC grants by respectively the total number of publications, the population of a country (in millions of inhabitants, total R & D expenditures, and public R & D expenditures. Figure 5.1 maps the four variables included in the analysis.

5.2. Multivariate analysis

The skewness and kurtosis levels for three of the four variables are within the generally acceptable range (absolute skewness<2 and kurtosis<3.5, see: Groeneveld and Meeden, 1984). The ERC variable still has outliers, notably Switzerland and Israel. In order to ensure that the outliers in the distribution of ERC grants per GDP do not drive the results of our composite indicator, we decided apply a logarithmic transformation of this variable. The log-transformation is used when data ranges are positively skewed and shrinks the right side of the distribution. An effect of this transformation is that score increases in lower values weigh more than identical score increases in higher ranges. After the log-transformation, both absolute skewness and kurtosis drop within the acceptable range (0 and 1.6 respectively).

Table 5.2 Pearson correlation coefficients for the Research Excellence variables

	HICIT	TOPINST	PCTPAT	ERC[a]
HICIT	1 ***	0.674 ***	0.745 ***	0.6320 ***
TOPINST	0.674 ***	1 ***	0.758 ***	0.526 ***
PCTPAT	0.745 ***	0.758 ***	1 ***	0.465 ***
ERC[a]	0.632 ***	0.526 ***	0.465 ***	1 ***

NB: *** = significant at 1 %; 40 countries (and EU27 weighted average) at 2 time points combined; a) correlation scores after outlier treatment

Composite indicators cannot be constructed based on anti-correlated or uncorrelated variables. As shown in Table 5.2, all variables correlate highly and positively. At the same time, none of the **correlation coefficients** are excessively high to make a variable redundant. ERC grants show the lowest correlation with the other three variables capturing research excellence (correlations ranging

([8]) The database contains information for 40 countries (and the EU27 average). The ERA countries covered are: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom, Croatia, Turkey, Iceland, Norway, Switzerland and Israel. The non-ERA countries covered are: Brazil, China, India, Japan, Russia, South Korea and the United States. Note that the underlying variables of the composite are not on par in terms of the exact years covered by the data. Highly cited publications, high-quality patents, top institutes, and ERC grants are available for respectively 2003 and 2007, 2004 and 2008, 2008 and 2010, and 2008 and 2010. As such we considered the composite indicator to roughly cover the period 2005-2010 whereby 2005 and 2010 ought to be placed within quotation marks.

between.47 and.63). This either indicates that what the ERC indicator captures is somewhat different from our phenomenon of interest (research excellence) or indicates that ERC captures a somewhat different dimension of research excellence. The correlations patterns shown in Table 5.2 are based on combining data for both time points.

5.2.1. Cluster analysis

Cluster analysis is a multivariate technique that allows for classifying a larger amount of data into meaningful piles based on similarity patterns within a dataset (Kaufman and Rousseeuw, 2009). We use this technique to analyse which countries can be seen as peers in terms of research excellence based on the different variables that populate the composite indicator. As for the composite indicator, we performed a cluster analysis on the sample of countries (n=40) including variables that do not contain missing values. This means that the cluster analysis is performed on three variables instead of four, since ERC grants are excluded due to missing data for non-ERA countries. All cluster analyses are performed on variables of the research excellence index '2010'.

Based on the information of the variables of research excellence, countries are statistically grouped in clusters in a way that the degree of association between the countries is maximal if they belong to the same cluster and minimal otherwise (Kaufman and Rousseeuw, 2009). Consequently, the members of each cluster are more similar to each other than to members of other clusters. We used a hierarchical clustering procedure (Ward's method) to determine the appropriate number of clusters and then performed k-means clustering methods to allocate the countries in these clusters.

Cluster analysis on the full sample of countries excluding ERC grants. The cluster analysis on the full sample of countries generates three clusters. A first cluster groups five countries containing Scandinavian countries (Sweden, Denmark), Switzerland, Israel, and the Netherlands. This group of countries corresponds to the best performing countries in terms of research excellence. Not surprisingly, these countries perform well on all the indicators of research excellence and report averages that outperform those of the other cluster means. A second cluster represents 14 countries including some West-European countries, Scandinavian countries (Norway, Finalnd), Eastern and South-European countries (Italy, Czech Republic) and the United States, Japan and South Korea as international benchmark countries. These countries report high to moderate performances in research excellence. Finally, a third cluster of countries (n=21) contains the bulk of Central and Eastern European countries, Mediterranean countries, BRIC countries and EU candidate countries. These countries are characterised by relatively poor performances in all variables of research excellence.

Rather than a mere identification of clusters, this analysis allows to benchmark a country with its immediate peers. For this purpose, for each cluster we list the country with the highest score on a particular variable measuring research excellence. Consequently, we suggest that countries lagging behind in terms of research excellence should first attempt to improve their performance up until the level of research excellence of their immediate peers prior in focusing on longer term achievements.

As argued before, the variables are proxies of the complex realities which we are trying to capture. Aiming at informing public policy, our intention is not to affect the performance of countries in terms of the variables that we propose. Rather, the intention is to affect the performance of countries in terms of the underlying phenomenon of interest. As such, from these benchmarks, we do not intend to suggest that countries lagging behind in terms of research excellence and hence belonging to the third cluster, should first attempt to improve for example the number of highly cited publications per GDP up to the best level within this specific cluster prior in focusing on longer term achievements of the second cluster. Hence, the countries listed in Table 5.3 do signal for each cluster the countries that score highest on particular variables and as such can be taken as immediate peers of the other countries in that cluster. For example, for countries in the least performing cluster (cluster 3) it makes little sense to take the overall best performing country (Switzerland) as their peer. Instead, depending on where they want to focus on, they could take top performing countries in their corresponding cluster as their benchmarks for improving their research excellence.

Table 5.3 Clusters of 40 countries based on 3 variables of research excellence in 2010

	Cluster 1 **n = 5 countries**	**Cluster 2** **n = 14 countries**	**Cluster 3** **n = 21 countries**	**Entire dataset** **n = 40 countries**
List of countries	CH, DK, IL, NL, SE	AT, BE, CZ, DE, FI, FR, GR, IT, JP, KR, NO, SK, UK, US	BG, BR, CN, CY, EE, ES, HR, HU, IE, IN, IS, LT, LU, LV, MT, PL, PT, RO, RU, SI, TR	
List of indicators	**Benchmark country with highest score per variable**			
Highly cited publications	CH	BE	IS	CH
PCT patents	SE	FI	IE	SE
Top institutes	CH	BE	BG	CH

NB: The cluster analysis is based on a k-means clustering method using 3 clusters (determined by Ward's clustering method). Countries within clusters are listed by alphabetical order. All variables are defined according Table 4.1.

5.2.2. Principal component analysis

Principal component analysis (PCA) was carried out in order to detect the number of latent phenomena described by the four research excellence variables. This is a crucial step before the aggregation of variables into a composite indicator, since a given set of variables that are associated with a latent factor should ideally show a unique, most relevant principal component which accounts for a large amount of variability. Such a result would provide further justification for summarising the variables by a single combined measure (a composite). In the ideal case where all variables are placed on equal footing with respect to the overall index, then they should contribute roughly in the same extent to the variation of the overall scores and with the same orientation to the first principal component). This type of information can be derived from the strength of the correlation between the variables and the first principal component. Note that the higher the correlation, the less influence any adjustment of weights may have on their aggregation (Hagerty and Land, 2007, Nardo et al., 2005, Michalos, 2011).

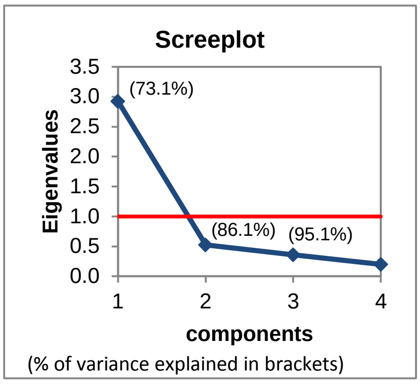
We carried out principal component analysis in order to test whether the four variables are associated with a single latent dimension. PCA results statistically confirm that there is a single latent dimension within the 4 variables, hence the variables express different aspects of the same phenomenon (see Figure 5.2). This result supports the aggregation of the four variables into a composite indicator.

Figure 5.2 Results of the principal component analysis carried out on 3 variables ('2005' & '2010')

				N=68
Principal Component	Eigenvalue	Diff.	Prop.	Cumm.
1	2.92	2.40	73.1 %	73.1 %
2	0.52	0.16	13.0 %	86.1 %
3	0.36	0.16	8.9 %	95.1 %
4	0.20		4.9 %	100 %

	Component Coefficients	Correlation
Variable	First Principal Component	First Principal Component
HICIT	0.510	0.872
TOPINST	0.507	0.867
PCTPAT	0.532	0.910
ERC	0.447	0.764

NB: 2 time points 2005 and 2010 considered simultaneously



We observe strong, positive correlation coefficients, ranging 0.77 to 0.91 between the variables and the first principal component, which explains 73 % of variability in the three variables (Figure 5.2). As the component coefficients suggest, the first principal component is a weighted average of the four variables, whereby the two strong variables of research excellence (HICIT and PCTPAT) as well as TOPINST receive a slightly higher weight than ERC. Overall, from a statistical perspective, we take all four variables to contribute to a composite indicator on research excellence as (i) they capture sufficiently different aspects of the same phenomenon whilst (ii) still being coherent enough for capturing the same phenomenon.

5.3. Normalisation, weighting and aggregation

Normalisation with the min-max method. The variables were normalised into a common scale of 10-100 by applying the 'min-max' method (across the two time points considered). The minimum was set to 10 rather than 0 in order to allow geometric aggregation. The further away a country scores with respect to a maximum of a given variable, the more it needs to catch-up in a certain aspect of research excellence. We note that an implication of this most commonly used normalisation method is that whenever new data becomes available that lies outside the previous minimum and maximum range will result in changes in the normalised scores.

Weighting. Weights of different underlying variables of a composite indicator can on the one hand express certain policy priorities associated with the concepts the various variables measure, and, on the other hand, can also serve as a statistical way to correct for certain properties of the variables. From a statistical point of view, the high correlation among the four variables and the rather balanced component coefficients justify the use of equal weights for all variables. At the same time, the robustness analysis for the composite indicator addresses the potential impact of applying different weights for the various variables. Consequently, we opted for equal weights. In an alternative calculation, we elicited weights using a cross-efficiency data envelopment analysis. This method is described in the section on sensitivity analysis later on.

Aggregation method. The four variables were aggregated using the geometric average. This method is often used for composite indicators where one does not allow countries to entirely compensate for a relatively weaker performance in one aspect of research excellence with a stronger performance in another aspect. A statistical consideration to this conceptually-driven choice on the arithmetic average comes from the principal component analysis, whereby although a single latent dimension in the four variables was identified, the correlation coefficients between the principal component and the variables were not entirely equal. Considerate of the somewhat arbitrary nature of the choice of aggregation method in the robustness analysis, we also address this issue.

6. Results

6.1. Country scores and rankings in their performance on research excellence

Figures 6.1 and 6.2 present the outcomes of the analysis for the composite indicator on research excellence. First, Figure 6.1 shows the scores of the composite indicator for the two time points ([9]). For both time points, among the most excellent countries in research are Switzerland, the Netherlands, Denmark, Sweden, and Israel. Countries that are ranked in the middle involve both big countries (like Germany, France, the United Kingdom and the United States) and smaller countries (like Belgium and Austria). Lower ranked countries are both emerging economies (like Brazil, India, and China) and Central and Eastern European countries (like Slovakia, Hungary, and Latvia). Overall, this ranking seems to be very much in line with the outcomes of the cluster analysis presented in section 5 of this report.

Figure 6.1 Composite Indicator scores for '2005' and '2010'

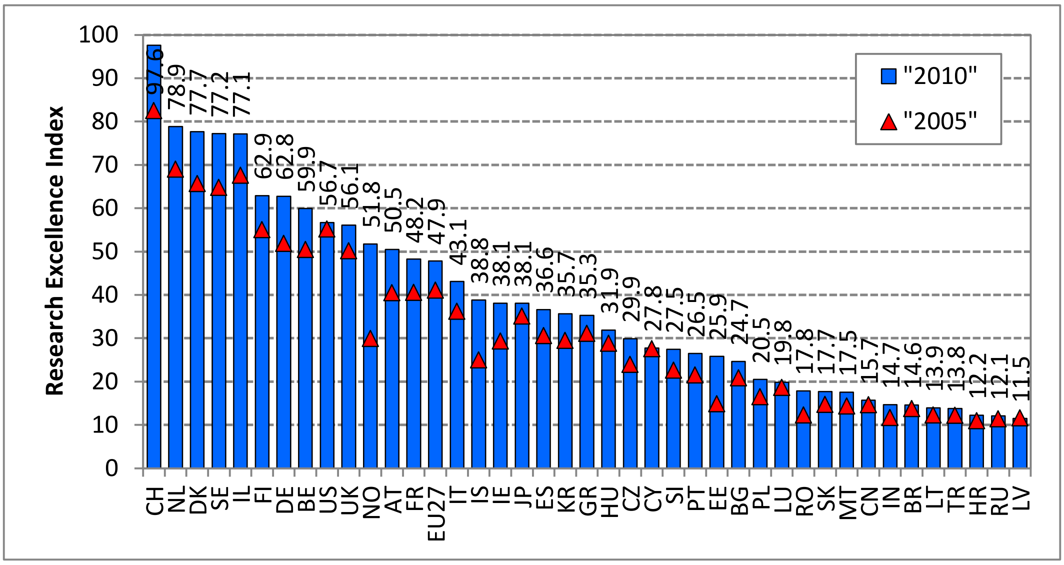
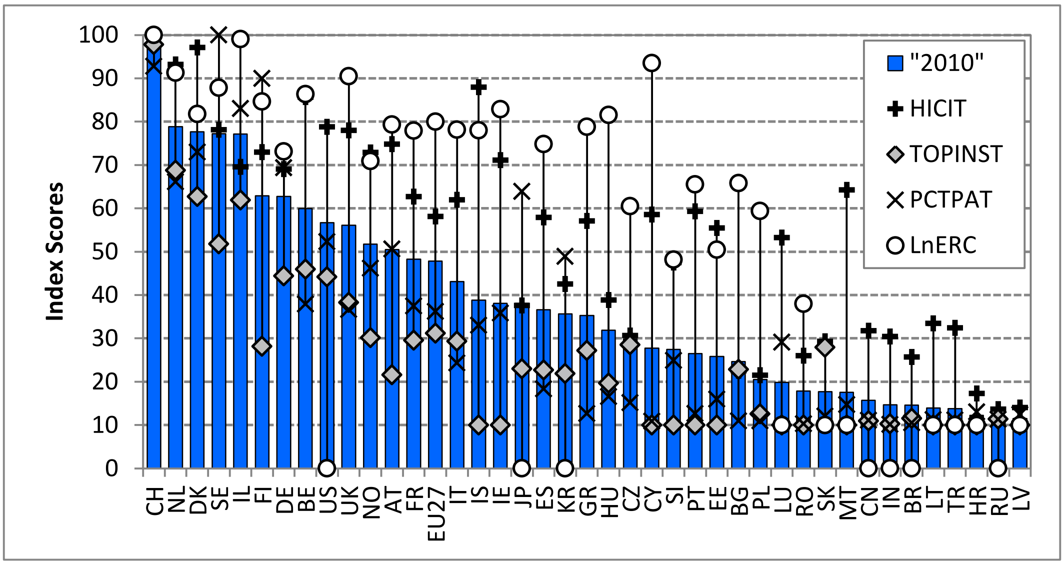


Figure 6.2 compares the overall scores of countries with the scores of these countries for the 4 individual variables. A number of patterns stand out from this figure. First, as was clear from the cluster analyses as well, lower ranked countries perform especially poorly in terms of the number of world class universities and research institutes located there and the number of PCT applications of those countries. Second, comparing only the strong variables (i.e. highly cited publications with PCT patents), we find that while some countries perform especially well in technological research excellence (e.g. Sweden, Israel, Japan, and Finland) and less so in scientific research excellence, for other countries it is the other way around (e.g. the Netherlands, Denmark, the United Kingdom, and the United States perform better in scientific research excellence than in technological research excellence). Finally, note that lower ranked countries perform better in terms of scientific research excellence than they do in technological research excellence.

