Imperial College London

NETWORKS OF INNOVATION: MEASURING, MODELLING AND ENHANCING INNOVATION IN SURGERY

Georgios Garas

BSc (Hons) MBBS (Dist) MRCS (Eng) FRCS (ORL-HNS) FEB ORL-HNS

Department of Surgery and Cancer

Imperial College London

2020

Thesis submitted to Imperial College London for the award of

Doctorate of Philosophy (PhD)

Supervisors:

Professor Thanos Athanasiou *MD PhD MBA FRCS*

Professor The Lord Ara Darzi *MD OM KBE PC FRCS*

Supported by:

The Royal College of Surgeons of England

Alexander S. Onassis Public Benefit Foundation

Declaration of Originality

I hereby declare that I am the sole author of this thesis and that all work within it is my own. Any individuals who carried out work in collaboration with the author are appropriately credited and all reviewed literature appropriately referenced.

Copyright Declaration

Abstract

The rate of innovation occurring in surgery is beyond our systemic capacity to quantify, with several methodological and practical challenges. The existing paucity of surgical innovation metrics presents a global healthcare problem especially as surgical innovations become increasingly costlier at a time when healthcare provision is experiencing a radical transformation driven by pressures to reduce costs, an ageing population with ever-increasing healthcare needs and patients with growing expectations.

This thesis aims to devise a novel, quantitative, network-based framework that will permit modelling and measuring surgical innovation to add the most value to patient care. It involves the systematic, graphical and analytical assessment of surgical innovation in a way that has never been done before. This is based on successful models previously applied in the industry with advanced analytical techniques derived from social science (network analysis). In doing so, it offers an exciting new perspective and opportunity for understanding how the innovation process originates and evolves in surgery and how it can be measured in terms of value and virality, a priority for the NHS, RCS, Imperial and the wider surgical community.

The ability to measure value and rank innovations is expected to play a fundamental role in guiding policy, strategically direct surgical research funding, and uncover innovation barriers and catalysts. This will ensure participation in the forefront of novel surgical technology and lay the scientific foundations for the development of improved healthcare models and services to enhance the quality of healthcare delivered.

Peer-reviewed publications

Published papers

Garas G, Cingolani I, Patel V, Panzarasa P, Darzi A, Athanasiou T. Evaluating the implications of Brexit for research collaboration and policy: A network analysis and simulation study. *BMJ Open*. 2019 Sep 10;9(9):e0250525.

Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Beyond IDEAL: the importance of surgical innovation metrics. *Lancet*. 2019 Jan 26;393(10169):315.

Garas G, Cingolani I, Patel V, Panzarasa P, Alderson D, Darzi A, Athanasiou T. Surgical innovation in the Era of Global Surgery: A Network Analysis. *Ann Surg.* 2018 Dec 20. [Epub ahead of print]

Garas G, Patel V, Cingolani I, Panzarasa P, Ashrafian H, Darzi A, Athanasiou T. Future of Surgery: Process Transformation is required to quantify Surgical Innovation in the era of Data Intelligence. *Future of Surgery.* The Royal College of Surgeons of England. 2018 Dec 7.

Garas G, Markar SR, Malietzis G, Ashrafian H, Hanna GB, Zacharakis E, Jiao LR, Argiris A, Darzi A, Athanasiou T. Induced Bias Due to Crossover Within Randomized Controlled Trials in Surgical Oncology: A Meta-regression Analysis of Minimally Invasive versus Open Surgery for the Treatment of Gastrointestinal Cancer. *Ann Surg Oncol.* 2018;25(1):221-230.

Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Network analysis of surgical innovation: Measuring value and virality of diffusion in robotic surgery. *PLoS One*. 2017;12(8):e0183332.

Published abstracts

Garas G, Markar S, Malietzis G, Darzi A, Athanasiou T. Induced bias due to crossover in randomised controlled trials in surgical oncology. *Br J Surg.* 2017;104(S6):36-37.

Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Networks of surgical innovation: Measuring value and the virality of diffusion The example of robotic surgery. *Br J Surg.* 2017;104(S6):229.

Fellowships, grants and prizes awarded

2017 **Royal College of Surgeons of England Doctoral Research Fellowship**

Grant value: £62,744

Fellowship awarding body: *The Royal College of Surgeons of England*

London, United Kingdom

2017 **ASGBI Best E-Poster Prize, Surgical Simulation and Technology**

Prize awarding body: *Association of Surgeons of Great Britain and Ireland*

International Surgical Congress

Glasgow, United Kingdom

2016 **Alexander S. Onassis Public Benefit Foundation Fellowship**

Grant value: £32,300

Fellowship awarding body: *Alexander S. Onassis Public Benefit Foundation*

Vaduz, Liechtenstein

Acknowledgments

I dedicate this thesis to my amazing wife, Chrysoula Palatianou-Garas, without whom, it would have not been possible to complete this work and to our magnificent son, Ioannis John Garas. I am eternally grateful for her endless patience over the years, her vital encouragement during difficult times, and for being such a supportive wife allowing me to pursue my ambitions, as well as for being a truly amazing mother to Giannis.

I also dedicate this thesis to my late father, Professor Ioannis Garas, a giant on whose shoulders I stood, and who will always be my inspiration, and to my mother, Aimilia Garas for providing my brothers and me with all the ingredients for success in life.

I am greatly indebted to my supervisors Professor Thanos Athanasiou and Professor The Lord Ara Darzi for their enduring support and guidance throughout the years of my study. It has been a pleasure and honour to work, learn and publish with them.

I would also like to express my gratitude to our collaborators Dr. Pietro Panzarasa, Reader in Organisational Theory and Behaviour at the School of Business and Management at Queen Mary, University of London and Dr. Isabella Cingolani, Research Associate at the Big Data and Analytical Unit at Imperial College London for their support with the network analysis. Without their unique expertise, it would have been very challenging to complete this work. Finally, I owe a big thank you to Mr. Vanash Patel and Mr. Hutan Ashrafian, for their valuable support and guidance all these years that we have been collaborating.

Table of Contents

List of figures

List of Tables

List of abbreviations and acronyms

AAP	Accelerated Access Pathway
AI	Artificial Intelligence
Altmetrics	Alternative Metrics
AR	Augmented Reality
aVR	Aortic Valve Replacement
Brexit	Britain's Exit (from the European Union)
CDF	Cumulative Distribution Function
CE	Conformité Européenne
CVA	CerebroVascular Accident
DL	Deep Learning
EU	European Union
FDA	Food and Drug Administration
G7	Group of Seven
GUDID	Global Unique Device Identification Database
IBM	International Business Machines Corporation
ICMJE	International Committee of Medical Journals Editors
IDEAL	Idea, Development, Exploration, Assessment, Long-term study
INTEMACS	Interagency Registry for Mechanically Assisted Circulatory Support
IP	Intellectual Property
IVC	Innovation Value Chain
MDEpiNET	Medical Device Epidemiology Network
MDR	Medical Device Regulation
MDT	Multi-Disciplinary Team
MeSH	Medical Subject Headings

MHRA	Medicines and Healthcare products Regulatory Agency
ML	Machine Learning
NHS	National Health Service
NIS	National (Nationwide) Inpatient Sample
NLP	Natural Language Processing
PIP	Poly Implant Prothèse
PMA	Pre-Market Authorisation
PMS	Post-Market Surveillance
PROMs	Patient Reported Outcome Measures
QoL	Quality of Life
R&D	Research & Development
RCS	Royal College of Surgeons of England
RCT	Randomised Controlled Trial
ROI	Return On Investment
RoW	Rest of World
SIF	Surgical Innovation Funnel
SMART	Specific, Measurable, Achievable, Results-based, and Time-bound
SV	Structural Virality
TAVI	Transcatheter Aortic Valve Implantation
TORS	TransOral Robotic Surgery
TPLC	Total Product LifeCycle
UDI	Unique Device Identification
UK	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
WOS	Web of Science®

1. Introduction

Surgery and innovation go hand-in-hand with the majority of surgeons "innovating on a daily basis".[1] Without innovation, surgery would be static with little room for improvement. Instead, surgery has witnessed unprecedented advances, most as a direct result of innovation, some of which have led to substantial improvement in patient outcomes and countless of lives saved.[1]

At present, the rate of innovation occurring in surgery is beyond our systemic capacity to quantify, with several methodological and practical challenges.[2] Earlier attempts to measure surgical innovation through publication and patent counts have had minimal uptake due to oversimplifying assumptions.[3] Existing frameworks are limited to qualitative models.[4] These include IDEAL, currently the most widely implemented surgical innovation framework with its own developers recently admitting that it has "probably only contributed in a minor way to the improvement seen so far".[5]

The qualitative nature of existing approaches to the study of surgical innovation has thwarted accurate evaluations within and across specialties. As a result, assessment and judgement are often used rather than actual measurement – a characteristic example is that of robotic surgery.[6] By looking outside healthcare, it becomes apparent that innovation metrics are extensively used. International corporations including innovation leaders such as Apple Inc. and Amazon.com, Inc. have been using innovation metrics for years, and heavily rely on these to systematically

evaluate their innovation efforts.[7] It is through innovation metrics that strategy is devised towards enhancing innovation and driving performance.[8]

The existing paucity of surgical innovation metrics presents a major healthcare problem on a global scale especially as surgical innovations become increasingly costlier.[9] At the same time, global healthcare provision (including in the NHS) is experiencing a radical transformation driven by pressures to reduce costs, an ageing population with ever-increasing healthcare needs and patients with growing expectations. It is for all these reasons that there is a growing sense of urgency to develop rigorous surgical innovation metrics, crucial for optimising patient care.[10]

Innovation - including the global spread of ideas and knowledge - is a complex, network-driven dynamic process; to be able to measure it, it must be modeled as such.[11] An effective way to model the footprints of innovation in surgery is through the use of network analysis.[4] Networks (e.g. citation, collaboration) enable the precise tracing of how surgeons discover, exchange and apply new information over time.

This thesis represents the first of its kind in surgical research. Through a number of studies, it illustrates how network analysis of big data – large, complex (multidimensional) datasets - can be utilised to study, measure, and 'visualise' the innovation process in surgery. These include the development and validation of novel surgical innovation metrics and demonstrating how this approach can be used to devise effective strategies towards the establishment of partnerships that can enhance research impact, facilitate innovation and advance patient care. Finally, the impact of marketing efforts and different regulatory systems on innovation and patient safety is

evaluated. All studies are derived from the real world and address issues of practical importance to surgeons universally.