([9]) Note that we cannot properly assess the growth of research excellence across time because the underlying variables are not on par in terms of years covered. As such Figure 6.1 should not be interpreted in terms of a development in research excellence for each individual country. Rather, this figure should be interpreted as a layered snapshot of two individual (i.e. non-comparable) time points.

Figure 6.2 Composite scores and individual variable scores for '2010'



Visualising the scores in a different way, Figure 6.3 presents the comparative performance of countries in the various dimensions of research excellence. We note the high number of zero values for the world class universities and research institutes variable. While this variable cannot distinguish the performance of countries with a score of zero, this does not mean that this variable cannot be interpreted. In other words, the zeros are meaningful from our definition of research excellence being about top-end quality research outputs in the sense that they convey the message that many countries do not have world class universities and research institutes at all. Subsequent updates of the variable with more recent data will be able to track if these countries improved (see also: Aghion et al., 2010).

Figure 6.3 Comparative performance of of countries

Country		Highly Cited Publications per Total Publications	Top Universities per GERD	PCT Patents per GDP	ERC Grants per public R&D	*Overall Score*
Switzerland	CH	100	98	93	100	*97.6*
Netherlands	NL	93	69	66	91	*78.9*
Denmark	DK	97	63	73	82	*77.7*
Sweden	SE	78	52	100	88	*77.2*
Israel	IL	70	62	83	99	*77.1*
Finland	FI	73	28	90	85	*62.9*
Germany	DE	69	44	69	73	*62.8*
Belgium	BE	86	46	38	86	*59.9*
United States	US	79	44	52	n.a.	*56.7*
United Kingdom	UK	78	38	37	90	*56.1*
Norway	NO	73	30	46	71	*51.8*
Austria	AT	75	22	51	79	*50.5*
France	FR	63	30	37	78	*48.2*
EU-27	EU27	58	31	36	80	*47.9*
Italy	IT	62	29	24	78	*43.1*
Iceland	IS	88	10	33	78	*38.8*
Ireland	IE	71	10	36	83	*38.1*
Japan	JP	38	23	64	n.a.	*38.1*
Spain	ES	58	23	18	75	*36.6*
Rep. of Korea	KR	43	22	49	n.a.	*35.7*
Greece	GR	57	27	13	79	*35.3*
Hungary	HU	39	20	17	82	*31.9*
Czech Republic	CZ	31	29	15	60	*29.9*
Cyprus	CY	59	10	11	93	*27.8*
Slovenia	SI	48	10	25	48	*27.5*
Portugal	PT	59	10	13	65	*26.5*
Estonia	EE	55	10	16	50	*25.9*
Bulgaria	BG	23	23	11	66	*24.7*
Poland	PL	21	13	11	59	*20.5*
Luxembourg	LU	53	10	29	10	*19.8*
Romania	RO	26	10	10	38	*17.8*
Slovakia	SK	29	28	12	10	*17.7*
Malta	MT	64	10	15	10	*17.5*
China	CN	32	11	11	n.a.	*15.7*
India	IN	30	10	10	n.a.	*14.7*
Brazil	BR	26	12	11	n.a.	*14.6*
Lithuania	LT	33	10	11	10	*13.9*
Turkey	TR	32	10	11	10	*13.8*
Croatia	HR	17	10	13	10	*12.2*
Russia	RU	14	11	11	n.a.	*12.1*
Latvia	LV	14	10	12	10	*11.5*
	Median:	*57.9*	*21.8*	*24.3*	*76.3*	*35.3*

Note: Bar lengths indicate country scores for the four research excellence indicators (highly cited publications, Top 250 Universities, PCT patent applications and ERC grants received). The minimum score for each indicator is 10, the maximum is 100.

Non-ERA countries are not assessed on the ERC grants indicator due to home region bias, thus marked by "n.a." Median scores within each aspect and the overall median are shown beneath the bars.

6.2. Robustness assessment of the research excellence index

Monitoring research excellence at the national scale across the European Union Member States and with respect to benchmark countries raises practical challenges related to the quality of data and the combination of these into a single number. This section discusses the assessment of the research excellence index along two main axes: the conceptual and statistical coherence of the structure and the impact of key modelling assumptions on the country ranks (see for example: Saisana et al., 2011, Saltelli et al., 2008).

These are necessary steps to ensure the transparency and reliability of the index, to enable policymakers to derive informed and meaningful conclusions, and to potentially guide choices on priority setting and policy formulation. The conceptual and statistical coherence is carried out using two statistical approaches, principal component analysis and reliability item analysis. The key modelling assumptions tested random weights, and alternative aggregation formulas (geometric versus arithmetic). The analysis complements the country rankings with confidence intervals, in order to better appreciate the robustness of these ranks to the index computation methodology. In addition, we include a discussion on the use of alternative variables for TOPINST, PCTPAT and ERC and alternative measures that could be used as denominators.

6.2.1. Conceptual and statistical coherence in the framework

The options for the research excellence index were assessed in an iterative process that aimed at setting the foundation for a balanced index. The process followed four steps (see Figure 6.4).

Step 1: Conceptual Consistency*.* Candidate variables were selected for their relevance to research excellence, on the basis of a literature review, expert opinion, country coverage, and timeliness (as discussed thoroughly in sections 2, 3, and 4). To represent a fair picture of country differences, all variables were denominated. We discussed and assessed alternative denominators, namely Gross R & D expenditure, GDP and Number of researchers as rescaling measures (not reported here).

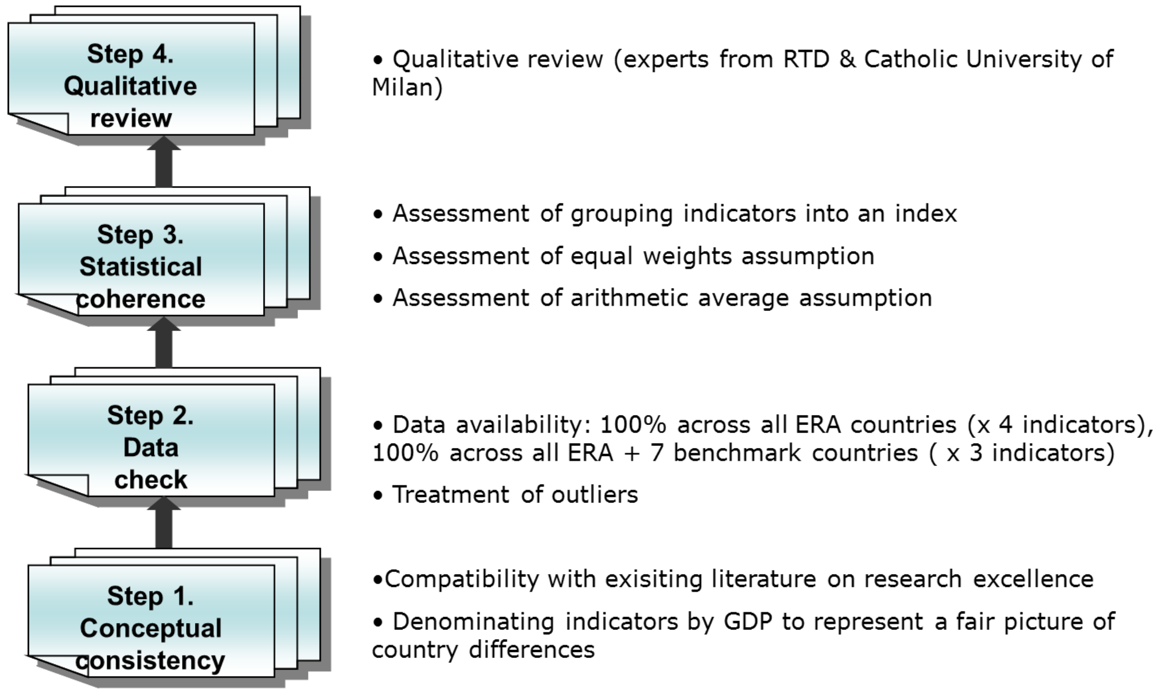
Step 2: Data Checks*.* The data availability in the two years across the four variables for the ERA countries is 100 %. For the remaining benchmark countries, data on ERC grants is absent. The distributions of scores for three of the four variables had skewness and kurtosis levels within the generally acceptable range (absolute skewness<2 and kurtosis<3.5, see: Groeneveld and Meeden, 1984). Instead, the ERC variable had extreme values for several countries. To avoid that the outliers in the distribution of ERC grants introduce a bias in the results of our composite indicator, we decided to log normalise this variable.

Step 3: Statistical Coherence*.* Principal component analysis confirms the presence of a single statistical component that captures 73 % of the variation in the four variables for the ERA countries (see Section 5.3 for more details). A further statistical justification that the selected variables measure the latent phenomenon well is offered by reliability item analysis using the coefficient Cronbach alpha (c-alpha) (Cronbach, 1951).

The c-alpha value is 0.88 for the research excellence index, calculated as the simple average of the four selected variables across the ERA countries only (and across the two time points). This suggests that the index has a high statistical reliability (well above the.7 recommended threshold; see e.g.

Nunnally and Bernstein, 1978). When either HICIT, TOPINST, PCTPAT or ERC is excluded from the framework, the reliability remains mainly unaffected. These results reveal that the choices made in building the research excellence index (choice of indicators, equal weights, geometric average formula) have assured the statistical coherence of the index.

Figure 6.4 Conceptual and statistical coherence in the research excellence framework



Step 4: Qualitative Review. Finally, the country scores and ranks for the research excellence index were evaluated by RTD and an external consultant from the Catholic University of Milan to verify that the overall results were, to a great extent, consistent with current evidence, existing research or prevailing theory. Notwithstanding these statistical tests and the positive outcomes on the statistical coherence of the proposed indicator, it is important to mention that it should remain open for future improvements as new relevant research studies become available. A potential revision of the framework in the upcoming years can thus be envisaged.

6.2.2. Impact of modelling assumptions on the aggregate results

Every country score on the research excellence index depends on modelling choices: variables' selection, normalisation, weights, aggregation method, among other choices. Robustness analysis is aimed at assessing the simultaneous and joint impact of these modelling choices on the rankings. The data are error-free since eventual errors and typos were corrected during the computation phase.

To explore the robustness of the research excellence composite indicator, we conduct an uncertainty and sensitivity analysis. In an uncertainty analysis we analyse the predicted imprecision of country rankings that is due to the overall uncertainty in modelling assumptions. Sensitivity analysis is then used to quantify how changes in particular modelling assumptions alter the value of the country rankings. Both methods are complementary and hence are often used together in assessing the robustness of a composite. The robustness assessment of the research excellence composite was based on a combination of a Monte Carlo experiment and a multi-modelling approach that dealt with two issues: (1) change of weights for all variables and (2) various aggregation methods. This type of assessment aims to anticipate eventual criticism that the aggregate scores were calculated under conditions of certainty (Saisana et al., 2005, Saisana et al., 2011).

The Monte Carlo simulation was played on the weights for the selected variables, and comprised 1 000 runs, each corresponding to a different set of weights, randomly sampled from uniform continuous distributions in the range 15 %-35 %. The sampled weights were then rescaled to unity sum. This choice of the range for the weights' variation ensures a wide enough interval to have meaningful robustness checks, whereby we consider a 10 percent weight below or above equal weights of 25 %.

The next modelling assumption considered relates to the use of the geometric average in the calculation of the index, a formula that received statistical support from principal component analysis and reliability item analysis. Decision-theory practitioners have challenged the use of simple arithmetic averages because of their fully compensatory nature, in which a comparative high advantage on a few indicators can compensate a comparative disadvantage on many indicators (Munda, 2008). Hence we chose to use the geometric approach. The geometric average ([10]) is a partially compensatory approach that rewards countries with similar performance on the underlying variables and motivates them to improve in the variables in which they perform poorly, and not just in *any* variable. Note however that apart from these statistical properties, there might in principle be good conceptual reasons to go for the arithmetic approach for measuring research excellence.

We only perform the robustness analyses on countries that have observations for the four variables. As such, this analysis only includes ERA countries. Four models for the ERA country comparison were tested based on the combination arithmetic versus geometric average of the selected variables and the years of observation of the index (year 2005 and 2010). Combined with 1 000 simulations per model for the random weights assigned to the variables, a total of 4 000 simulations for the ERA comparison were carried out (see Table 6.1 for a summary of the uncertainties considered during this testing phase of the index).

Table 6.1 Uncertainty analysis for the research excellence index

1. Uncertainty in the weights for the variables		
Reference		***Alternative***
HICIT	25 %	[15 %-35 %]
TOPINST	25 %	[15 %-35 %]
PCTPAT	25 %	[15 %-35 %]
ERC	25 %	[15 %-35 %]
2. Uncertainty in the aggregation formula		
Reference		***Alternative***
geometric average		arithmetic average

([10]) In the geometric average, indicators are multiplied as opposed to summed in the arithmetic average. Indicator weights appear as exponents in the multiplication.

Uncertainty analysis results. The main results of the uncertainty analysis, accounting for the two issues on the random weights for the selected variables and the aggregation formula are summarised in Table 6.2 for the comparison of ERA countries, which reports the country ranks of the Research Excellence Index '2005' and the Research Excellence index '2010' and the respective 90 % confidence intervals. We observe that all the country ranks of the research excellence indicator lay within the ranking intervals simulated with the above-mentioned Monte Carlo method. While the uncertainty intervals are relatively stable for the top twelve countries, the range in ranks seems to be much more volatile for the remaining countries. The most volatile countries in the ERA comparison appear to be Cyprus and Iceland with interval ranges of 14 positions in the Research Excellence index '2005' and respectively 13 and 9 in the Research Excellence index '2010'. Also Czech Republic, Luxembourg and Norway seem to be fairly volatile. We can conclude that for the top countries, country ranks on research excellence are relatively robust to the particular methodological choices made, while for the remaining countries more caution should be taken in interpreting their scores and rankings on research excellence.

Table 6.2 ERA country ranks and 90 % intervals for the composite indicators on research excellence

	RE Index "2005"		RE Index "2010"	
	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	1	[1, 1]
Netherlands	2	[3, 3]	2	[2, 4]
Israel	3	[2, 2]	5	[2, 3]
Denmark	4	[4, 5]	3	[4, 5]
Sweden	5	[4, 5]	4	[3, 5]
Finland	6	[6, 8]	6	[6, 8]
Germany	7	[7, 10]	7	[7, 9]
Belgium	8	[6, 9]	8	[6, 8]
United Kingdom	9	[7, 9]	9	[7, 9]
EU-27	10	[10, 14]	13	[10, 16]
France	11	[11, 15]	12	[12, 15]
Austria	12	[11, 13]	11	[10, 13]
Italy	13	[12, 18]	14	[14, 17]
Greece	14	[14, 18]	18	[16, 19]
Spain	15	[15, 20]	17	[15, 19]
Norway	16	[11, 19]	10	[10, 13]
Ireland	17	[17, 20]	16	[13, 19]
Hungary	18	[14, 20]	19	[16, 23]
Cyprus	19	[7, 21]	21	[12, 25]
Iceland	20	[10, 24]	15	[11, 20]
Czech Republic	21	[18, 25]	20	[18, 26]
Slovenia	22	[20, 24]	22	[21, 26]
Portugal	23	[22, 26]	23	[21, 26]
Bulgaria	24	[21, 27]	25	[21, 28]
Luxembourg	25	[21, 27]	27	[20, 28]
Poland	26	[26, 30]	26	[26, 31]
Estonia	27	[24, 28]	24	[23, 27]
Slovakia	28	[24, 29]	29	[22, 28]
Malta	29	[26, 29]	30	[20, 29]
Romania	30	[29, 34]	28	[30, 33]
Lithuania	31	[30, 32]	31	[29, 31]
Turkey	32	[30, 34]	32	[30, 32]
Latvia	33	[30, 33]	34	[34, 34]
Croatia	34	[31, 34]	33	[32, 33]

Sensitivity analysis results. Sensitivity analysis (Saltelli et al., 2008) has been used to identify which of the modelling assumptions have the highest impact on country ranks, and thereafter to help focus the discussion on those uncertainties. Figure 6.5 presents the box plots of ranking shifts for the two assumptions tested for the ERA comparison based on 4 variables. The median shift in rank across all simulations is the red segment. The vertical boxes show the 75 % of the distributions (percentiles P25 and P75 are the horizontal edges of the boxes) and vertical lines extend from minimum to maximum values.

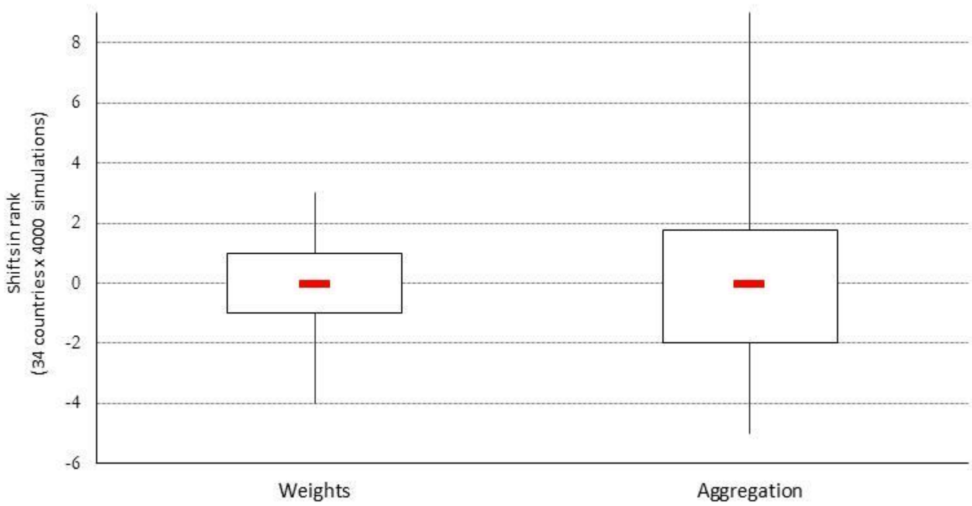
The two assumptions are influential in the ERA country comparison based on 4 variables, although this influence remains quite small for the majority of countries. The impact of the aggregation method seems to be higher compared to the variance created by random weights ([11]). If geometric averages were used instead of arithmetic ones, more than seventy percent of the countries move

([11]) Note that this last is an average shift in rank across all other uncertainties.

less than 3 positions in either year. For the other thirty percent, the most notable impact would be for Cyprus gaining 8 positions on the Research Excellence index '2005' and Iceland gaining 10 on the Research Excellence Index '2005', while Bulgaria and Czech Republic both lose 6 positions in the Research Excellence Index '2010'. The influence of the random weights is smaller. Almost forty percent of the countries are left unaffected due to changes in random weights, while the remaining countries face a moderate impact of 4 positions shift in the most extreme cases.

The results of sensitivity analysis confirm how important it is to take into account the influence of the different choices of the model specifications. In line with this sensitivity analysis it is important to open the discussion towards the choice of alternative variables used to develop a composite indicator on research excellence and the issues of the denominators used to scale variables.

Figure 6.5 Sensitivity analysis: Impact of assumptions on the composite Indicator ranks for the ERA country comparison on 4 variables



In principle we can use a number of variables to denominate our four variables on research excellence. As such, the potential list of denominators could be, among others: e.g. Gross Domestic Product (GDP), Population, Total R & D expenditure (GERD), number of researchers, total number of publications in a country, total number of patents in a country. In addition, researcher may prefer to use a single variable as denominator (e.g. scaling all the variables with GDP) rather than using a different denominator as we choose in this report. In the same vein, the numerators used to construct the composite indicator may be subject to various choices. In this regard, triadic patent families could be used as alternative for PCT data, while the TOPINST variable could have been constructed based on the Shanghai ranking instead of the SciMago ranking. We refer to Sections 4.3 and 4.4 in which we discussed the potential of each alternative in more detail. The choices of numerators and denominators seem to play an important role in the construction of composite indicators as we observed in various initial tests that we performed the research excellence index. Consequently it should receive particular attention on the future research agenda and in upcoming revisions of this framework. In the following uncertainty analysis however, we take for granted the structure of the composite indicator composed of HICIT, TOPINST, PCTPAT and ERC.

6.3. The link between research excellence and related phenomena

6.3.1. Assessing research excellence vis-à-vis innovation and competitiveness

This section discusses the validity of the composite indicators on research excellence by comparing them with composite indicators that capture innovation and competitiveness. The rationale underlying this validation builds upon the literature that research (excellence) is closely related to phenomena such as competitiveness and innovation (see also sections 2 and 3 of this report). For this purpose we will analyse scatter plots between the research excellence composite and these indicators and, given the alleged relationship between research excellence and these other phenomena, we expect positive correlations. However, we do not expect that the correlation will be too high as the competitiveness and the innovation performance of countries are influenced by many other factors. While too high correlation would indicate that composites of research excellence capture the same phenomenon as composites on competitiveness and innovation; a too low correlation would cast doubts about the validity of the composites on research excellence altogether. In addition, we do not suggest any causality between research excellence and these indicators as correlation analyses do not allow for the distinction of a cause-and-effect relationship.

First we chose to compare the results of the composite indicator of research excellence with mostly widely used indicators of innovativeness and competeviveness at the country level: the Global Innovation Index (GII), the Summary Innovation Index (SII), and the Global Competitiveness Index (GCI). As has been put forward by many scholars in the innovation literature (Freeman, 1995, Lundvall et al., 2002, Zucker et al., 2002, Fleming and Sorenson, 2004), both scientific and technological research have led to new innovations on the one hand, while these innovations have in turn fostered new lines of (excellent) research. Given that research and innovation activities are closely intertwined, we expect the indicator of research excellence to be correlated with these two innovation indices, without digging into the causal relationship among these phenomena.

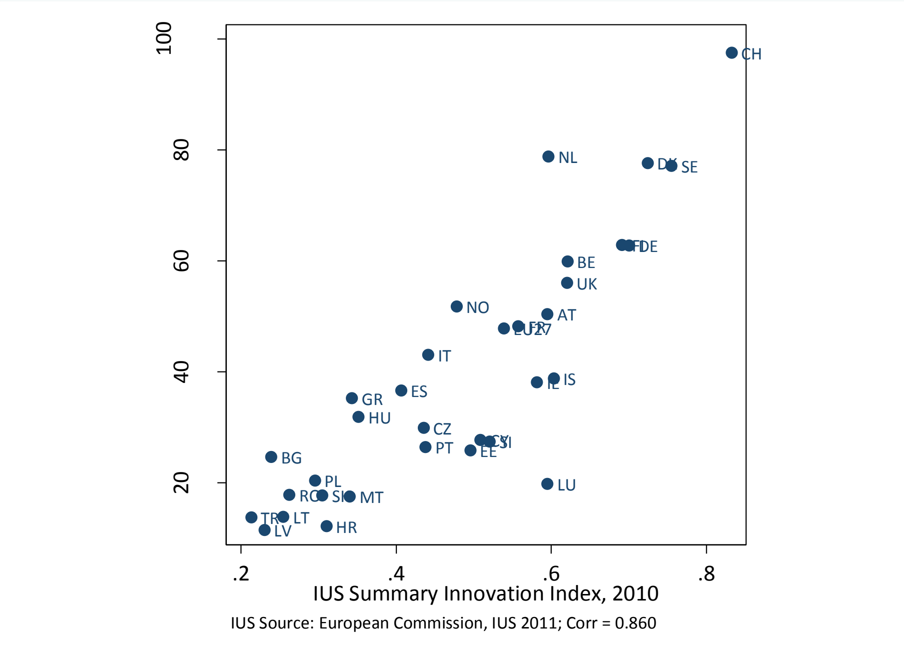
The summary innovation index (SII) is a composite indicator that measures the innovative performance of EU27 countries and its international competitors. It is based on a weighted average of three main pillars of innovation indicators: enablers (indicators capturing the main drivers of innovation such as financial support, openness of research systems and the availability of human capital), firm activities (indicators on entrepreneurship, firm investments and intellectual assets), outputs (indicators on innovators and economic effects of innovation). We refer to the Innovation Union Scoreboard report of 2011 for a more extensive explanation on the construction of this indicator (Commission, 2011). ([12])

The correlation between the research excellence composite indicator of '2010' and the SII index of 2011 is presented in Figure 6.6. As expected, the SII and research excellence indicator are positively and highly correlated (Corr. =.86). This result emphasise the fact that although countries' innovative

([12]) We used the 2011 scoreboard's '2010' figures as they are the closest in time (mostly referring to data of 2008 and 2009) to the indicators of our composite indicator of research excellence. The latest edition of the Innovation Union Scoreboard was released in March 2013, at the time of writing this report, and shows very similar correlation patterns.

performance and research excellence might be related to each other, both phenomena are influenced by other factors as well, which is potentially linked to entrepreneurship, industrial structure and other systemic innovation features.

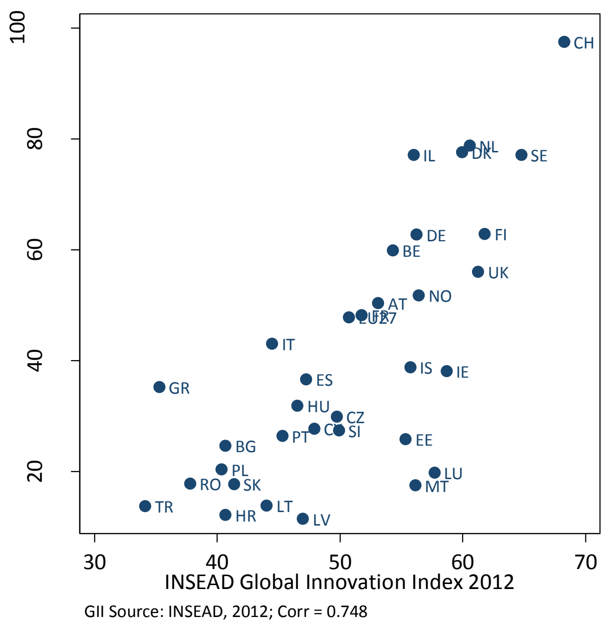
Figure 6.6 Research excellence ('2010') versus Summary Innovation Index (2010)



In the same line as the Summary Innovation Index, the Global Innovation Index (GII 2012) published by INSEAD in collaboration with (among others) WIPO, measures the innovative performance of countries by including both input and output measures of innovation (Dutta, 2012). Besides a broader scope of variables on human capital and financial support for innovation, this index also includes two additional pillars on the input side of innovation: (i) an institutional pillar (taking into account variables on the political, business and regulatory environment that can foster or hinder the ease of innovation) and (ii) an infrastructure pillar (reflecting on the availability of ICT or other general infrastructures such as transport and electricity). Also from the output side, this index includes more variables that capture the outcomes of innovative efforts (e.g. variables on knowledge diffusion). In addition, the scope of countries that are covered by this index is much broader as the index is also calculated for a sample of Latin-American and African countries ([13]). The correlation of the Global Innovation Index 2012 with the research excellence indicator of '2010' is presented in Figure 6.7. As with the previous index, we note a positive correlation.