This research offers an exciting new perspective and opportunity for understanding how the innovation process originates and evolves in surgery and how it can be measured in terms of value and virality, a priority for the NHS, RCS, Imperial and the wider surgical community. The ability to measure value and rank innovations is expected to play a fundamental role in guiding policy, strategically direct surgical research funding, and uncover innovation barriers and catalysts. This will ensure participation in the forefront of novel surgical technology and lay the scientific foundations for the development of improved healthcare models and services to enhance the quality of healthcare delivered.

2. The need for process transformation and novel surgical innovation metrics in the era of data intelligence

2.1 Introduction

Innovation has been defined as *"the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation, or external relations"*.[12] This definition has relevance to surgical innovation as it emphasizes the two key characteristics of innovation: novelty ("new") and added value ("improvement"). An additional and important point highlighted in this definition is that innovation may not necessarily be a product but can relate to a process, organisational change, or even a new marketing strategy (**Fig. 1**). Thus, as long as novelty and added value are present, the implementation of any change can be regarded as innovation.[12]

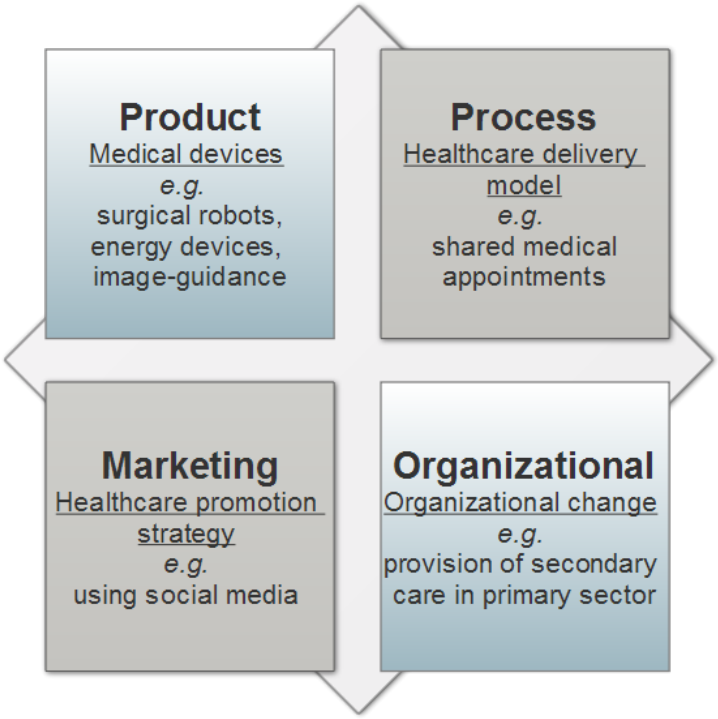


Figure 1 Examples of the four different types of surgical innovation. A 'hybrid' innovation relates to the combination of two (or more) innovation types, e.g., virtual shared medical appointments (process innovation) conducted through Skype® (product innovation).

Novelty may not necessarily refer to 'first ever appearance' of a product, service or process, but can also relate to 'first ever translation' of a product from one sector to another, for example the implementation of an established organisational model from the corporate world to healthcare. Similarly, novelty may relate to the introduction of an established product (e.g., a medical device) to a new anatomic location or patient group (e.g., first application in the paediatric population).[13] One core definition of value can derive from the quantifiable factor of financial potential so that in the healthcare sector, value has been defined as a *"meaningful outcomes achieved for a patient relative to the money spent on his or her care"*[14] (**Fig. 2**).

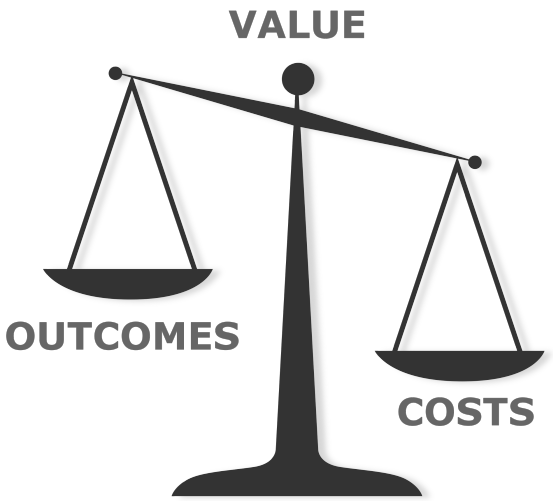


Figure 2 Value in healthcare relates to balance between improved patient outcomes and reduction in associated costs.

2.2 Value in Healthcare

The definition of value in healthcare may at first appear simple (essentially represented by the ratio between outcomes and costs),[15] however, measuring value - and thus innovation - can be challenging. This is evident by the absence of innovation metrics in many medical fields, which lies in direct contrast to other industries (such as financial technologies and commercial online sales platforms) where the strengths and weaknesses of new innovations are formally characterised through specific innovation metrics.[7,8]

Surgery is one such field where innovation metrics are lacking and here smart metrics are urgently needed if innovation is to flourish.[16] SMART, relates to an acronym commonly used in management for optimal performance indicators and stands for

Specific, Measurable, Achievable, Results-based, and Time-bound.[17] After all, *"we can only be sure to improve what we can actually measure"*.[18,19] In the era of big data, artificial intelligence (AI), and an unceasing technological revolution, the lack of surgical innovation metrics can no longer be considered sustainable. Beyond the obvious implications to patients and the wider society, there are real financial risks that are becoming increasingly apparent even in G7 country-members, the world's seven largest and most advanced economies.[20,21]

These include ageing populations and Western sedentary lifestyles that are powerful strains on country-level healthcare budgets.[22] This is further compounded by an increasing pressure to adopt expensive medical devices; many of which are promoted and often perceived as superior despite no clear evidence to support this.[23,24] Without properly measuring innovation, there is a real risk that marketing and financial factors as well as intuition rather than actual value will increasingly drive the way demand is met.

Robotic surgery is a prominent example that is characterised by marked diffusion across a number of surgical specialties in the absence of demonstrable value beyond minimally invasive prostatectomy.[25] Recent studies have shown that despite over 30 years of incremental advances, robotic surgery remains a surgical intervention of *"low value"*, representing *"a potential candidate for disinvestment"* and an *"unfulfilled promise"*.[20,26,27] Yet, despite no clear demonstration of improved outcomes over conventional laparoscopic surgery and substantially increased costs (estimated at around 13% across a range of surgical specialties), the market share of robotic surgery has continuously been expanding for over a decade now.[28] This can

only be explained by non-clinical factors driving its diffusion, such as the industry's intense marketing strategy targeting not only surgeons and other healthcare providers (*business-to-business model*) but also patients directly (*business-to-consumer model*) to promote interest and sustain demand. [29,30]

It thus becomes apparent that, to ensure progress in healthcare, the ability to measure innovation (through value created or added) is crucial.[31] It is only through innovation metrics that benchmarking, comparisons, and rankings can become possible, all critical for informing healthcare policy and ensuring the selective adoption of those innovations of highest value.[32] To do so, both components of value – outcomes and costs – must be incorporated in those metrics.

What is also important to appreciate is that outcomes should be *"meaningful"* and, as *"value should always be defined around the customer"*,[32] what should actually be measured is what patients deem important (rather than what healthcare providers and/or policy-makers choose to measure - their perspectives and priorities often differ substantially from those of patients).[14]

A variety of measures for patient outcomes have been developed and validated for different diseases and patient populations. Examples include overall and disease-specific survival rates primarily used in oncology, Quality of Life (QoL) scores and Patient Reported Outcome Measures (PROMs) used in a range of diseases to measure the impact of treatments on patients' wellbeing.[33] More recently, with an increasing trend towards value-based purchasing, composite quality measures of surgical performance have been developed and validated for certain surgical operations.[34]

Similarly, costs too should be defined around the patient, and should thus take into consideration the 'full cycle of care'.[32] This is paramount as healthcare delivery often takes place across several organisational units (e.g., primary and secondary care) that therefore need to be properly coordinated.[32] It is this fragmented nature of healthcare with costs often dispersed across a variety of healthcare providers and organisational units that substantially adds to the challenges of measuring the 'true' value of healthcare innovations.[9]

Unless a systematic framework is employed that will take into account all of the above, the likelihood is that any measurement risks being inaccurate and misleading with potentially detrimental consequences on healthcare delivery. It is also crucial that any evidence utilised in an innovation benchmarking process is derived from adequately designed sources that accommodate sound methodological comparison methods (such as demographics, economics, population size, geography, etc).

2.3 Conceptualizing surgical innovation

To conceptualize the innovation process in surgery, it is important to look outside healthcare, at sectors where innovation has been extensively studied. The *Innovation Value Chain* (IVC), a concept introduced and applied over a decade ago,[35] depicts innovation as a continuous process consisting of three successive phases linked to each other in the following order: 'idea generation', 'conversion' (also referred to as 'translation' or 'materialisation'), and 'diffusion' (relating to the uptake of innovation and associated 'market success') (**Fig. 3**).[35]

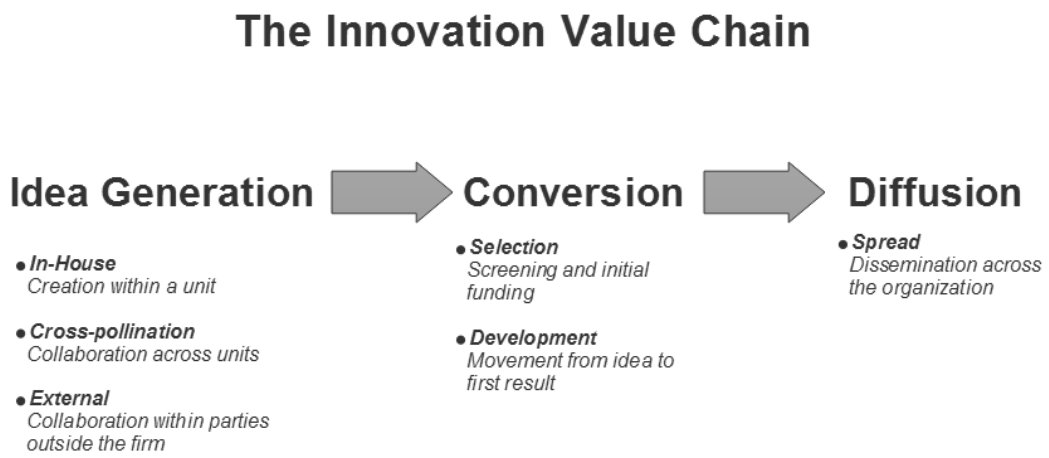


Figure 3 The three phases of the innovation process according to the innovation value chain.