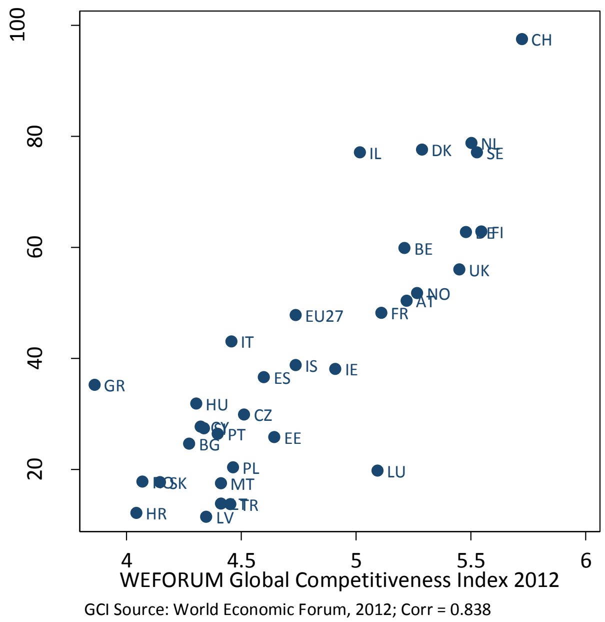
([13]) More information on the methodology of the GII can be found at www.globalinnovationindex.org.

Figure 6.7 Research excellence ('2010') versus Global Innovation Index (2012)



Finally, the Global Competitiveness Index (GCI 2012) intends to capture micro- and macroeconomic foundations of national competitiveness. Competitiveness can be defined as those amenities of a country that contribute to a country's productivity (Boschma, 2004, Krugman, 1996). This index is a weighted average of many different components that measure various aspects of competitiveness. Without going into detail, we give a brief overview of how this index is constructed (Sala-i-Martín et al., 2012). Roughly speaking, the index includes twelve pillars that are grouped in three main building blocks, each focusing on a different aspect of competitiveness. A first building block concerns the basic requirements a country needs to be competitive. It includes pillars (and underlying variables) on a country's infrastructure, institutions, macroeconomic environment, health and primary education. A second block focuses on variables that enhance the efficiency of a country, containing pillars on e.g. good market efficiency, labour market efficiency, technical readiness of a market and financial market development. A third and last block includes pillars on the innovative performance and the business sophistication of a country. Following the rationale of section 2, research excellence can be said to be related to a country's (sustained) competitiveness, although other factors may play a role as well. We explore this relationship by plotting the research excellence of '2010' on the competitiveness index of 2012 in Figure 6.8. Although the plot reveals a positive correlation between the two phenomena, many countries seem to differ quite extensively in terms of competiveness while having a similar score on research excellence (e.g. Greece versus Luxembourg, Israel versus Finland) and vice versa.

Figure 6.8 Research excellence ('2010') versus Global Competitiveness Index (2012)



Overall, given the distinct positive correlations found between research excellence and the various indices on innovation and competitiveness, we can conclude that the composite indicator on research excellence is valid. The correlation analysis shows that our Research Excellence Index is akin to but measuring a somewhat different phenomenon than the innovativeness and competitiveness of a country.

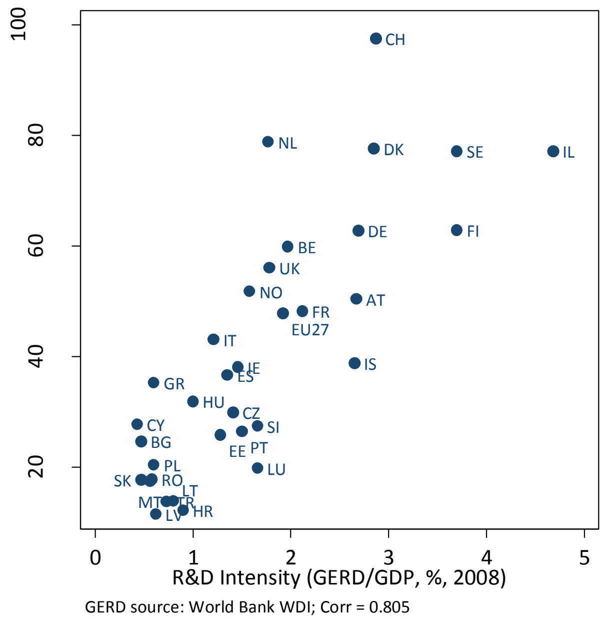
6.3.2. Assessing research excellence in the broader context of national research systems

Research excellence and research assets

In the following sub-sections we relate the research excellence composite to variables that are associated with the first building block of national research systems which we defined earlier as research assets. Here, we plot research excellence against financial research assets first.

Figure 6.9 plots the research excellence composite in '2010' against total R & D expenditures (GERD) per GDP in 2008. Most of the countries reporting a low level on GERD/GDP are also low performers in terms of research excellence. However, when looking at countries with higher actual levels of research excellence, they seem to be more scattered in their levels of GERD/GDP. Sometimes, countries with relatively high levels of GERD/GDP report the same level of research excellence as countries with much lower level of GERD/GDP (compare e.g. Austira with Norway or Germany with Finland). Alternatively, countries with similar levels of GERD/GDP report vastly different levels of research excellence (compare e.g. Croatia with Hungary or Slovenia with the United Kingdom).

Figure 6.9 Research excellence versus Gross R & D expenditure



Similar patterns can be discerned for the association between research excellence and private R & D expenditures (BERD/GDP), plotted in Figure 6.10. Concerning the plots of research excellence against public R & D expenditures (Figure 6.11), we note again a positive correlation. When comparing the vertical axis of this plot with the previous ones, the differences in actual levels in public R & D/GDP seem to be much lower across countries than for BERD/GDP or GERD/GDP.

Figure 6.10 Research excellence versus Business R & D expenditure

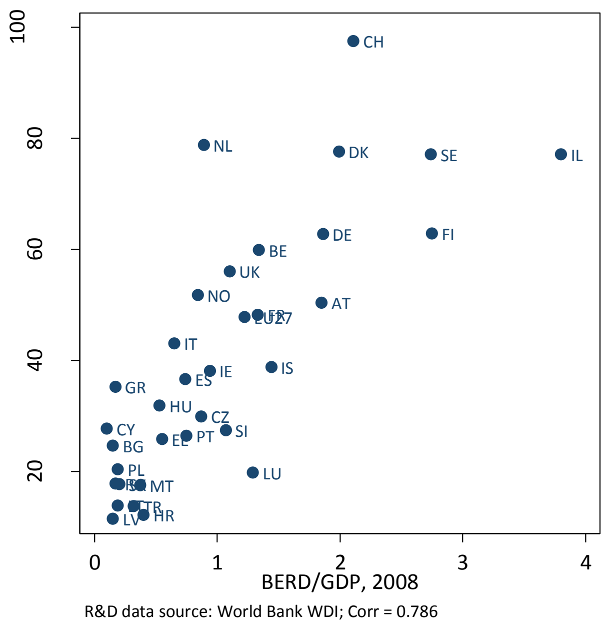
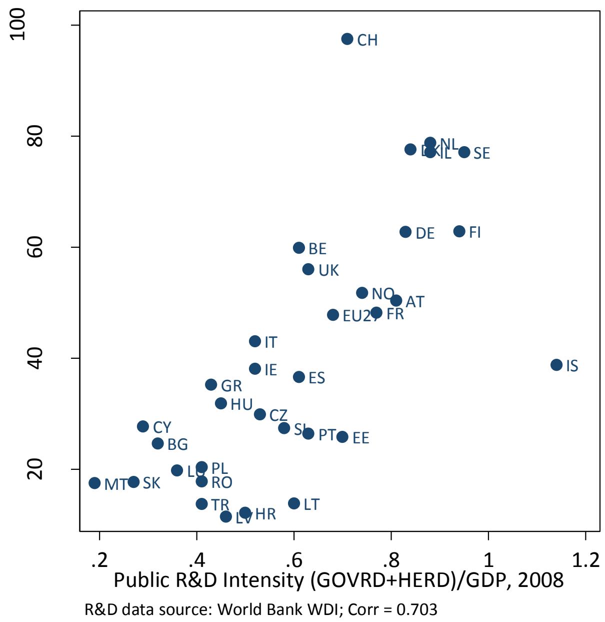
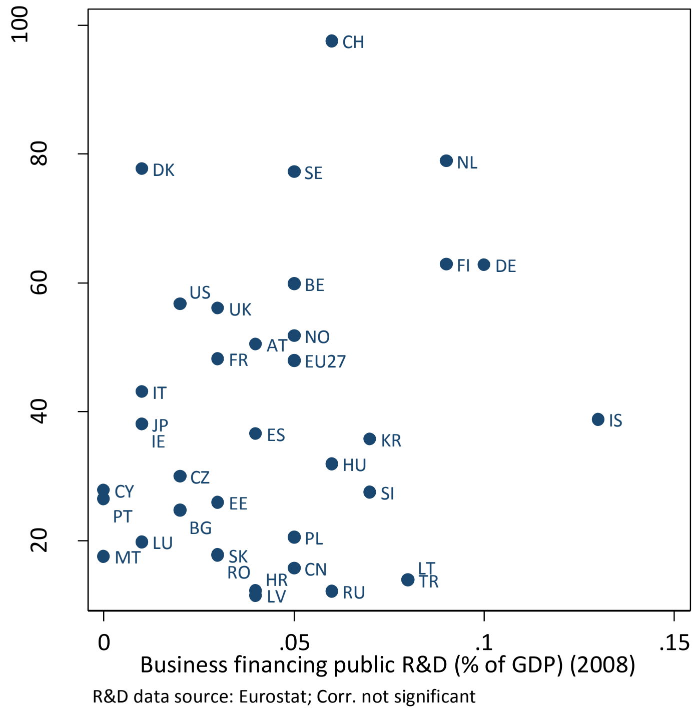


Figure 6.11 Research excellence versus Public R & D expenditure



Employing a bar plot, Figure 6.13 provides an alternative way to compare the actual levels of the research excellence composite indicator of '2010' with the respective levels of public R & D per GDP and business R & D per GDP for 2008. A number of patterns stem out from this figure. First, for most of the countries that perform relatively poor in terms of research excellence, the lion's share of R & D spending stems from public R & D. In sharp contrast with the previous pattern, the lion's share of R & D expenditure per GDP shifts from public to private in countries showing medium to high performances in research excellence. Nevertheless, many leaders in research excellence are also the leaders in public R & D spending per GDP (e.g. Sweden, Finland, Israel, and the Netherlands). Also, although some countries report high R & D investments (both in terms of public and private spending) they do not correspond to high research excellence (e.g. Japan and Korea). However, when we relate research excellence to public R & D expenditures financed by business we do not observe a significant positive correlation; suggesting that research excellence is not associated with this particular funding source (Figure 6.12).

Figure 6.12 Research excellence ('2010') versus public R & D expenditure financed by business



Alternatively, R & D expenditure can be broken down by basic and applied research and experimental development, as done in Figure 6.14. We find the highest correlation between excellence scores and the percentage of GDP spent on basic research in 2008 (0.73), a somewhat lower coefficient for experiment development (0.64) and substantially lower coefficients for applied research (0.50). Apart from the UK and Slovenia – which seem to compensate relatively low basic research spending with relatively high expenditures on applied research – few countries combine high research excellence with low expenditures in basic research. It is interesting to see that the disproportionately high expenditure on experimental development in Israel is coupled with very low basic and applied research spending but high excellence scores – a result potentially driven by unmeasured defense spending. We also note Japan and South Korea, two among the three largest spenders on experimental development are relatively weaker performers in terms of research excellence ([14]).

([14]) Unfortunately, data breakdown was missing for a number of EU countries, therefore an EU27 – US comparison could not be provided.

Figure 6.13 Research excellence ('2010') versus sources of R & D expenditure (2008)

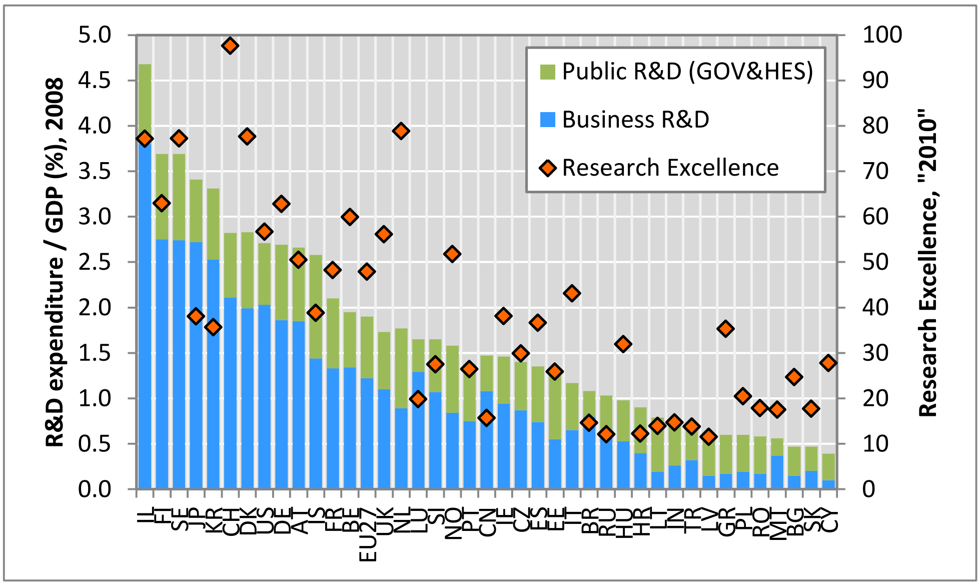
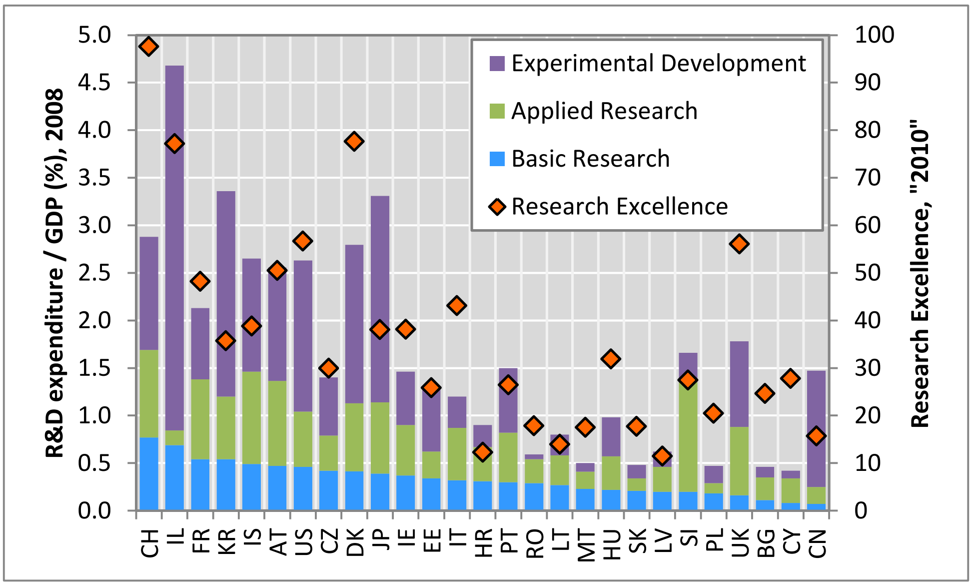
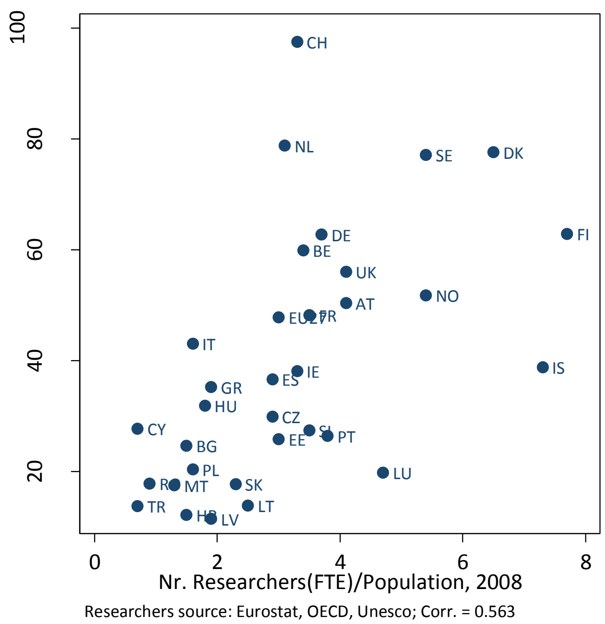


Figure 6.14 Research excellence ('2010') versus Gross R & D expenditure broken down by basic, applied research and experimental development (2008)



Besides financial resources in R & D we also relate the composite indicator of research excellence with the amount of **human research assets** as defined by the number of researchers per thousand inhabitants. Comparable to the correlation plots with financial resources, countries with a lower number of are associated with lower levels of research excellence. However, countries with medium to high numbers of researchers, have much more dispersed scores on research excellence. While relatively high performing countries in research excellence such as Sweden, Finland and Denmark belong among the countries with the highest number of researchers, other top performers (e.g. Switzerland, the Netherlands) show much lower levels of researchers. These findings would suggest a more efficient use of human resources in the latter class of countries.

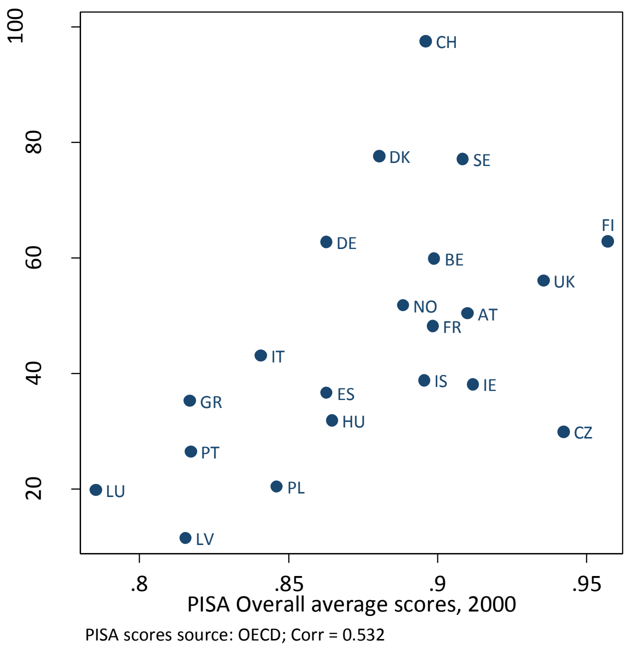
Figure 6.15 Research excellence ('2010') versus number of researchers per thousand inhabitants (2008)



Finally, we relate the composite of research excellence in 2008 with the level of skills and competencies of the population that can be measured by scores of the Programme for International Student Assessement (PISA) ([15]). We plot the research excellence composite of '2010' against the PISA scores of 2000 in figure 6.17. Although the figure reveals a positive relationship between the PISA scores and research excellence, this relation does not hold strongly for the top performing countries in terms of average PISA scores. Countries with the highest PISA scores do not only report high performances on research excellence (Sweden and Finland), but also show medium to low performances (Ireland and Czech Republic). More consistency can be found for the countries with low levels on their PISA scores. These countries are generally associated with low values on research excellence (with the most notable case of Brazil).

([15]) PISA is a worldwide study conducted by the OECD that aims to evaluate 15-year-old school pupils' competencies in three main subjects: reading, mathematics and science.[15] This study was launched in 1997, the first comprehensive data collection carried out in 2000, and is repeated every three years. To date more than 70 countries have been participating worldwide. The PISA scores are weighted averages of the results on the tests in these three competence fields. The tests are identical across countries and aim to analyze the concept of literacy, 'which is concerned with the capacity of students to apply knowledge and skills and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations' (see OECD 2004. *PISA Learning for Tomorrow's World: First Results from PISA 2003*, OECD Publishing.). Literacy should not be understand as the historical notion of the ability to read or write but should be interpreted in a broad sense as the students competencies acquired through life and education. As such, the aim is not to evaluate subject-specific knowledge, but rather to examine competencies across disciplinary boundaries and applied to real life situations. The PISA scores have been widely used to monitor and to assess the education quality and to improve education policies. In a sense, they provide robust measures of students' educational competencies that are internationally comparable, and therefore they can be seen as proxies for countries' future research capabilities (disregarding student and researcher mobility). Hence, excellence in PISA may be related to a country's future research excellence.

Figure 6.16 Research excellence ('2010') versus PISA average scores (2000)



Research excellence and structural capabilities

In this sub-section we relate the research excellence composite indicator with variables that measure countries' structural capabilities. For this purpose we plot the research excellence composite against respectively the share of knowledge intensive activities (KIA) in a country's economy and the share of business R & D investments from abroad.

The research excellence composite is plotted against the share of knowledge intensive activities (KIA) as a fraction of a country's GDP in Figure 6.17. ([16]) The figure shows a rather scattered picture and does not allow us to draw any conclusions on the association between the two analysed factors. In sharp contrast with the previous figure, a clear positive association can be observed when analyzing a scatter plot on the level of research excellence in '2010' with the percentage of business R & D investments from abroad in 2008 (Figure 6.18). Overall, countries with a low (high) percentage of business R & D that is invested from abroad, report low (high) performances on research excellence. Note however, that this plot does not contain all the countries due to data availability.

([16]) We follow the classification method of Eurostat to define activities as knowledge intensive. An activity is classified as knowledge intensive if tertiary educated persons employed (according ISCED97, levels 5+6) represent more than 33 % of the total employment in that activity. The definition is built based on average number of employed persons aged 25-64 at aggregated EU27 level in 2008 and 2009 according to NACE Rev. 2 at 2-digit, using EU Labour Force Survey data.

Figure 6.17 Research excellence ('2010') versus percentage of knowledge intensive activities per GDP (2008)

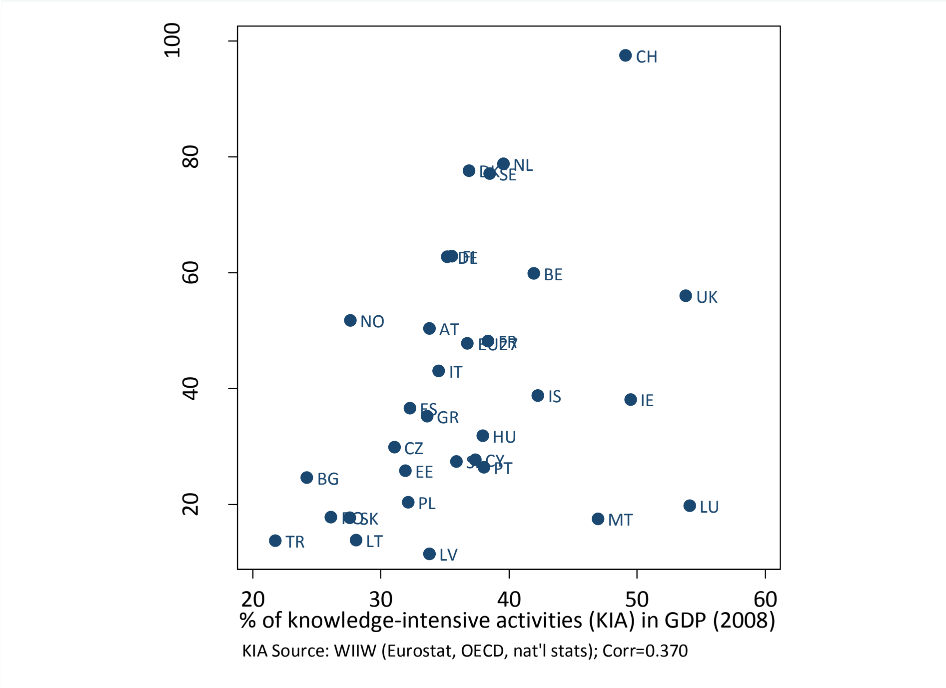
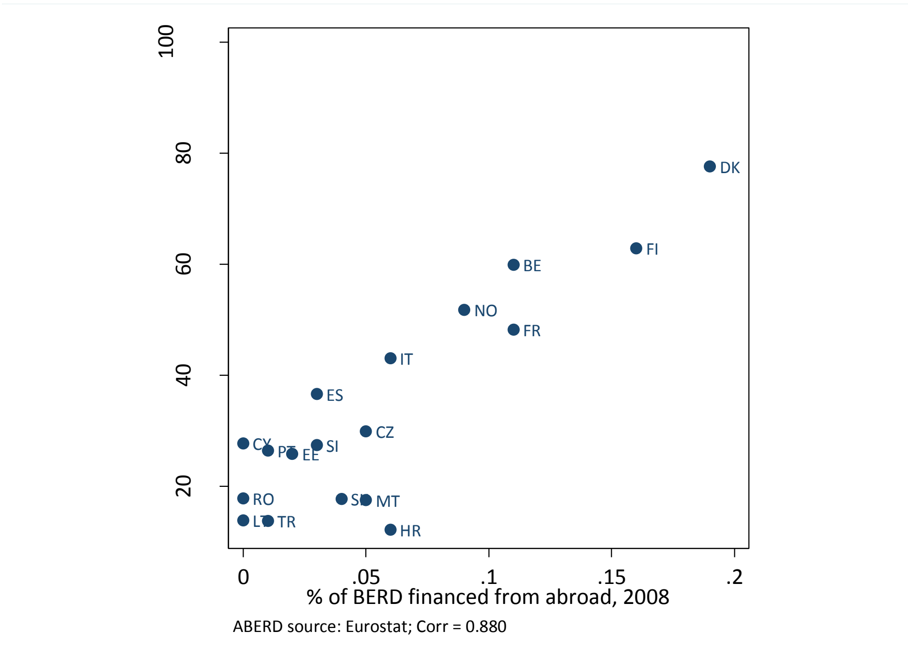


Figure 6.18 Research excellence ('2010') versus percentage of Business R & D expenditure financed from abroad (2008)

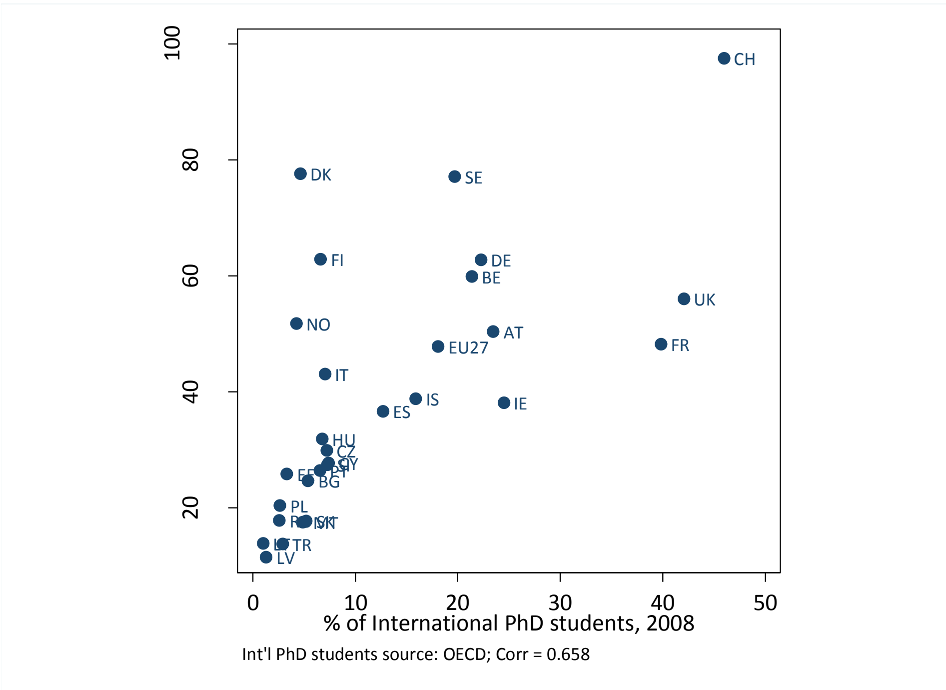


Research excellence and research interactions

In order to relate the composite indicator of research excellence with the third building block of national research systems related to interactions among researchers, we include plots with respectively the mobility of doctoral students, the share of co-publications within the European Research Area (ERA), and the share of co-publications with international (inside or outside ERA) partners.

In Figure 6.19 we plot the research excellence composite against the percentage of international doctoral students in levels for 2008. The plot shows a positive correlation among the two indicators (although it might be largely driven by Switzerland's performance in both). Countries with lower (higher) percentages of international doctoral students in 2008 are associated with lower (higher) levels of research excellence in the same year. However, some countries do not follow this positive trend. First, most Scandinavian countries report low (Denmark, Finland, Norway) to medium (Sweden) percentages in international doctoral students, while their score on research excellence is relatively high. Second, although UK and France belong to the countries with the highest percentages of international students, their research excellence score is relatively low.

Figure 6.19 Research excellence ('2010') versus percentage of international PhD students (2008)



We further explore the interactions among researchers by looking at the share of co-publications with ERA partners. The share of co-publications with ERA partners is defined as the number of co-publications with at least two ERA countries divided by the total number of co-publications, using fractional counting of co-publications. In order to analyse the relationship with the research excellence performance of countries we plot both indicators against each other in Figure 6.20. The plot does not reveal a clear pattern: countries with highest shares of co-publications with ERA partners report high (Switzerland), medium (Iceland) and even low (Luxembourg) performances on research excellence.