All links between those three phases must be present and strong if innovation is to occur. If any link is deficient or even weak, it does not matter how strong the others are, the 'chain' will break and innovation will be compromised. Hence, it is the weakest link that dictates innovative ability.[35]

For innovation to transpire, a flow of novel ideas is vital but insufficient. Mechanisms are needed to capture and filter those ideas worthy of further evaluation. The ability to convert (translate) those ideas into actual products, processes (or other novel changes), and subsequently diffuse these across the market is paramount. Diffusion (dissemination) of innovations does not happen passively nor does it happen by chance. On the contrary, it is an active process that requires carefully planned and executed strategies for boosting the visibility of novel products, services or processes, and thus optimising their chances of uptake and becoming a success.[35]

The generation, development and diffusion of innovations are subject to a natural selection type process ('survival of the fittest').[36,37] This natural selection becomes

even more prominent in the context of surgical innovation as a result of the strict monitoring and regulatory frameworks set in place to safeguard patient safety and address ethical concerns.[38] This is particularly true when it comes to 'first-in-human' applications or when proposing the expansion of small studies to larger ones or to a new patient cohort (e.g., from an adult to a paediatric population).

To exemplify this natural selection process, this thesis introduces the original concept of the *Surgical Innovation Funnel* (SIF).[39] This consists of the different stages of implementation of surgical innovation including the description of ideas, and the pre-clinical (i.e., laboratory, animal, and cadaveric studies) as well as clinical (based on the levels of evidence) phases. Innovations progressively and selectively flow along these stages, with only the fittest ones reaching the final stage (**Fig. 4**).[39]

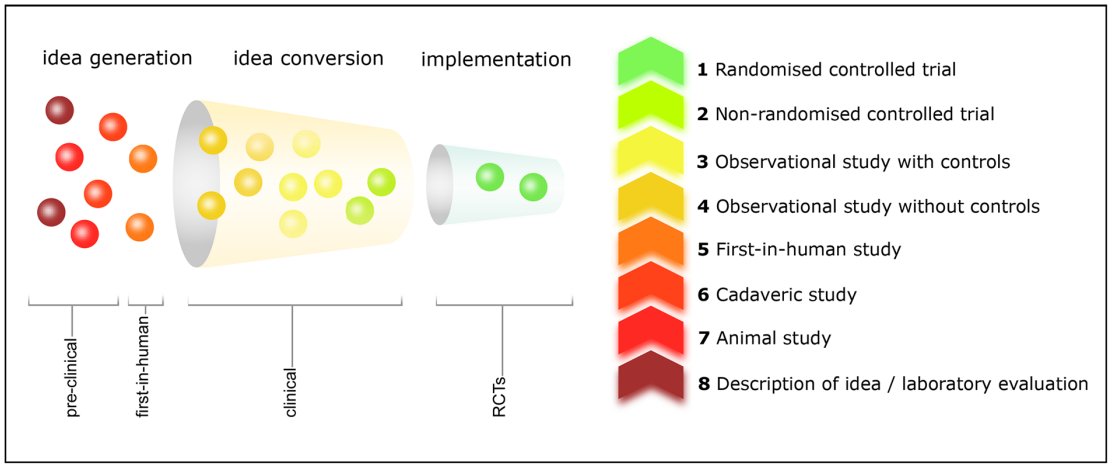


Figure 4 The surgical innovation funnel. Left-hand panel: the surgical innovation funnel illustrating the different stages through which innovation in surgery typically progresses. Right-hand panel: the different stages of innovation implementation according to the level of evidence.

2.4 The need for surgical innovation metrics

Having defined, classified, and described the stages of innovation, an important question ensues: how can such a complex, multi-faceted, and dynamic process be measured? Answering this question is critical for (i) informing healthcare policy and (ii) guiding clinical practice towards the strategic adoption of innovations with the highest value. At the same time, innovation metrics can also enable (iii) the early dismissal of low value pseudo-innovations before substantial financial, human, and social capital resources are further consumed. This strategic and timely disinvestment will likely reduce waste in healthcare thus optimising efficiency in terms of resource utilisation.[20,40]

In addition to optimising patient care, patient safety, and resource utilisation, innovation metrics can be valuable in (iv) strategically directing research funding through comparative analysis and rankings to uncover superiority and enable prioritisation. This is fundamental as innovation in surgery is entirely dependent on research. Surgical research whether involving basic science, engineering, clinical trials, or big data analytics is typically associated with high costs.[41] In the presence of finite resources and competing national interests for investment, the need for innovation metrics is evident.[42,43] Furthermore, as surgical research represents such a diverse range of themes and complex data types (that can be multi-dimensional and multi-relational), the introduction of innovation metrics would (v) offer a common platform to appraise and compare new surgical developments irrespective of their nature.

By looking outside healthcare, it becomes apparent that innovation metrics are extensively used. International corporations including innovation leaders such as Apple Inc. and Amazon.com, Inc. have been using smart innovation metrics for years, and heavily rely on these to systematically evaluate their innovation efforts.[7] As stated in the Introduction, it is through innovation metrics that strategy is devised towards enhancing innovation and driving performance.[8] Whilst most metrics used in these industries cannot be directly translated to healthcare (as the outcomes of interest differ), they nonetheless offer important lessons[7] that can be adopted to enhance medical and surgical practice.

There are two main categories of innovation metrics - those measuring 'activity' (input) and those measuring 'impact' (output).[8,44] For innovation to take place, the output must be shown to offer added or new value (**Fig. 5**). Common metrics used for measuring input in corporate innovation include expenditure towards Research & Development (R&D) and marketing, whilst for output, sales and profits are often used.[8] Another metric is Return On Investment (ROI), the ratio between output and input.[8]

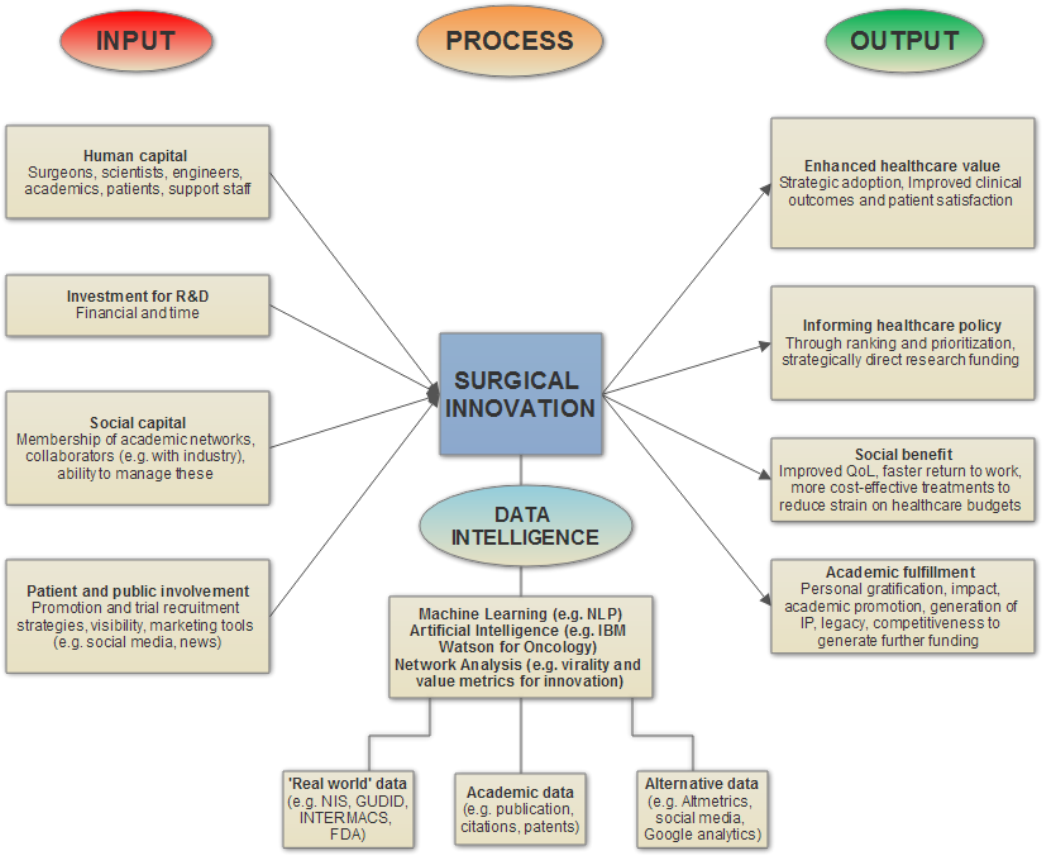


Figure 5 The innovation process in surgery. R&D = Research & Development; NLP = Natural Language Processing; NIS = National (Nationwide) Inpatient Sample database (HCUP); GUDID = Global Unique Device Identification Database (FDA); INTEMACS = Interagency Registry for Mechanically Assisted Circulatory Support; FDA = Food and Drug Administration.

Managing healthcare innovation is complex due to the volume and speed of new evidence generation and the wide variation of clinical practice, so that it is considered "the most complex and fast moving industry".[45] It is the unique Volume, Veracity, Velocity, and Variety of data evidence; collectively referred to as the 'four Vs' that has led to the title of 'big data'.[46,47] A key challenge relates to its fragmented nature involving a multitude of stakeholders (e.g., patients, surgeons, policy-makers, medical device manufacturers, health insurance companies), each with their own perspective of what constitutes value.[9]

The need for standardised metrics of global surgical surveillance has been widely emphasised,[48] and a framework for evaluating surgical innovation (IDEAL) developed.[49] However, despite its universal adoption underlining the importance of evaluating innovation in surgery ("no surgical innovation without evaluation"),[50] this framework is primarily qualitative in nature, lacking quantitative components. It is precisely in this area where innovation metrics can offer substantial value in objectively assessing new and current innovations.[31]

In the past, attempts have been made to use patent and publication counts to quantify innovations in surgery.[3] Though this data may be of interest, they both measure activity rather than impact, and more importantly rely on the assumption that the acquisition of Intellectual Property (IP) will eventually translate to new products and revenues.[51] However, this is seldom the case.[52]

The rise of big clinical databases offers a unique opportunity for measuring surgical innovation and conducting comparisons and rankings applicable to the real world.[53] Innovation metrics based on big data can provide a novel insight into surgical innovation and permit meaningful comparisons essential for both clinical practice and policymaking. Conventional statistical techniques can suffer from some limitations when handling some forms of big data. The 'four Vs' necessitate the use of substantially more powerful analytical tools to capture underlying complex, non-linear relationships.[54] Moreover, data derived from patient registries are often less structured and uncategorized since these are usually set up for different purposes (e.g., billing), and thus difficult to analyse unless thoroughly curated. This data processing is not only very expensive (owing to the expertise and time required) but, more

importantly, too slow when compared with the rate of production (deposition) of new data. This inevitably makes a curated big dataset likely to be already outdated before it is even ready for analysis.[54,55] This data complexity therefore requires the application of *Data Intelligence* – the utilisation of advanced analytical processes such as machine learning, network approaches and artificial intelligence to manage big data.

Machine Learning (ML) represents a collection of computational tools and techniques with the potential of overcoming these limitations. As an example, Natural Language Processing (NLP) permits data mining, i.e., the automatic categorisation and labeling of previously unstructured data based on content.[56] Another example is Deep Learning (DL), which by implementing algorithms based on a hierarchical multi-level learning approach to recognise patterns (in a fashion similar to the human brain),[57] offers researchers the possibility to extract meaningful abstract representations from heterogeneous, uncategorised big data.[55] Importantly, this can be performed in real time to provide continuously updated results based on the latest evidence and individual patient data.[57]

In addition to the aforementioned techniques falling under the category of 'unsupervised learning' (i.e., discovering structure in unlabeled data), 'supervised learning' (i.e., the training of the DL algorithm to make predictions from novel data based on existing data and outcomes) can prove essential for evaluating surgical innovation.[58] Again, looking outside healthcare, innovation leaders such as Google Inc. and Microsoft Corporation routinely use DL algorithms to analyse big data and facilitate decision-making based on predictive models.[55] By utilising this approach on

individual patient data, the promises of precision medicine can be achieved by selecting the best innovations applied to the populations to which they are most suited.

In healthcare, IBM has recently implemented AI to create IBM Watson for Oncology a DL predictive algorithm trained by Memorial Sloan Kettering Cancer Center (MSKCC) physicians to assist in the decision-making of the multi-disciplinary team (MDT).[59] IBM Watson for Oncology ranks treatment options and offers evidence-based, personalised advice by integrating and analysing data from a variety of sources including not only the literature, but also the individual patients' laboratory results and physicians' notes.[60] This can be achieved within seconds and shown to correlate by 90% with the actual recommendations of the MDT.[61] This exceptional ability of real-time, multi-dimensional analytics is likely to be of special salience when it comes to measuring surgical innovation.

To measure surgical innovation, modeling the underlying process is required. As the global spread of ideas, knowledge and innovation is a complex, network-driven dynamic process,[4] an effective way to model the footprints of innovation is through the use of network analysis.[62] Networks (e.g., citation and collaboration networks) enable the precise tracing of how all stakeholders (i.e., the nodes of the networks) involved in the innovation process forge, maintain, and sever connections (i.e., links between nodes) over time, and thus discover, exchange and apply new information.

A promising first step in this direction in described in detail in Chapter 3 with the development of the first two network-based, surgical innovation-specific metrics,

namely the *structural virality* and *innovation index* aimed at measuring the diffusion of surgical innovation and evidence-based innovation value, respectively.[39] As also demonstrated in Chapter 3, these indices have been validated against real world data (from the National Inpatient Sample-NIS database) showing a strong, statistically significant correlation.[39] Of note, no correlation was found when traditional metrics (including number of publications and citations) were studied alongside the same real world data. This further supports the need for the introduction and adoption of novel innovation-specific metrics rather than traditional metrics (that were not developed for directly assessing innovation).[39]

Novel innovation metrics can be applied to all relevant data available and should not be confined to Randomized Controlled Trials (RCTs). This is especially relevant to surgery where the conduction (and even analysis on certain occasions) of RCTs remains problematic, as is their practicality when it comes to assessing harm.[43,63] The latter is of particular importance as surgical innovation inherently involves risk.[13] Surgical innovation metrics should complement the traditional evidence base with real world evidence incorporating multiple additional sources including those derived from administrative healthcare databases.[64]

In addition to more traditional publication and patent data as well as financial data, recently emerging alternative metrics (altmetrics) are required for the evaluation of surgical innovation in an 'all-rounded' perspective.[65] In the modern digital world, these represent "indicators of different types of visibility" and offer the ability to measure both input/activity (in terms of online presence, social media promotion and digital marketing) and output/impact (in terms of references in news media, Internet

citations, search engine hits, downloads, and other metrics of impact relating to Patient and Public Involvement, PPI).[65]

It thus becomes apparent that SMART surgical innovation metrics are essential and can only be developed through data intelligence where big data from a number of diverse sources are collected, linked and analysed in real time using an array of advanced computational tools, including ML, neural networks and AI. It is important to appreciate that with these immense opportunities, new challenges emerge. Examples relate to data governance; in particular issues concerned with privacy, access, storage, database linkage, and online safety. Patient confidentiality and the application of cybersecurity principles are paramount and cannot be compromised. Technology to address these challenges is being developed in parallel. Examples include high security cloud storage (e.g. Iron Mountain®)[66] and distributed ledger technologies such as Blockchain, a digital network originally designed to ensure safety in online transactions and more recently introduced to healthcare by IBM to manage clinical trial data and electronic medical records while maintaining regulatory compliance.[67]

2.5 Conclusion

Data intelligence, enabling the automated collection, processing and analysis of large, complex (multi-dimensional and multi-relational) datasets to measure innovation remains largely underutilised in healthcare. Surgical innovation metrics will enable quantified benchmarking, comparisons, and rankings of innovations according to their individual value. This is fundamental for informing policy and guiding clinical

practice in view of the strategic adoption of innovations of highest value. The future in this field is to engage with healthcare leaders and policy makers to highlight the strengths and powerful advantages of introducing quantifiable innovation indices and to setup both national and international strategies with which to implement them.

2.6 Chapter Summary

Novel surgical innovations are being introduced at a rate and volume that surpasses our capacity to quantify them. The current system for assessing innovations is primarily qualitative so that there is confusion in the selection of new surgical devices. As a result, the selection of innovations is subject to exploitative marketing efforts, which can result in the diffusion of costlier technologies that lack evidence of added value over established practice. This chapter aimed to highlight why surgical innovation metrics are urgently needed and how data intelligence can facilitate their development.

Data intelligence enables the collection and analysis of big data using machine learning, artificial intelligence, and network analysis. When looking outside healthcare, it becomes apparent that innovation leaders across a number of fields in the corporate world have been heavily relying on SMART (Specific, Measurable, Achievable, Results-based, and Time-bound) metrics to systematically evaluate their innovation efforts. It is through metrics that strategy is directed towards enhancing innovation and driving performance. These metrics are lacking in surgery and there is a growing sense of urgency to develop rigorous surgical innovation metrics, crucial for optimising patient care.

There is an implicit need for quantifiable innovation metrics for surgical innovation. Current global healthcare pressures are straining budgets, which are further depleted by an increasing emphasis on expensive medical devices without clear evidence of added value. Unless surgical innovation metrics are introduced, there is a real danger that marketing factors and intuition alone rather than value will inform assessment. The relevant data is increasingly abundant and the necessary tools for analysis are available and waiting. Innovation metrics will enable benchmarking, comparisons, and rankings of innovations by value. This is fundamental for informing policy, regulation, and guiding clinical practice towards the strategic adoption of innovations of highest value.

The next chapter, Chapter 3, presents in detail the development and validation of the first two network-based, surgical innovation-specific metrics, namely the *structural virality* and *innovation index* aimed at measuring the diffusion of surgical innovation and evidence-based innovation value, respectively.

3. Network analysis of surgical innovation: The development and validation of novel surgical innovation metrics

3.1 Introduction

Innovation has long occupied centre stage in the medical and health sciences. In surgery, it has been held up as a catalyst of unprecedented advances that have led to substantial improvements in healthcare delivery and patient outcomes.[1] A variety of surgery-specific innovation frameworks have been proposed, among which the IDEAL is the most widely implemented paradigm that categorises surgical innovation into distinct stages.[68] Most existing frameworks, however, suffer from a unifying limitation. Their qualitative nature has thwarted comparative assessments of the innovation-generating potential of individual scientists, research groups, institutions, and medical specialties. Moreover, earlier attempts to assess surgical innovation solely through publication and citation counts were premised on the oversimplifying assumption of equating innovation value with short-lived surges in popularity, and failed to capture the long-term impact of innovation upon healthcare delivery.[3]

To address these shortcomings, we propose a novel and quantitative network-based framework for measuring the value of surgical innovation. To this end, we leverage on the structure of the adoption-induced fingerprints produced by diffusion processes as they unfold over time.[69] Indeed, innovation typically triggers complex diffusion processes, driven by social contagion mechanisms, in which individuals' adoption is a

function of their exposure to others' knowledge, attitude, or behaviour. Diffusion can therefore be mapped out as a time-varying cascade of adoptions that propagate from individual to individual over potentially many generations of adopters. Any attempt to capture how increasingly popular innovations ultimately transform medical care ought to explicitly account for the size, structural depth, and breadth of the whole adoption cascade underpinning diffusion.[70]

This chapter aims to make a contribution in this direction by using citation networks to study the structural foundations of innovation diffusion in surgery. Drawing on a unique and comprehensive dataset on robotic surgery, we propose a novel set of network-based measures for uncovering the virality of adoption cascades. We demonstrate how these network measures facilitate comparative assessments of different robotic surgical procedures in terms of how they diffuse and implement innovation. Our framework can therefore play a fundamental role in guiding and assisting policymakers, funding bodies, and healthcare providers.

3.2 Materials and methods

3.2.1 The dataset

Drawing upon SciVerse Scopus® (Elsevier®, Amsterdam, The Netherlands),[71] another researcher and I independently extracted all articles concerned with robotics (i.e., containing the MeSH terms 'robot' and 'robotic' in their title, abstract, or keywords) across all medical specialties from the start of the database (1974) until December 2014. We limited the dataset to articles published in scientific journals, and restricted the scope of the analysis to the subject area 'Medicine'. Therefore, any article that did

not represent original research studies on robotic surgery was excluded (e.g., review article or conference article).

The initial search produced 13,240 publications, of which 9,423 were articles that either received or made at least one citation within the broad medical field of robotics. Among these, 5,961 articles received at least one citation from another article within the field, and can therefore be regarded as 'sources' of innovation, whereas 8,158 articles made at least one citation to another article within the field, and can therefore be regarded as 'adopters' of innovation (**Fig. 6**).