Figure 6.20 Research excellence ('2010') versus % of co-publications with ERA partners (2008)

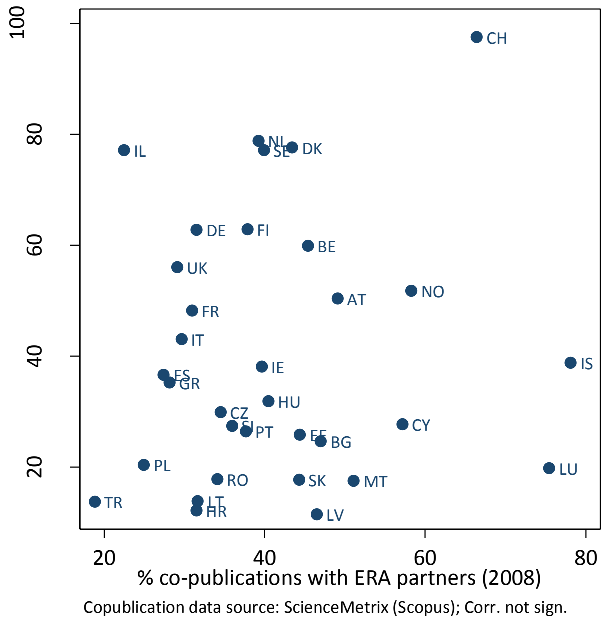
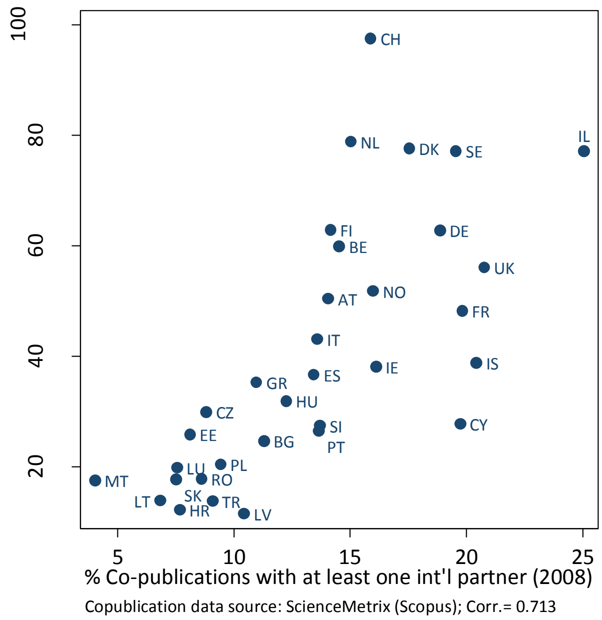


Figure 6.21 Research excellence ('2010') versus % of co-publications with international partners (2008)

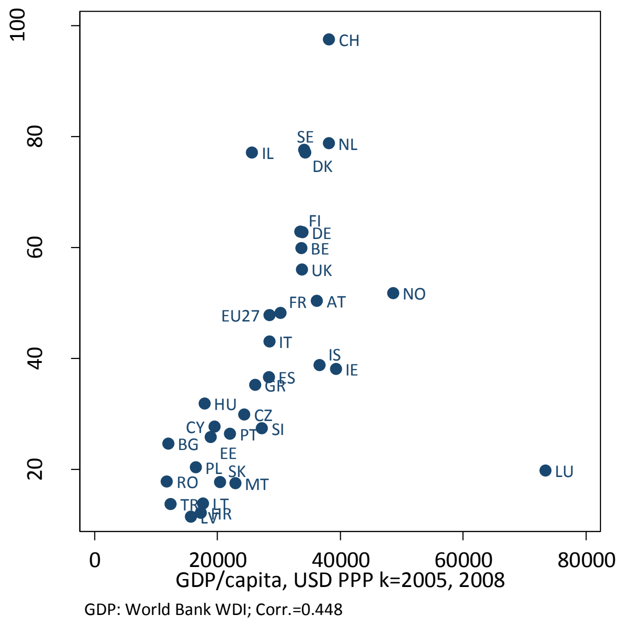


Besides plots with co-publications with ERA partners, we also compare the research excellence indicator with a more general variable: the share of co-publications with international partners (Figure 6.21). This share is defined as the number of co-publications listing at least one ERA country and one non-ERA country divided by the number of co-publications (fractional counting). Compared to the previous plot, the pattern in Figure 6.21 is more distinct for the low performing countries. All countries with poor performance on research excellence report lower scores on the share of co-publications with international partners. Turning to medium and high performers on research excellence, the pattern becomes more scattered again. Countries with seemingly comparable values for one factor may diverge substantially on the other (e.g. Switzerland versus Ireland).

6.3.3. Assessing research excellence vis-à-vis the size of a country's economy

In Figure 6.22 the research excellence composite is related to GDP/capita and reveals a positive correlation: countries with a lower (higher) score on GDP/capita generally report a lower (higher) performance in research excellence. However, a few countries show relatively low levels of research excellence while belonging to the countries with the highest GDP/capita (Norway and Luxembourg). In addition, some countries with similar levels of GDP/capita seem to diverge extensively in terms of their research excellence (e.g. Switzerland and the Netherlands).

Figure 6.22 Research excellence ('2010') versus GDP per capita



7. Conclusion

7.1. Summary

It is widely acknowledged that many European countries are outperformed by countries like the United States when it comes to technological and scientific research (Pavitt, 2000, Dosi et al., 2006). To remedy this situation, the European Commission aims at stimulating research excellence by increasing competition among researchers at a European level. The results reported in this report follow from a project initiated by the Directorate-General for Research and Innovation of the European Commission (DG RTD) within the context of developing indicators for the Innovation Union. The main objective of the overall project is to develop indicators that are capable of measuring and monitoring patterns and trends in research across countries. As such, this report assesses the performance of countries in terms of their record in producing state-of-the-art scientific and technological outcomes; that is, research excellence.

The main contribution of this report lays in a proposal for measuring research excellence at the country level using a composite indicator approach that is related to but distinct from other (composite) indicators available in the field of science and technology assessment. To accomplish this, we followed ten steps (see also: Nardo et al., 2005). First, we developed a conceptual framework that provides the basis for the selection of variables to include in a composite indicator on research excellence. To this aim we positioned the notion of research excellence within the larger context of national research systems. Research excellence is then defined in terms of those aspects of systematically performed creative activities that are about the production of new knowledge whose products are characterised by their high-end quality which can be expressed along various dimensions (i.e. scientific, technological, cultural, etc.) and embodied in both people and artifacts.

Second, following a threefold structure of extent, reliability, and validity, we set out our rationale for choosing particular variables measuring research excellence. Overall, the data needs to conform to at least two requirements. One is that they have to reflect on excellence as defined by high-end quality outputs of research. Another is that we do not consider research input data which represent inputs to research for which we can measure the outputs. In other words, if we make use of input data these inputs need to cover aspects of research that are not already covered by the output data that are available to us. Following this rationale, we measure research excellence using four variables: the number of highly cited publications published by a country, the number of high-quality (i.e. PCT) patents on which a country is listed, the number of world class universities and research institutes in a country, and the number of ERC grants received by a country. While highly cited publications and high-quality patents reflect on new knowledge from respectively science and technology that is attributable to a country and inscribed in texts, the number of world class universities and research institutes and ERC grants are measures of new knowledge from science that is embodied in (groups of) people in a country.

In a third step we assessed the quality profile of the four variables thus selected. Despite the fact that the number of ERC grants received by a country only applies to a limited set of countries (i.e. ERA countries), we choose to construct only one composite indicator measuring research excellence at the country level that includes all 4 variables. Descriptive statistics show that the distributions of the underlying variables are extremely skewed. In general: bigger countries produce more research excellence. Given that countries differ in size, we denominated all 4 variables as to make them scale-independent.

Fourth, given that the number of ERC grants received per public R & D expenditure shows skewness and kurtosis levels that are not within the generally acceptable range, we decided to log-transform this variable. After log-transformation, both skewness and kurtosis of the ERC variable drop to reach generally accepted levels.

Fifth, given that composite indicators cannot be constructed based on anti-correlated or uncorrelated variables, we applied several multivariate statistical techniques to check for statistical coherence among the four variables measuring research excellence. Correlation analysis showed that correlations among the four variables are neither too low (correlation >>.5) nor too high (correlation >>.9) While the former is important as to make sure that the variables are related at all, the latter is important as to make sure that none of the variables is redundant.

In addition, we performed cluster analysis to analyse which countries can be seen as peers in terms of research excellence based on the different variables that populate the composite indicator. The results of the cluster analysis shows that – in general – all countries are grouped following a positive trend in all three variables simultaneously; that is, none of the clusters shows underperformance on one variable vis-à-vis another cluster while over-performing that cluster with respect to another variable.

Principal component analysis (PCA) was carried out in order to detect the number of latent phenomena described by the four research excellence variables. PCA results statistically confirm that there is a single latent dimension within the overall index. Hence, the variables express different aspects of the same phenomenon which supports the aggregation of the four variables into one composite indicator. In a sixth step, given the high correlation among the four variables and the rather balanced component coefficients we opted to weigh the four variables equally in the composite. All four variables were next normalised using a min-max approach (across the two time points considered). The four variables were aggregated using the geometric average.

Following the multivariate statistical analysis, we present the scores and rankings of the composite indicator. Among the most excellent countries in research are Switzerland, Sweden, Denmark, The Netherlands, and Israel. Countries that are ranked in the middle involve both big countries (like Germany, Japan, France, and the United States) and small countries (like Belgium and Austria). Lower ranked countries are both emerging economies (like Brazil, India, and China) and central –and Eastern European countries (like Latvia, Croatia, and Russia). Overall, this ranking seems to be very much in line with the outcomes of the cluster analysis.

After the presentation of the scores and rankings, we performed sensitivity analysis to assess how volatile the proposed composite indicator is to the particular methodological choices made throughout its construction. As it turns out, in general the proposed composite indicator is not particularly sensitive to the methodological choices made. However, the composite indicator does seem highly volatile as to the choice of denominator; that is, dividing highly cited publications, high-quality patents, world class universities and research institutes, and ERC grants by respectively the total number of publications, the number of inhabitants, total R & D expenditures, and public R & D expenditures or other denominators such as GDP. Also, the aggregation method (geometric rather than arithmetic) sems to matter a great deal.

Finally, we assessed the relations between on the one hand, the proposed composite indicator measuring research excellence and, on the other hand variables or composite indicators that measure phenomena with which research excellence is often associated. In a first analysis we assessed the validity of the composite indicators on research excellence by comparing them with composite indicators that capture innovation and competitiveness. Given the distinct positive correlations found between research excellence and the various indices on innovation and competitiveness, we conclude that the composite indicator on research excellence is valid. The correlation analysis shows that the composite indicators on research excellence are akin to but measuring a somewhat different phenomenon than indicators measuring the innovativeness and competitiveness of a country. Likewise, the composite indicators on research excellence generally correlate positively with variables measuring other dimensions of national research systems (i.e. research assets, structural research capabilities, and research interactions) and GDP per capita. For both, the correlations are generally positive and significant.

In conclusion, we believe that the proposed composite indicator fills a gap in measuring research excellence at the country level and have an added value on top of other country-level performance indicators in the field of science and technology assessment. Measured against the most commonly used composite indicators of innovation and competitiveness it is shown that the proposed composite indicator on research excellence is akin to but also different from measurements dealing with the phenomena of innovation and competitiveness. In addition, the results from the correlations among the composite indicator on research excellence with on the one hand variables measuring other features of national research systems and on the other hand macroeconomic variables provide interesting and promising avenues for further research.

7.2. Discussion

Before turning to the implications of our work, at least two issues require further deliberation. First, it should be noted that the proposed composite indicator on research excellence at the country level is scarcely populated in terms of the number of underlying variables included; especially when it comes to the number of variables that cover the technology dimension (only one) and the number of variables that cover new knowledge that is embodied in people therein (no variables). This is not problematic in the sense of the proposed indicator being susceptible to biases in the direction of science or texts. Different weighting and aggregation schemes account for the potential issue of a composite indicator being biased towards one or more of the included variables at the cost of other included variables. Rather, it is problematic in the sense of this indicator being susceptible to what might be called omitted variable bias. The results of the analysis might be driven by variables not included in the analysis. That is, would these potential variables be included in the analysis the results obtained could be completely different. In the absence of such variables it is hard, if not impossible, to affirm or reject such a bias. One can do it on theoretical grounds, but not entirely on empirical grounds.

This leaves us to argue that we do not expect that the proposed composite indicator on scientific and technological research excellence suffers considerably from omitted variable bias. The reason to believe so is primarily twofold. On the one hand we observe that both variables capturing new knowledge that is embodied in texts are highly correlated and on the other hand we observe that all variables on scientific excellence correlate highly with the one variable on technological research excellence as embodied in texts. Although this need not necessarily be so, these twin observations make us belief that any variable capturing new technological knowledge that is embodied in people is likely to behave grossly in accordance with the variables already included in our analysis rendering large changes less likely.

Second, following up on the first discussion point, although the proposed composite indicator is, excepted for the choice of denominators and the aggregation method, not extremely sensitive to the particular methodological choices made throughout its construction (as shown by the results of the sensitivity analysis presented in section 6.2 of this report), this need not imply that the proposed composite indicator on research excellence is not sensitive at all. For one thing, in defining research excellence with reference to science and technology other dimensions to research are excluded from the assessment. Also, some data that were not available might shed a different light on the phenomenon of research excellence altogether. That is, even if we rephrase our composite indicator to say something about scientific and technological research excellence only (as we did in section 3.2 of this report), measuring this restricted phenomenon is still conditional upon our inclusion of some variables and not others (either because these variables are not available to us or not available at all). What is of immedtiate concern though, are the choice of denominators and aggregation method, as the sensitivity analysis clearly showed most variability here.

On top of performing sensitivity analysis, what is needed in addition is performing sensitivity auditing that goes beyond an assessment of technical (i.e. mathematical and statistical) uncertainties to include an exploration of the space of uncertain assumptions underlying the particular conceptual models and data used (Saltelli et al., 2012). Hence, we take the proposed composite indicator on research excellence as a necessary step to inform research policymakers, but also as a first and preliminary step in the ongoing debate on measuring research excellence and informing policymakers therewith. As such, we agree with Barré (2001, p. 264) who argued that:

> '*quantitative indicators are the starting point for the discussion, with their raison d'être being to be criticised in terms of their (limited) relevance and (limited) comparability. Indicators are considered thus, not as a final result to be accepted, but as an entry point for debate. This is an excellent way to enter an exercise of learning-by-comparing, which is what benchmarking is about. Criticism must be careful and positive, because the purpose is not to dismiss the legitimacy of the exercise, but to help dig further for a better understanding of the situation.*'

In all, the validity of the proposed composite indicator on research excellence does not just depend on its statistical soundness rather than on the indicator being accepted by the community of people it seeks to address.