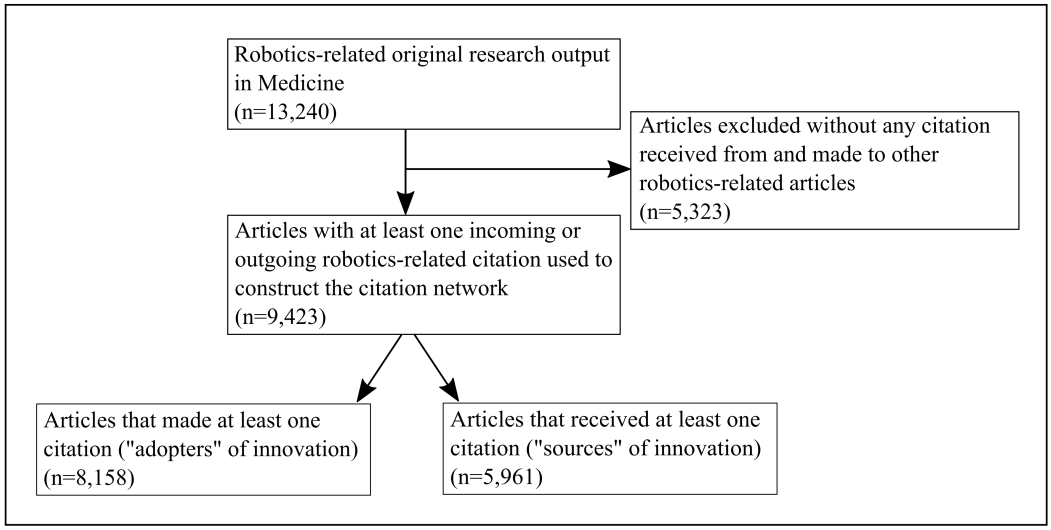


Figure 6 Flow diagram illustrating the search strategy used for generating the citation networks. Note that the intersection between the two sets of articles at the bottom used to construct the citation networks is non-empty as an article can both cite and be cited.

Using the MeSH terms reported in Table S1 (Appendix 1), each of the 5,961 articles that received at least one citation was allocated to one of 16 categories (based on surgical specialty and/or procedure). We limited our study only to the ten specialties (the terms specialties and procedures are used interchangeably) with at least 100 of

the original 13,240 publications. In total, the final dataset includes 4,860 articles. Among these, there are 2,101 articles across the ten specialties that received at least one citation, and 2,759 articles that do not necessarily belong to any of the ten specialties, but are part of chains of citations leading to articles in those specialties (**Fig. 7**, and Appendix 1).

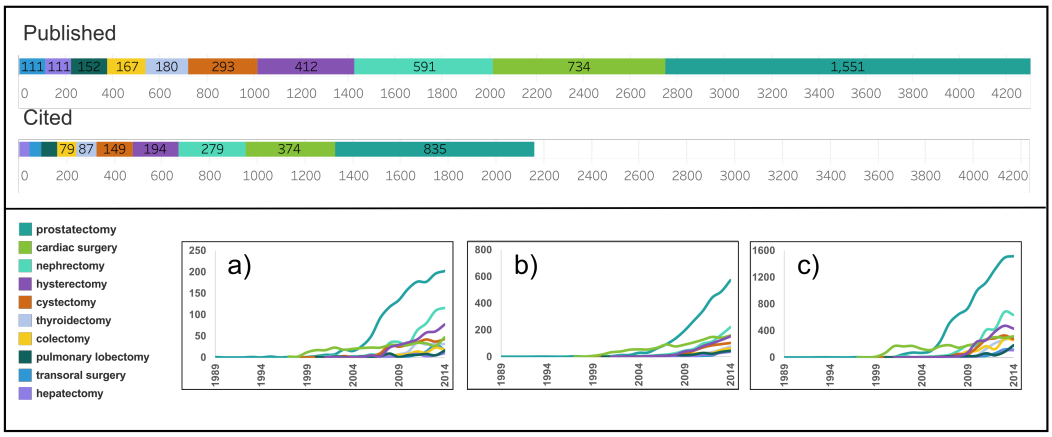


Figure 7 Scientific production and citations across surgical specialties. Published and cited articles per surgical specialty (top panel). Cumulative number of published and cited articles: (a) number of published articles over time, (b) number of cited articles over time, (c) number of citations received over time by published articles.

3.2.2 Citation networks and diffusion cascades

In a citation network, the nodes are the articles, and a directed link is established from one article to another if the former cites the latter in its bibliography.[43,72] A citation network can therefore be thought of as a diffusion cascade along which information spreads and adoption of innovation propagates.

For each of the 9,423 articles that received or made at least one citation, we constructed the diffusion cascade based on the corresponding citation network (**Fig. 8**).[73,74] In these cascades, each citation-based chain of adoption can be traced back to

the seed node representing the original article in which a given innovation was introduced (as an idea description or laboratory evaluation) in the first place. The structure of these cascades can therefore shed light on the depth and breadth of the diffusion process through which innovations, once proposed, built up momentum over time.[35]

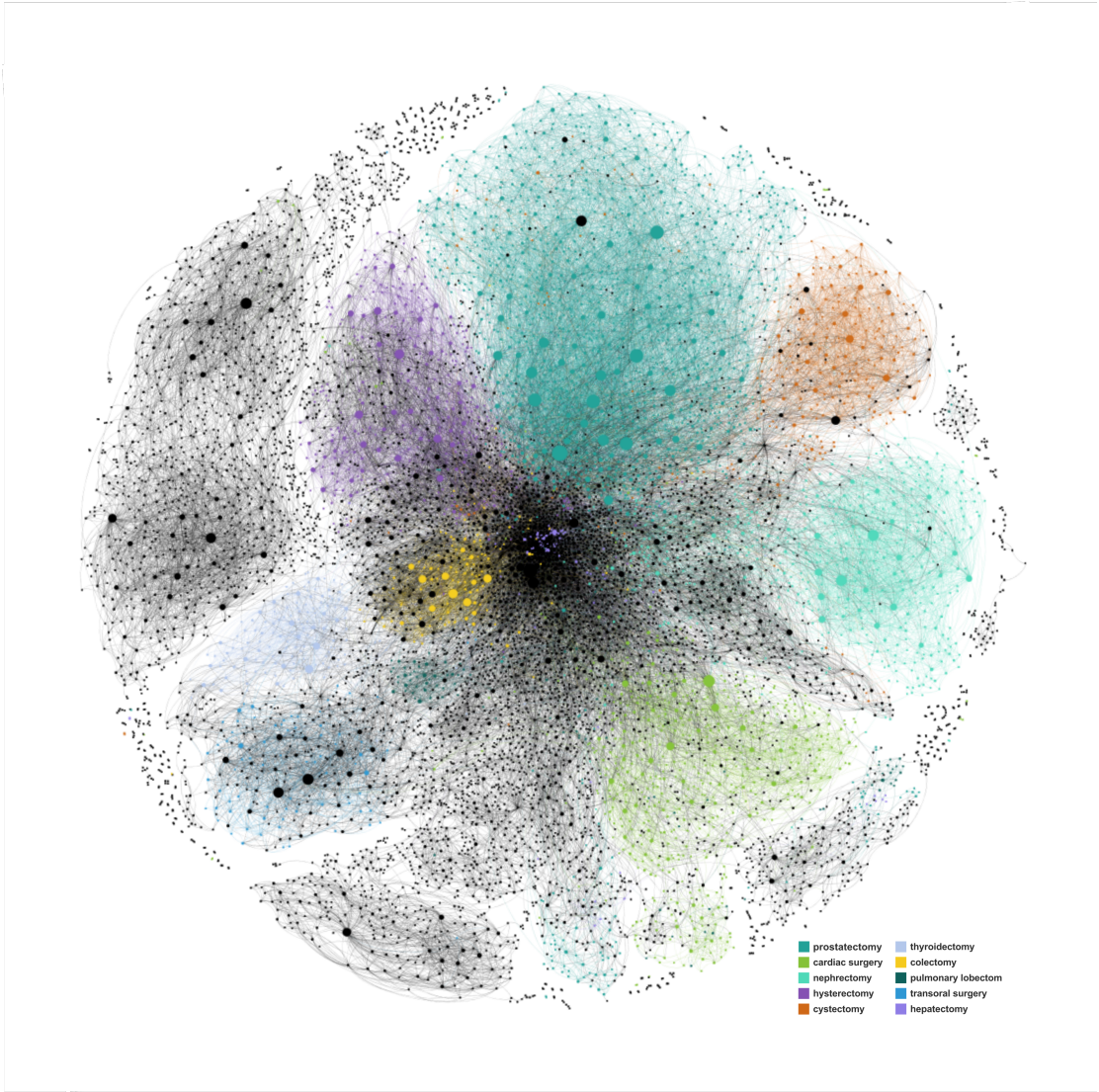


Figure 8 Network of citations. The network includes 9,423 articles that received or made at least one citation. The visualisation of the network was obtained through the network visualisation software Gephi© (Gephi© Consortium, Compiegne, France). The size of each node i is proportional to its in-degree k_i^{in} (i.e., the number of citations received), and the colour associated with each node denotes the surgical specialty it belongs to. The network is partitioned into topological communities that are coextensive with surgical specialties (Appendix 1).