7.3. Recommendations

From the analysis and results presented in this report we make three recommendations. One recommendation involves the use of the proposed composite indicator on research excellence as an input to the broader debate on measuring and monitoring research activities at the country level. As argued, we take the proposed composite indicator on research excellence as a necessary but also preliminary step to inform research policymakers. From the growing concerns on possible inadvertent consequences of taking the measure on science and technology (Weingart, 2005, Burrows, 2012) and its related concerns with the neoliberalisation of research (Moore et al., 2011, Mirowski, 2011), we deem it necessary if not inevitable to include such voices in the construction of a valid indicator on research excellence. This is and needs to be an ongoing project. As argued, the validity of an indicator does not just depend on its statistical soundness rather than on the indicator being accepted by the community of people it seeks to address.

Another, related recommendation concerns the necessity of collecting and using alternative data and methods for the analysis. Some data might be nearby; others further away, not to say out of range altogether. For example, some data such as those on science citations in patents and patent citations in scientific publications might be readily used provided that we have access to them. Other data such as science's appearances in the media might be further away (science in the media, impact of technology on the arts); not to speak of including measurements of research excellence that go beyond the domains of science and technology. However, and notwithstanding the difficulties in collecting alternative data that capture the phenomenon of research excellence at the country level, measuring and monitoring research excellence appropriately would greatly benefit from alternative data becoming available. As to using alternative methods, given that the results of the sensitivity analysis show that the proposed composite indicator is sensitive to particular methodological choices, these choices need to be discussed more thoroughly and might need to be revised in the future (especially when it comes to the choice of denominators and aggregation method).

As to our recommendations for policy in the field of science and technology, we stress to note that the results presented in this report are conditional upon the conceptual choices that we made (and sometimes were forced to make due to our ignorance) and the data that we used (and sometimes were restricted in using due to a lack in available alternatives). Likewise, any policy recommendation that follows from the results thus presented is conditional upon the errors made throughout the process of constructing the proposed composite indicator on research excellence. Are these errors big? As to the statistical errors we are confident that these are small. The results of the sensitivity analysis are telling in this respect. As to the errors that stem from the underlying conceptual framework and the use of particular variables; frankly, we do not know how big these errors are or can be. These could be infinitely small or infinitely large. Isn't there anything then that we can recommend from our assessment of countries' research excellence to policy-makers in the field of science and technology? Well, we can. However, we need to learn to live with the highly provisional character of such recommendations.

Having said that, what we can recommend from the outcomes of our analysis that we have now is that (i) relatively poorly performing countries need to focus more on establishing research excellence embodied by people and (ii) for most countries there is room for improvement in overall research excellence by either improving their scientific or technological research excellence. The first recommendation follows from the observation that some countries' relative under-performance is mainly due to their relatively low scores for the number of world class universities and research institutes and the number of received ERC grants (both variables measuring (scientific) research excellence embodied in people) and not so much for the number of highly cited publications and the number of high-quality patents (both indicators of research excellence embodied in texts and artifacts). The second recommendation follows from the observation that few countries perform equally good in both scientific and technological research excellence. That is, scientific and technological research excellence do not necessarily go hand in hand. Here, we do not favour one form of research excellence over the other. What holds then is that, either way, there is room for improving research excellence in most if not all countries.

Overall, we stress to note that research excellence is a complex, multidimensional phenomenon; especially when positioned within the context of national research systems and the economy at large. Correlations with other dimensions of national research systems are generally high (though never extremely high). Yet, from these correlations we can make no inferences about the causal nature of these relationships

References

ABRAMOVITZ, M. 1956. Resource and Output Trends in the United States Since 1870. *The American Economic Review,* 46**,** 5-23.

AGHION, P., DEWATRIPONT, M., HOXBY, C., MAS-COLELL, A. & SAPIR, A. 2010. The governance and performance of universities: evidence from Europe and the US. *Economic Policy,* 25**,** 7-59.

AMIN, A. 1994. *Post-Fordism: a reader,* Oxford; Cambridge, Mass., Blackwell.

ARROW, K. 1962. Economic welfare and the allocation of resources for invention. *The rate and direction of inventive activity: Economic and social factors.* Nber.

BALCONI, M., BRUSONI, S. & ORSENIGO, L. 2010. In defence of the linear model: An essay. *Research Policy,* 39**,** 1-13.

BARRÉ, R. 2001. Sense and nonsense of S&T productivity indicators. *Science and Public Policy,* 28**,** 259-266.

BARRÉ, R. 2005. S&T indicators for policy-making in a changing science-society relationship. *Handbook of quantitative science and technology research.* Springer.

BARRÉ, R., HOLLANDERS, H. & SALTER, A. 2011. Indicators of research excellence. Retreived from: http://ec.europa.eu/commission_2010-2014/geoghegan-quinn/hlp/documents/20120309-hlp-research-excellence_en.pdf.

BATTY, M. 2003. The geography of scientific citation. *Environment and Planning A,* 35**,** 761-765.

BAUMAN, Z. 2000. *Liquid modernity,* Cambridge, UK/Malden, MA, Polity Press.

BECK, U. 1992. *Risk society: towards a new modernity,* London; Newbury Park, Calif., Sage Publications.

BELL, D. The coming of the post-industrial society. 1976 1976. Taylor & Francis, 574-579.

BONACCORSI, A. & DARAIO, C. 2005. Econometric approaches to the analysis of productivity of R & D systems. *Handbook of quantitative science and technology research.* Springer.

BORNMANN, L. 2012. Measuring the societal impact of research. *EMBO reports,* 13**,** 673-676.

BORNMANN, L. & DANIEL, H.-D. 2008. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation,* 64**,** 45-80.

BORNMANN, L., LEYDESDORFF, L., WALCH-SOLIMENA, C. & ETTL, C. 2011. Mapping excellence in the geography of science: An approach based on Scopus data. *Journal of Informetrics,* 5**,** 537-546.

BOSCHMA, R. 2004. Competitiveness of Regions from an Evolutionary Perspective. *Regional Studies,* 38**,** 1001-1014.

BOSCHMA, R. 2005. Proximity and Innovation: A Critical Assessment. *Regional Studies,* 39**,** 61-74.

BOWKER, G. C. & STAR, S. L. 1999. *Sorting Things Out: Classification and Its Consequences*, MIT Press.

BRISTOW, G. 2005. Everyone's a `winner': problematising the discourse of regional competitiveness. *Journal of Economic Geography,* 5**,** 285-304.

BROOKS, T. A. 1986. Evidence of complex citer motivations. *Journal of the American Society for Information Science,* 37**,** 34-36.

BURROWS, R. 2012. Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review,* 60**,** 355-372.

CAIRNCROSS, F. 1997. *The death of distance: how the communications revolution will change our lives,* Boston, Mass., Harvard Business School Press.

CARLSSON, B., JACOBSSON, S., HOLMÉN, M. & RICKNE, A. 2002. Innovation systems: analytical and methodological issues. *Research Policy,* 31**,** 233-245.

CASTELLACCI, F. & ARCHIBUGI, D. 2008. The technology clubs: The distribution of knowledge across nations. *Research Policy,* 37**,** 1659-1673.

CASTELLACCI, F. & NATERA, J. M. 2011. A new panel dataset for cross-country analyses of national systems, growth and development (CANA). *Innovation and Development,* 1**,** 205-226.

CASTELLS, M. 1996. *The rise of the network society,* Malden, Mass., Blackwell Publishers.

CHESSA, A., MORESCALCHI, A., PAMMOLLI, F., PENNER, O., PETERSEN, A. M. & RICCABONI, M. 2013. Is Europe Evolving Toward an Integrated Research Area? *Science,* 339**,** 650-651.

CIMOLI, M., DOSI, G. & STIGLITZ, J. E. 2009. The political economy of capabilities accumulation: The past and future of policies for industrial development. *Mario Cimoli, Giovanni Dosi and Joseph E. Stiglitz*.

COLE, S. 1989. Citations and the evaluation of individual scientists. *Trends in Biochemical Sciences,* 14**,** 14.

COLE, S. & COLE, J. R. 1967. Scientific Output and Recognition: A Study in the Operation of the Reward System in Science. *American Sociological Review,* 32**,** 377-390.

COLLINS, H. 2010. *Tacit and Explicit Knowledge*, University of Chicago Press.

COMMISSION, E. 2011. Innovation Union Scoreboard. ENTERPRISE & INDUSTRY MAGAZINE.

COOKE, P. 2001. Regional Innovation Systems, Clusters, and the Knowledge Economy. *Industrial and Corporate Change,* 10**,** 945-974.

COOKE, P., GOMEZ URANGA, M. & ETXEBARRIA, G. 1997. Regional innovation systems: Institutional and organisational dimensions. *Research Policy,* 26**,** 475-491.

COWAN, R. & FORAY, D. 1997. The Economics of Codification and the Diffusion of Knowledge. *Industrial and Corporate Change,* 6**,** 595-622.

CRONBACH, L. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika,* 16**,** 297-334.

DASGUPTA, P. & DAVID, P. A. 1994. Toward a new economics of science. *Research Policy,* 23**,** 487-521.

DAVID, P. A. & FORAY, D. 2002. An introduction to the economy of the knowledge society. *International Social Science Journal,* 54**,** 9-23.

DERNIS, H. 2007. Nowcasting Patent Indicators. *OECD Science, Technology and Industry Working Papers* [Online]. Available: http://dx.doi.org/10.1787/055888673307.

DIMAGGIO, P. J. & POWELL, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review,* 48**,** 147-160.

DOSI, G. 1988. *Technical change and economic theory,* London; New York, Pinter Publishers.

DOSI, G., LLERENA, P. & LABINI, M. S. 2006. The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called 'European Paradox'. *Research Policy,* 35**,** 1450-1464.

DRORI, G. S. 2003. *Science in the modern world polity: institutionalization and globalization,* Stanford, Calif., Stanford University Press.

DRUCKER, P. F. 1969. *The age of discontinuity; guidelines to our changing society,* New York,, Harper & Row.

DUTTA, S. 2012. *The Global Innovation Index 2012: Stronger Innovation Linkages for Global Growth*, INSEAD.

EDGE, D. 1979. Quantitative measures of communication in science: A critical review. *History of Science Cambridge,* 17**,** 102-134.

EDQUIST, C. 1997. *Systems of innovation: technologies, institutions, and organizations,* London; Washington, Pinter.

EDQUIST, C. 2005. Systems of innovation. *The Oxford handbook of innovation***,** 181-208.

EDQUIST, C. & ZABALA, J. M. 2009. Outputs of innovation systems: a European perspective.

ELZINGA, A. 2012. Features of the current science policy regime: Viewed in historical perspective. *Science and Public Policy,* 39**,** 416-428.

ETZKOWITZ, H. 2003. Research groups as 'quasi-firms': the invention of the entrepreneurial university. *Research Policy,* 32**,** 109-121.

ETZKOWITZ, H. & LEYDESDORFF, L. 2000. The dynamics of innovation: from National Systems and 'Mode 2' to a Triple Helix of university–industry–government relations. *Research Policy,* 29**,** 109-123.

FEALING, K. H. 2011. *The science of science policy: a handbook,* Stanford, Calif., Stanford Business Books.

FELDMAN, M. P. 1999. The New Economics Of Innovation, Spillovers And Agglomeration: Areview Of Empirical Studies. *Economics of Innovation and New Technology,* 8**,** 5-25.

FILIPPETTI, A. & ARCHIBUGI, D. 2011. Innovation in times of crisis: National Systems of Innovation, structure, and demand. *Research Policy,* 40**,** 179-192.

FLEMING, L. & SORENSON, O. 2004. Science as a map in technological search. *Strategic Management Journal,* 25**,** 909-928.

FORAY, D. 2004. *Economics of knowledge,* Cambridge, Mass., MIT Press.

FREEMAN, C. 1987. *Technology, policy, and economic performance: lessons from Japan,* London; New York, Pinter Publishers.

FREEMAN, C. 1995. The 'National System of Innovation' in historical perspective. *Cambridge Journal of Economics,* 19**,** 5-24.

FRENKEN, K., HARDEMAN, S. & HOEKMAN, J. 2009. Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics,* 3**,** 222-232.

FRIEDMAN, T. L. 2005. *The world is flat: a brief history of the twenty-first century,* New York, Farrar, Straus and Giroux.

FUNTOWICZ, S. O. & RAVETZ, J. R. 1990. *Uncertainty and Quality in Science for Policy*, Springer.

FUNTOWICZ, S. O. & RAVETZ, J. R. 1993. Science for the post-normal age. *Futures,* 25**,** 739-755.

GIBBONS, M., LIMOGES, C., NOWOTNY, H., SCHWARTZMAN, S., SCOTT, P. & TROW, M. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, SAGE Publications.

GIDDENS, A. 1990. *The Consequences of Modernity*, Stanford University Press.

GILBERT, G. 1978. Measuring the growth of science. *Scientometrics,* 1**,** 9-34.

GODIN, B. 2001. Defining R & D: Is research always systematic? *Project on the History and Sociology of S&T Statistics 7* [Online]. Available: http://www.csiic.ca/PDF/Godin_7.pdf.

GODIN, B. 2006. The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology & Human Values,* 31**,** 639-667.

GODIN, B. & DORÉ, C. 2004. Measuring the impacts of science: Beyond the economic dimension.

GODIN, B. & GINGRAS, Y. 2000. The place of universities in the system of knowledge production. *Research Policy,* 29**,** 273-278.

GRILICHES, Z. 1986. Chapter 25 Economic data issues. *In:* ZVI GRILICHES AND MICHAEL, D. I. (ed.) *Handbook of Econometrics.* Elsevier.

GROENEVELD, R. A. & MEEDEN, G. 1984. Measuring Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician),* 33**,** 391-399.

GRUSS, P. 2012. Driven by Basic Research. *Science,* 336**,** 392-392.

HAGERTY, M. R. & LAND, K. C. 2007. Constructing Summary Indices of Quality of Life: A Model for the Effect of Heterogeneous Importance Weights. *Sociological Methods & Research,* 35**,** 455-496.

HALL, B. H. & JAFFE, A. B. 2012. Measuring Science, Technology, and Innovation: A Review.