3.2.3 Measuring broadcast and viral diffusion processes: Citations, cascade size, structural depth, and structural width

Popularity can be gained through two main modes of diffusion: broadcast and viral spreading (**Fig. 9, panels a and b**). While broadcast spreading is dominated by processes of bursty adoptions from a single seed article, viral spreading is typically characterised by multi-generational branching processes in which any article receives a citation from only few others, thus yielding multiple, widespread, and long chains of citations extending over time.[4]

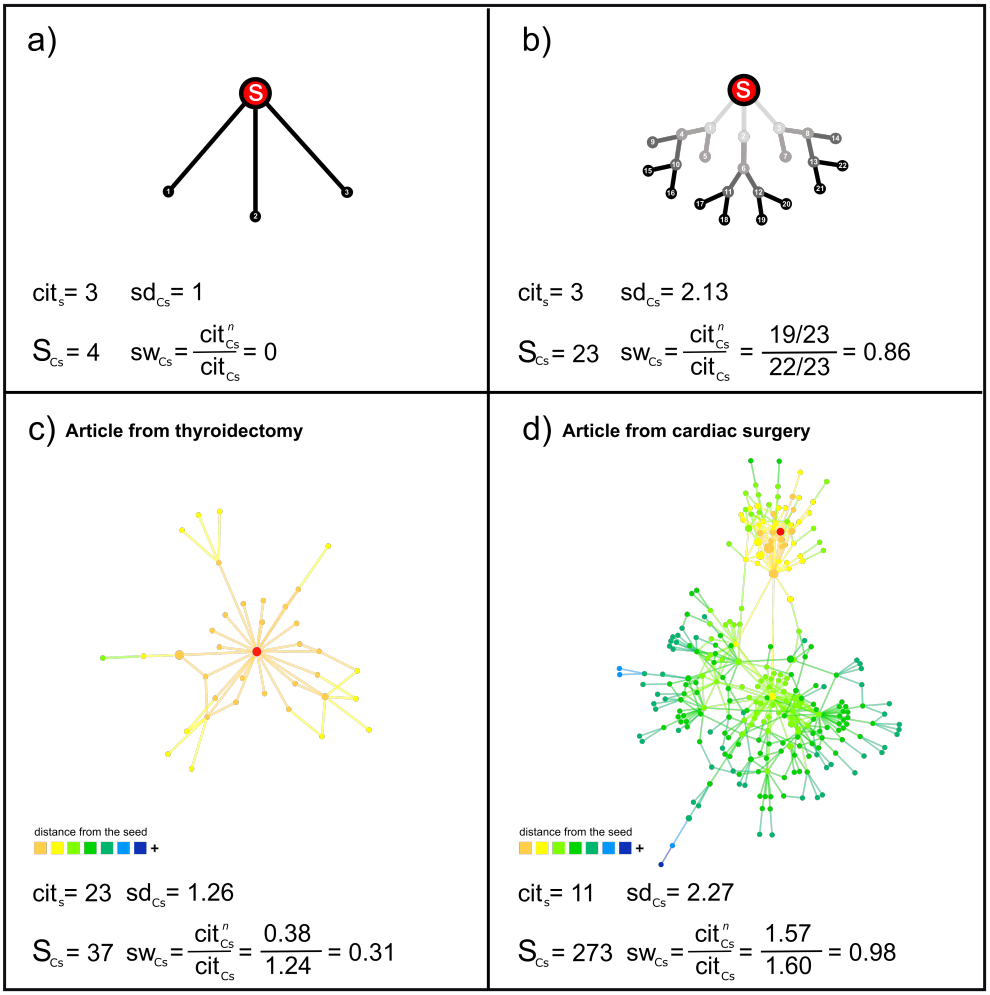


Figure 9 Schematic representation of cascades and real world examples. Top panels: (a) broadcast-driven diffusion; (b) viral diffusion. In both cases, we calculated the number of citations received by the seed node, the size of the cascade, the structural depth, and the structural width of the cascade. Bottom panels: two real examples of cascades within (c) thyroidectomy and (d) cardiac surgery. For each cascade, we calculated the number of citations to seed article, the cascade size, and structural depth and width. The colour of nodes denotes their distances from the seed article (red node). Links between nodes carry the colour of the citing node. The size of each node is proportional to the ratio between the node's second-step citations and first-step citations.

To investigate broadcast-driven diffusion processes, two measures were quantified. First, for each seed article s that received at least one citation, we calculated the total number of citations Cit_s received from any other article in the corresponding cascade C_s. Second, for each seed article s, we calculated the fraction of citations $Cit_{s,year}$ received from any other article in C_s within one year since the date of publication of s.

To uncover the virality of diffusion processes, we computed three measures (Appendix 1). First, for each seed article s, we calculated the size S_{C_s} of the corresponding cascade C_S, namely the total number of articles in C_S (including seed article s). Second, drawing on a classical graph property[75] and recent work on online diffusion,[75] we propose a measure for quantifying the multi-generational nature of adoption cascades. Specifically, for each seed article s we define the structural depth sd_{C_s} of cascade C_S, as the average length of the shortest directed paths between pairs of articles in C_s. Formally, for $n>1$ articles,

$$sd_{C_s} = \frac{1}{l}\sum_{i\in C_s}\sum_{j\in C_s} d_{\overrightarrow{i,j}}, i \neq j,$$

where $d_{\overrightarrow{i,j}}$ denotes the length of the directed shortest path from article i to article j, C_s is the set of nodes belonging to the cascade originating from article s, and l is the number of directed paths connecting pairs of articles in C_s. So constructed,

sd_{C_s} becomes larger as the adopters of the original idea proposed by seed article *s* are farther apart from one another and from *s*, thus producing a multi-generational cascade. An important property of structural depth is that it varies as a function of the size of the cascade only under certain conditions (Appendix 1).

Finally, we introduce a measure for quantifying the branching structure of a diffusion cascade.[70,72,74,76] Specifically, we define the structural width sw_{C_s} of cascade C_s as the ratio between the average number of second-step citations ($cit^n_{C_s}$) and the average number of citations (cit_{C_s}) accrued by articles in C_s. Formally, for $n>1$ articles,

$$sw_{C_s} = \frac{cit^n_{C_s}}{cit_{C_s}}.$$

Thus, the larger sw_{C_s}, the more widespread and locally dense the diffusion cascade is. In particular, sw_{C_s} ranges between zero in the limiting case of the star graph (with no second neighbours) and indefinitely large values in cases of branching cascades with highly widespread tree-like local structures.

Combined, cascade size, depth and width enable us to capture the multi-faceted nature of viral innovation.[72,75,76] While cascade size captures the *overall popularity* of innovation based on total number of adopters, structural depth and width shed light on how popularity is gained. Specifically, depth captures the *multi-generational character* of diffusion processes. In this sense, innovation becomes viral not simply because it is widely adopted, but because it propagates further beyond the first generation of initial adopters, typically targeted by extensive media coverage and large advertising efforts. Finally, structural width uncovers the *branching tree-like character* of diffusion. In this sense, innovation becomes viral when a large

population of adopters 'infect' a much larger population, with ripple effects producing a rapid, large scale increase in popularity as in viral disease spreading. Thus, these three measures jointly quantify virality as a function of: (a) the overall number of adopters; (b) the number of generations of adopters; and (c) the contribution of each adopter to overall diffusion. This enables us to create ranking lists from the most viral innovations to those that only generate short-term surges in popularity and then quickly die out. **Fig. 9 (panels c and d)** shows two real cascades that differ in structural virality.

3.2.4 Measuring the value of surgical innovation: The implementation-based innovation index

We propose a novel metric aimed at capturing the intrinsic value of a surgical innovation as a function of the degree to which it has reached an implementation stage. To this end, a score was attributed to each seed article based on its corresponding level of evidence (Appendix 1).[64] The US Department of Health and Human Services evidence levels were used as they only include numerical values (with no lettered subcategories) facilitating classification.[64] Further stages of implementation were added for pre-clinical categories (description of idea/laboratory evaluation, animal, and cadaveric studies), as illustrated in the surgical innovation funnel (SIF) in **Fig. 4** (Chapter 2). The SIF shows that innovation follows a trajectory akin to natural selection whereby as "the going gets tough, the tough get going" (i.e., only the fittest among attention-seeking ideas will survive as they move along the SIF).[77,78] All studies were scored by two independent academic surgeons (G.G. and T.A.), and disagreements arbitrated by a third academic surgeon (A.D.). Formally, for each surgical specialty $g \in G$, we define the innovation index as

$$i_g = \frac{\sum_{c \in S} \frac{1}{c} \times \frac{p_c^g}{P_c}}{\max_{g \in G} \left(\sum_{c \in S} \frac{1}{c} \times \frac{p_c^g}{P_c} \right)} \times 100,$$

where $c \in [1, \ldots, 8]$ is an integer value labeling the ordinal category associated with the implementation stage according to level of evidence, p_c^g is the count of publications by surgical specialty g appearing in category c, and $P_c = \sum_g p_c^g$ is the total number of publications appearing in category c across all surgical specialties in G (Appendix 1).

To validate our innovation index, additional data from the National (Nationwide) Inpatient Sample (NIS) database were collected and ranked in terms of the actual numbers of robotic procedures performed in the US in 2012 (most recent publically available database at the time of analysis). As NIS approximates a 20% stratified sample of all discharges from US community hospitals containing data from more than seven million hospital stays per year, it can reasonably be used to provide empirical support in favour of findings based on our innovation index (Appendix 1).[79]

3.3 Results

3.3.1 Ranking surgical innovations by broadcast-driven popularity

Fig. 10 shows the rankings of robotic surgical procedures in terms of the medians of the distributions of: (i) citations received by seed articles (panel a); and (ii) fractions of citations received by seed articles within one year since publication (panel b). Results from Mann-Whitney U (for citations) and Kolmogorov-Smirnov (for fraction

of citations) tests for comparing pairs of such distributions are not statistically significantly different.

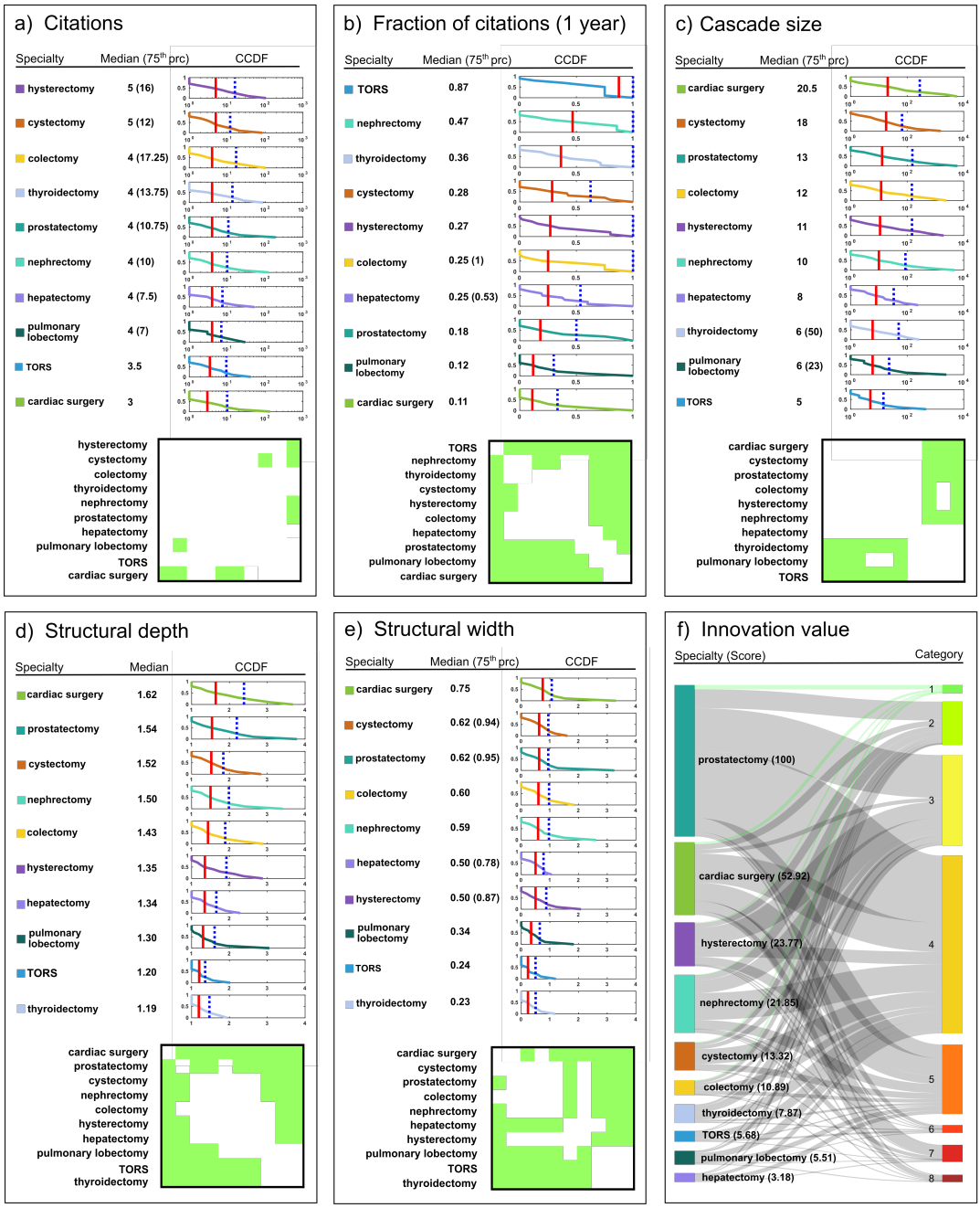


Figure 10 Rankings of surgical specialties. (a) Ranking by citations to seed articles; (b) ranking by fraction of citations to seed articles within one year; (c) ranking by cascade size; (d) ranking by structural depth; (e) ranking by structural width. Rankings were obtained by using the medians (red lines) of the complementary cumulative distribution functions (CCDFs). In case of ties between medians, the 75th percentiles (dotted blue lines) were used. The bottom of each panel shows results from Mann-Whitney U tests (for citations and size) and Kolmogorov-Smirnov tests (for fraction of citations, depth, and width) of independence between pairs of distributions (green color: $p<0.05$). (f)

Sankey diagram representing the ranking of surgical specialties in terms of innovation index. Specialties are listed in the left column, with the top-ranked scoring the highest by innovation index. The height of coloured bars is proportional to the number of published articles in the corresponding specialty. Levels of evidence are listed in the right column, with the top-ranked being the final implementation stage. The height of coloured bars is proportional to the number of published articles at the corresponding level of evidence across all specialties. The width of the lines connecting a given specialty to a given level of evidence is proportional to the number of articles published by the specialty at that level of evidence. Green lines refer to contributions of specialties to the final stage of implementation.

3.3.2 Ranking surgical innovations by virality

For each robotic surgical procedure, we measured the size, structural depth and width of each cascade, and produced the corresponding frequency distributions. We then ranked these in terms of the medians of such distributions (**Fig. 10, panels c, d, e**). Cardiac surgery occupies the top of the rankings, followed by urological procedures, specifically prostatectomy, cystectomy, and nephrectomy. Mann-Whitney U test (for size) and Kolmogorov-Smirnov tests (for depth and width) were used for comparing each pair of distributions (Appendix 1). Cascade size, structural depth, and width of cardiac surgery are statistically significantly different from those of all other specialties, as occurs with second-ranking prostatectomy (except when compared to colectomy). At the other end of the ranking list, thyroidectomy occupies the last position, overtaken by transoral robotic surgery (TORS) though the difference between the two does not reach statistical significance ($p>0{\cdot}05$). Notice that differences between any of the three highest-ranking specialties and any of the three lowest-ranking ones are all statistically significant ($p<0{\cdot}05$).

3.3.3 Ranking surgical innovations by evidence-based innovation value

Fig. 10 (panel f) reports a Sankey diagram illustrating the contribution of each surgical specialty/procedure to each level of evidence, and the ranking by innovation index (Appendix 1). Results suggest that prostatectomy ranks first with the greatest

number of Randomized Controlled Trials (RCTs), followed by cardiac surgery, hysterectomy, nephrectomy and cystectomy. At the bottom of the ranking list are thyroidectomy, TORS, pulmonary lobectomy, and hepatectomy. To validate our measure of innovation, we compared the ranking of innovation across surgical specialties by innovation index with the ranking based on real world evidence from the NIS database, and found a statistically significant correlation (Spearman's rank correlation coefficient = 0·673, p=0·033; Kendall's tau coefficient = 0·511, p=0·04).

3.3.4 Comparisons across surgical innovations' rankings

The ranking of robotic surgical procedures by innovation index closely matches the rankings by cascade size, structural depth, and width. We tested the similarity between these three pairs of rankings using the Spearman's rank correlation coefficient (respectively, 0·758 (p=0·011), 0·782 (p=0·008), 0·624 (p = 0·054)). Interestingly, no statistically significant correlation (p>0·05) was found between ranking by either citations or fraction of citations received within one year, on the one hand, and ranking by either cascade size, structural depth or width, on the other (Appendix 1). Most importantly, there is no statistically significant correlation (p>0·05) between the ranking by either citations or fraction of citations within one year, on the one hand, and the ranking by innovation index on the other (Appendix 1). We also tested the similarity between all pairs of distributions by using the Kendall's tau coefficient, and obtained similar results (Appendix 1).

3.4 Conclusion

Despite the prominent role of innovation in surgery, only limited attention has been paid to quantifying its value as a function of diffusion processes and implementation.

The study described here has made a three-fold contribution in this direction. First, we proposed a novel measure for the value of surgical innovation that directly accounts for the evidence-based implementation stage reached in clinical practice. The ranking of surgical specialties by this measure was found to closely match the one based on real word data (i.e., the NIS database). For example, it is no coincidence that robotic prostatectomy and cardiac surgery, widely supported by RCTs, have the highest score in the ranking, while robotic thyroidectomy, characterised by a remarkably poor uptake in the Western world, occupies the lowest position.[80] Second, we introduced a novel network-based framework for assessing the structure of adoption cascades. Specifically, we proposed measures for quantifying the virality of these cascades. Third, we demonstrated that the ranking of robotic surgical procedures by innovation value positively correlates with rankings by virality, but not with rankings by broadcast-driven popularity.

We focused on robotic surgery for a number of reasons. First, it is a sufficiently recent innovative technology so as to play a salient role in most surgeons' work, at least in the Western world. Second, it also boasts a sufficiently long history so as to enable the tracing of adoption cascades over time. Moreover, as the robotic surgery market is dominated by the da Vinci® surgical robot (Intuitive Surgical®, Inc., Sunnyvale, CA), it lies at the interface between different surgical specialties, most of which use different versions of the same robot (e.g., standard, S, Si, Xi, SP), thus providing a 'common comparator' when it comes to ranking (the same) innovation as applied to different surgical specialties.[81]

Our results can have a number of implications for research, clinical practice, and policy. First, by using citation networks to uncover the footprints of adoption behaviour, our study opens up new avenues for future work on knowledge transfer and sharing, and the way healthcare providers discover, combine, and apply new information over time. Second, our study complements and extends existing frameworks for surgical innovation, and enables suitable combinations of quantitative and qualitative assessments of innovation value. Third, our framework can be easily extended to also produce rankings of institutions, research centers, academic surgical groups, and even individual surgeons in terms of their ability and potential to produce pioneering innovation. From this perspective, our network-based framework can play a fundamental role in guiding policy, strategically directing medical research funding, and assisting healthcare providers in their efforts to optimise resource allocation and improve the quality of healthcare delivery and patient outcomes.[82] For example, our findings suggest that recently introduced innovations (e.g., percutaneous valve implantation technology), with only limited adoption history and yet highly viral diffusion cascades, have the potential of redirecting clinical practice, and should therefore be sustained by policymakers.

Most importantly, our findings on short-term broadcast-driven popularity should alert healthcare providers and policy-makers to the dangers of using mere citation counts as predictors of value-generating potential. Seemingly successful research efforts, with a disproportionally large number of citations, may quickly die out leaving no influential trace over time. By contrast, research that is only mildly successful in the short run may well gain in popularity over time, yield wide and multi-generational

cascades of adoptions, and ultimately provide the foundation for successful medical practice.

A number of limitations in our study should be noted. First, citation networks represent but one type of innovation networks. Other networks include co-authorship and collaboration networks where nodes represent individual surgeons or institutions. Drawing on these networks, one could investigate the structural foundations of innovation at the individual level, and the relationship between scientific collaboration and innovation diffusion. Second, our measures do not directly account for variations in time scales across cascades. Innovations may trigger cascades of adoption stretching over various intervals of time, but nonetheless characterised by the same degree of virality. Future work shall extend our measures to also account for temporal variations in adoption behavior. Finally, a broader family of network metrics, including degree correlations among adjacent nodes, will be required to fully assess the structure of adoption cascades and the multi-faceted value of surgical innovation.

In conclusion, our study has demonstrated that network analysis offers unique new opportunities for understanding, modeling and measuring surgical innovation, and ultimately for assessing and comparing the generative value of different specialties. When evidence-based data are difficult to collect or not yet available, a suitable methodological substitute is needed to inspire and guide the decisions of policymakers, funding bodies, physicians, and healthcare providers. Our study is an important first step in this direction.

3.5 Chapter summary

The qualitative nature of existing approaches to the study of surgical innovation has thwarted accurate evaluations within and across specialties. Earlier studies based on publication and citation counts failed to offer measures predicated on the role innovation plays in advancing healthcare practice. A more rigorous framework is needed in which an implementation-based value of innovation can be uncovered based on the structure of the entire diffusion processes triggered by adoption behaviour.

Our study draws on a comprehensive dataset including all articles concerned with robotic surgery (1974 – 2014). For each of the 9,423 articles that received or made at least one citation, we constructed the citation network underlying the time-varying cascade of adoptions. We assessed the virality of each cascade by measuring its size, structural depth and width. Each cascade was then associated with an innovation value reflecting the stage occupied by the corresponding seed article in the surgical innovation funnel.