HALL, B. H., MAIRESSE, J. & MOHNEN, P. 2010. Chapter 22 — Measuring the Returns to R&D. *In:* BRONWYN, H. H. A. N. R. (ed.) *Handbook of the Economics of Innovation.* North-Holland.

HARDEMAN, S. 2012. The distributed organization of science: with empirical illutrations from the field of diabetes medicine. Eindhoven: Eindhoven University of Technology.

HARVEY, D. 1989. *The Condition of Postmodernity: An Enquiry Into the Origins of Cultural Change*, Wiley.

HEIMERIKS, G. & BOSCHMA, R. 2013. The path- and place-dependent nature of scientific knowledge production in biotech 1986–2008. *Journal of Economic Geography*.

HESSELS, L. K. & VAN LENTE, H. 2008. Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy,* 37**,** 740-760.

HICKS, D. 2012. Performance-based university research funding systems. *Research Policy,* 41**,** 251-261.

HICKS, D. M. & KATZ, J. S. 1996. Where Is Science Going? *Science, Technology & Human Values,* 21**,** 379-406.

HOEKMAN, J. 2012. Science in an age of globalisation: the geography of research collaboration and its effect on scientific publishing. Eindhoven: Eindhoven University of Technology.

HOEKMAN, J., FRENKEN, K. & OORT, F. 2009. The geography of collaborative knowledge production in Europe. *The Annals of Regional Science,* 43**,** 721-738.

HOEKMAN, J., FRENKEN, K. & TIJSSEN, R. J. W. 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy,* 39**,** 662-673.

JAFFE, A. B., TRAJTENBERG, M. & ROMER, P. M. 2002. *Patents, Citations, And Innovations: A Window On The Knowledge Economy*, Mit Press.

KAPLAN, N. 1965. The norms of citation behavior: Prolegomena to the footnote. *American Documentation,* 16**,** 179-184.

KATZ, J. S. 2006. Indicators for complex innovation systems. *Research Policy,* 35**,** 893-909.

KAUFMAN, L. & ROUSSEEUW, P. J. 2009. *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley.

KITSON, M., MARTIN, R. & TYLER, P. 2004. Regional Competitiveness: An Elusive yet Key Concept? *Regional Studies,* 38**,** 991-999.

KLINE, S. J. & ROSENBERG, N. 1986. An overview of innovation. *The positive sum strategy: Harnessing technology for economic growth,* 14**,** 640.

KNIGHT, F. H. 1921. *Risk, Uncertainty and Profit*, Hougthon Mifflin Company.

KRUGMAN, P. 1996. Making sense of the competitiveness debate. *Oxford Review of Economic Policy,* 12**,** 17-25.

LAREDO, P. 2007. Revisiting the Third Mission of Universities: Toward a Renewed Categorization of University Activities? *Higher Education Policy,* 20**,** 441-456.

LERNER, J. & TIROLE, J. 2002. Some Simple Economics of Open Source. *The Journal of Industrial Economics,* 50**,** 197-234.

LEYDESDORFF, L. 2006. *The Knowledge-based Economy: Modeled, Measured, Simulated*, Universal Publishers.

LEYDESDORFF, L. & ETZKOWITZ, H. 1996. Emergence of a Triple Helix of universityindustrygovernment relations. *Science and Public Policy,* 23**,** 279-286.

LEYDESDORFF, L. & RAFOLS, I. 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology,* 60**,** 348-362.

LIPOVETSKY, G., CHARLES, S. B. & BROWN, A. 2005. *Hypermodern times,* Cambridge, UK; Malden, MA, Polity.

LUNDVALL, B.-A. 1988. Innovation as an interactive process: from user-producer interaction to the national system of innovation. *Technical change and economic theory,* 369.

LUNDVALL, B.-A., JOHNSON, B., ANDERSEN, E. S. & DALUM, B. 2002. National systems of production, innovation and competence building. *Research Policy,* 31**,** 213-231.

LUNDVALL, B. Å. 2007. National Innovation Systems—Analytical Concept and Development Tool. *Industry and Innovation,* 14**,** 95-119.

LUNDVALL, B. Å. 2010. *National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning*, Anthem Press.

MACILWAIN, C. 2010. Science economics: What science is really worth. *Nature,* 465**,** 682-684.

MACROBERTS, M. & MACROBERTS, B. 1996. Problems of citation analysis. *Scientometrics,* 36**,** 435-444.

MACROBERTS, M. H. & MACROBERTS, B. R. 2010. Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology,* 61**,** 1-12.

MARKUSEN, A. 1999. Fuzzy Concepts, Scanty Evidence, Policy Distance: The Case for Rigour and Policy Relevance in Critical Regional Studies. *Regional Studies,* 33**,** 869-884.

MARTIN, B. R. & IRVINE, J. 1983. Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy,* 12**,** 61-90.
MARTY, T. 2012. Research efficiency: Clean up the waste. *Nature,* 484**,** 27-28.
MCCANN, P. 2008. Globalization and economic geography: the world is curved, not flat. *Cambridge Journal of Regions, Economy and Society,* 1**,** 351-370.
MERTON, R. K. & STORER, N. W. 1979. *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press.
MICHALOS, A. 2011. What Did Stiglitz, Sen and Fitoussi Get Right and What Did They Get Wrong? *Social Indicators Research,* 102**,** 117-129.
MIROWSKI, P. 2011. *Science-Mart*, Harvard University Press.
MOED, H. F. 2005. *Citation Analysis in Research Evaluation*, Springer.
MOORE, K., KLEINMAN, D., HESS, D. & FRICKEL, S. 2011. Science and neoliberal globalization: a political sociological approach. *Theory and Society,* 40**,** 505-532.
MUNDA, G. 2008. Social Multi-Criteria Evaluation for a Sustainable Economy.
NARDO, M., SAISANA, M., SALTELLI, A., TARANTOLA, S., HOFFMAN, A. & GIOVANNINI, E. 2005. Handbook on constructing composite indicators: methodology and user guide. OECD publishing.
NAUWELAERS, C. & WINTJES, R. 2008. *Innovation Policy in Europe: Measurement and Strategy*, Edward Elgar.
NEDEVA, M. & STAMPFER, M. 2012. From 'science in Europe' to 'European science'. *Science,* 336**,** 982-983.
NEFFKE, F., HENNING, M. & BOSCHMA, R. 2011. How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography,* 87**,** 237-265.
NELSON, R. R. 1959. The Simple Economics of Basic Scientific Research. *Journal of Political Economy,* 67**,** 297-306.
NELSON, R. R. 1962. The rate and direction of inventive activity. *Princeton, NJ*.
NELSON, R. R. 1993. National innovation systems: A comparative study. Oxford university press, Oxford.
NELSON, R. R. & NELSON, K. 2002. Technology, institutions, and innovation systems. *Research Policy,* 31**,** 265-272.
NELSON, R. R. & WINTER, S. G. 1982. *An Evolutionary Theory of Economic Change*, Belknap Press.
NOWOTNY, H. 2006. Real science is excellent science–how to interpret post-academic science, Mode 2 and the ERC. *JCOM,* 5**,** 4.
NOWOTNY, H., SCOTT, P. & GIBBONS, M. 2001. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*, Wiley.
NUNNALLY, J. C. & BERNSTEIN, I. H. 1978. Psychometric theory. New York: McGraw-Hill.
OECD 2002. *Frascati Manual 2002*, OECD Publishing.
OECD 2004. *PISA Learning for Tomorrow's World: First Results from PISA 2003*, OECD Publishing.
OECD 2012. Measuring patent 'quality': refined indicators for cross-country evidence Paris: OECD Publishing.
PAVITT, K. 2000. Why European Union funding of academic research should be increased: a radical proposal. *Science and Public Policy,* 27**,** 455-460.
PEREZ, C. & SOETE, L. 1988. Catching up in technology: entry barriers and windows of opportunity. *Technical change and economic theory***,** 458-479.
PIORE, M. J. & SABEL, C. F. 1984. *The Second Industrial Divide: Posibilities for Prosperity*, Basic Books.
POWELL, W. W. & SNELLMAN, K. 2004. The Knowledge Economy. *Annual Review of Sociology,* 30**,** 199-220.
POWER, D. & MALMBERG, A. 2008. The contribution of universities to innovation and economic development: in what sense a regional problem? *Cambridge Journal of Regions, Economy and Society,* 1**,** 233-245.

POWER, M. 1994. *The Audit Explosion*, Demos.
POWER, M. 1997. *The Audit Society: Rituals of Verification*, Oxford University Press.
PUTNAM, J. 1996. The value of international patent protection. *Unpublished Ph. D. Dissertation, Yale University*.
RALLET, A. 1993. Choix de proximité et processus d'innovation technologique. *Revue d'économie régionale et urbaine,* 3**,** 365-386.
RAYMOND, E. S. 2008. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly Media.
ROSENBERG, N. 1982. *Inside the Black Box: Technology and Economics*, Cambridge University Press.
SAISANA, M., D'HOMBRES, B. & SALTELLI, A. 2011. Rickety numbers: Volatility of university rankings and policy implications. *Research Policy,* 40**,** 165-177.
SAISANA, M., SALTELLI, A. & TARANTOLA, S. 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 168**,** 307-323.
SALA-I-MARTÍN, X., BILBAO-OSORIO, B., BLANKE, J., CROTTI, R., HANOUZ, M. D., GEIGER, T. & KO, C. 2012. The Global Competitiveness Index 2012–2013: Strengthening Recovery by Raising Productivity. *The Global Competitiveness Report 2012–2013***,** 1-48.
SALTELLI, A., RATTO, M., ANDRES, T., CAMPOLONGO, F., CARIBONI, J., GATELLI, D., SAISANA, M. & TARANTOLA, S. 2008. *Global Sensitivity Analysis: The Primer*, Wiley.
SALTER, A. J. & MARTIN, B. R. 2001. The economic benefits of publicly funded basic research: a critical review. *Research Policy,* 30**,** 509-532.
SLAUGHTER, S. & RHOADES, G. 1996. The Emergence of a Competitiveness Research and Development Policy Coalition and the Commercialization of Academic Science and Technology. *Science, Technology & Human Values,* 21**,** 303-339.
SOLOW, R. M. 1957. Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics,* 39**,** 312-320.
STARK, D. 2011. *The Sense of Dissonance: Accounts of Worth in Economic Life*, Princeton University Press.
STEPHAN, P. 2012. *How economics shapes science*, Harvard University Press.
STEPHAN, P. E. 1996. The Economics of Science. *Journal of Economic Literature,* 34**,** 1199-1235.
STOKES, D. E. 1997. *Pasteur's quadrant: Basic science and technological innovation*, Brookings Inst Press.
STOKOLS, D., HALL, K. L., TAYLOR, B. K. & MOSER, R. P. 2008. The science of team science: overview of the field and introduction to the supplement. *American journal of preventive medicine,* 35**,** S77-S89.
THUROW, L. 1992. Head to head: The coming battle among Japan, Europe, and America. *New York: Morrow*.
TIJSSEN, R., VISSER, M. & VAN LEEUWEN, T. 2002. Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics,* 54**,** 381-397.
TIJSSEN, R. J. W. 2003. Scoreboards of research excellence. *Research Evaluation,* 12**,** 91-103.
VAN DIJK, P. J. A. G. M. 2005. *The Network Society: Social Aspects of New Media*, SAGE Publications.
VAN LOOY, B., DEBACKERE, K., CALLAERT, J., TIJSSEN, R. & VAN LEEUWEN, T. 2006. Scientific capabilities and technological performance of national innovation systems: An exploration of emerging industrial relevant research domains. *Scientometrics,* 66**,** 295-310.
VAN RAAN, A. F. J. 2005. For Your Citations Only? Hot Topics in Bibliometric Analysis. *Measurement: Interdisciplinary Research and Perspectives,* 3**,** 50-62.
VAN RAAN, A. F. J. 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment</p> for 147 chemistry research groups</p> </p>. *Scientometrics,* 67**,** 491-502.

VÉRTESY, D. & TARANTOLA, S. 2012. Composite indicators of research excellence. Ispra: Joint Research Centre of the European Commission.
WEBSTER, F. 2006. *Theories of the information society*, Taylor & Francis.
WEINGART, P. 2005. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics,* 62**,** 117-131.
WHITLEY, R. & GLÄSER, J. 2007. *The changing governance of the sciences: The advent of research evaluation systems*, Springer Science+ Business Media.
WICHMANN MATTHIESSEN, C., WINKEL SCHWARZ, A. & FIND, S. 2002. The Top-level Global Research System, 1997-99: Centres, Networks and Nodality. An Analysis Based on Bibliometric Indicators. *Urban Studies,* 39**,** 903-927.
WITT, U. 2009. Propositions about novelty. *Journal of Economic Behavior & Organization,* 70**,** 311-320.
ZIMAN, J. M. 1994. *Prometheus bound*, Cambridge University Press.
ZUCKER, L. G., DARBY, M. R. & ARMSTRONG, J. S. 2002. Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. *Management Science,* 48**,** 138-153.
ZUNIGA, P., GUELLEC, D., DERNIS, H., KHAN, M., OKAZAKI, T. & WEBB, C. 2009. OECD patent statistics manual. Paris: OECD Publications.

Acknowledgements

The authors express their gratitude for valuable comments made by Marco Vivarelli (Catholic University of Milano), Johan Stierna (DG RTD), Raquel Saiz Tome (DG RTD), Andrea Saltelli (JRC) and Stefano Tarantola (JRC) throughout the whole project. Comments made by the participants of the workshop on 'Indicators for the Innovation Union' (Ispra, Italy; October 25-26, 2012) are highly appreciated. As always, all remaining errors are those of the authors.

European Commission
EUR 26093 EN – Joint Research Centre – Institute for the Protection and Safety of the Citizen

Title: An analysis of national research systems (I): a composite indicator for scientific and technological research excellence

Author(s): Daniel Vertesy, Sjoerd Hardeman, Vincent van Roy, Michaela Saisana

Luxembourg: Publications Office of the European Union

2013 – 88 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424

ISBN 978-92-79-32527-4

doi: 10.2788/95887

Abstract

This paper first develops a framework for the analysis of national research systems, with particular focus on the excellence of scientific and technological research, a central topic for the current European research and innovation policy discourse. Next, after carefully considering measurement and data issues, it proposes a set of strong and weaker country-level indicators of research excellence, based on which a composite indicator of scientific and technological research excellence is proposed and tested for robustness and sensitivity.
The research was conducted on behalf of DG RTD within the framework of the “Composites 4IU” project.

LB-NA-26093-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-32527-4