Rankings of surgical specialties by cascade size, structural depth and width were found to correlate closely with the ranking by innovation value (Spearman's rank correlation coefficient = 0·758 (p=0·01), 0·782 (p=0·008), 0·624 (p = 0·05), respectively) which in turn matches the ranking based on real world data from the National Inpatient Sample (Spearman's coefficient = 0·673; p=0·033). No statistically significant correlation was found between ranking of specialties by

innovation value and rankings by mere citation counts and fraction of citations received within one year since publication.

The virality of adoption cascades is the structural signature of truly innovative procedures, while conventional measures based on short-term popularity are poor predictors of innovation value. Our study can assist policymakers and funding bodies in their efforts to optimise resource allocation and improve healthcare delivery.

The next chapter, Chapter 4, presents a different type of surgical innovation networks, namely collaboration networks. By building and analysing the real world collaboration network in robotic surgical research, the aim is to demonstrate how this can be used in practice to nurture collaborations that foster surgical innovation.

4. Surgical innovation in the era of global surgery: Network analysis of real world global research collaboratives

4.1 Introduction

As a result of recent technological and computational advances and the exponential rate at which new knowledge is generated, the 'lone innovator' model no longer represents a suitable paradigm for scientific production.[83] In the modern world, innovation is increasingly the outcome of a collaborative process.[84] This is especially the case for surgical innovation that has been unfolding at a global scale, within a complex network of international collaborations.[85-87]

The aim of collaborative efforts is the generation of social capital. The definition and generative mechanisms of social capital have long been the subject of debates and controversies within the social sciences.[88,89] Typically, scholars tend to converge on the idea that social capital refers to the value that individuals, groups, or organisations can derive from the underlying social relations.[90] In the context of surgical research, social capital can relate to accessing data, expertise, knowledge, or any other type of resources that become available through specific collaborative patterns and facilitate knowledge creation and innovation.

Surgical innovation networks are complex systems, typically consisting of hundreds or thousands of organisations dispersed across the globe forging various types of relationships with one another.[87] Network science offers a theoretical and methodological backdrop that has recently been widely used to study collaboration and innovation in a variety of fields. However, a network approach to examining surgical innovation has been largely neglected.[31,91]

The study in this chapter applies network analysis to a real world global collaboration network in robotic surgery. The aim is to familiarise surgeons with network analysis (of collaboration networks this time) and demonstrate how this approach can be used to devise effective strategies towards the establishment of partnerships that can enhance research impact, facilitate innovation and advance patient care.

4.2 Materials and methods

4.2.1 The dataset

This study draws on the *Web of Science*® (WOS) platform (Clarivate Analytics, Philadelphia, PA). All articles on robotic surgery were extracted through the use of the MeSH terms: *"robot OR robotic OR robot assisted OR robotic assisted OR robotically assisted OR robot-assisted OR robotic-assisted OR robotically-assisted"*. The *Research Area* was confined to *"Surgery"* and the *Document Types* to *"Article"*.

The search was performed on 17th of January 2017, and produced 3,889 publications (peer-reviewed articles) published between July 1988 and January 2017 (Fig. S1,

Appendix 2). These were all used for constructing the co-authorship network. Articles were generated from 1,700 organisations nested within 62 countries, in turn nested within 6 geographical regions (Appendix 2).

4.2.2 The surgical collaboration network

The collaboration network was constructed using VOSviewer© (Leiden University, Leiden, the Netherlands), a software developed specifically for the study of scientific collaboration networks. In the network, each node represents an organisation, and a link between two nodes represents collaboration between the corresponding organisations. As co-authorship has been shown to be a good proxy of collaboration,[92,93] links between nodes were based on co-authorship of articles. The resulting network is weighed: the value (or weight) of a collaborative link increases as a function of the intensity of collaboration (Fig. S2, Appendix 2).

4.2.3 Outcome measures

For each organisation, two outcome measures were computed: (i) the research impact; and (ii) the innovation index. These measures were computed as follows:

- Research impact: Research impact was measured as the sum of normalised citations received by all articles (co-)authored by scholars affiliated with a given organisation in each year. To obtain normalised citations, the citation count for each publication in a given year was divided by the average number of citations obtained by all articles published in the same year. The greater the sum of normalised citations for a given organisation, the greater the

organisation's research impact (i.e., normalised citations are used to measure impact).

- Innovation index: The innovation index represents a recently validated metric used to evaluate and rank surgical innovation (see Chapter 3 and Table S2, Appendix 2).[39] It captures the value of the innovative output produced by an organisation as a function of the degree to which it reached an implementation stage (Figs. S3-S4, Appendix 2). Thus, the greater the innovation index of a given organisation, the more innovative the organisation's surgical research output.

  For a detailed description of how each performance metric was calculated, please see Chapter 3 and sections S.3.1 and S.3.2 in Appendix 2.

4.2.4 Network measures

Two established network measures were computed: (i) the clustering coefficient; and (ii) closeness centrality. These measures were defined as follows:

- Local clustering coefficient: The local clustering coefficient quantifies how closed an organisation's ego-centered network is (i.e., the network including connections between the organisation and its partners as well as connections between these partners) enabling assessment of the extent to which an organisation's collaborators also collaborate with each other or, alternatively, the extent to which an organisation spans structural holes separating collaborators. The higher the local clustering coefficient of an organisation,

the more closed the organisation's ego-centered network is. More specifically, the local clustering coefficient of an organisation was defined as the ratio between the number of actual triangles containing the organisation and its neighbours, and the maximum possible number of such triangles. A generalised weighted clustering coefficient was calculated to take into account the weights of links (see Appendix 2). To facilitate interpretation, all values of the generalised weighted clustering coefficient were standardised.[94] Measuring the density of triangles in an organisation's local network uncovers how open or closed the network is, and the extent to which the organisation acts as the knowledge broker between otherwise disconnected organisations in the collaboration network.

- Closeness centrality: The closeness centrality of an organisation measures how close the organisation is to all other organisations in the collaboration network. The higher an organisation's closeness centrality, the greater the organisation's access to the knowledge (or data, or any other resource) provided by other organisations in the collaboration network, and thus the greater the organisation's influence on others as a result of its structural position.[95] The generalised weighted version of closeness centrality was used to account for the weights of links (see Appendix 2).[96]

For a detailed description of how each network metric was calculated, please see sections S.3.3 and S.3.4 in Appendix 2.

4.2.5 Measuring the geographical dispersion of collaborations

For each organisation, the *geographical entropy* of collaborations was computed to capture the geographical dispersion of the organisation's collaborators. An organisation's geographical entropy increases as the organisation collaborates with other organisations located in more countries and devotes an equal amount of collaborative effort towards each of these countries (Appendix 2). For a detailed description of how geographical entropy was measured, please see section S.3.5 in Appendix 2.

4.2.6 Measuring academic-industry collaborations

All organisations that were publicly registered as companies and classified as "corporate" entities in the WOS platform (through the InCites© intelligence tool) were identified.[97,98] For each organisation, the sum of the organisation's collaborative efforts towards other industrial (corporate) partners during the study period were calculated. For a detailed description of how the strength of industrial collaborations was measured, please see section S.3.6 in Appendix 2.

4.2.7 Control variables

Many other organisational characteristics may influence scientific performance. This study controlled for the following two additional variables: (i) each organisation's institutional type (e.g., academic, corporate, health, etc. as classified by the InCites© intelligence tool in the WOS platform) and (ii) a measure of volume, here referred to as 'number of articles in WOS', given by the number of all articles, beyond robotic surgery, published by each organisation that the InCites© intelligence tool could retrieve in the WOS database (see section S.3.7 in Appendix 2). Controlling for

research volume enables the association between collaboration network and both citation count and innovation to be investigated by keeping the number of publications constant.

4.2.8 Statistical analysis

Maximum-likelihood estimates of two hierarchical three-level random-intercept models were computed in which organisations were nested within countries, in turn nested within geographical regions. A random-intercept structure was combined with heteroskedastic level-1 residuals by letting the variances of these residuals be a function of the organisation's institutional type. All models were estimated using STATA 15 (StataCorp LP, College Station, TX, USA). The significance threshold was set at $p < 0.05$.

4.3 Results

Of the 1,700 organisations, 1,543 were connected through at least one collaboration. The overall network comprised of 6,000 collaborative links. **Fig. 11** shows two extreme examples of ego-centered network, one characterised by a closed structure ($cc_{org} = 1$) rich in third-party relationships (**Fig. 11A**), and the other by an open structure ($cc_{org} = 0$), rich in brokerage opportunities (**Fig. 11B**).

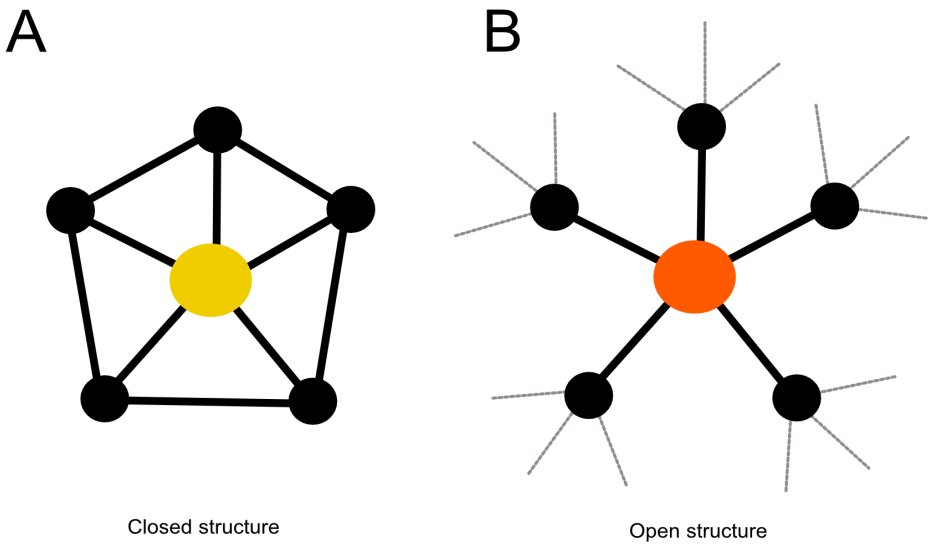


Figure 11 Illustrative examples of the two extreme cases of a closed ego-centered network (panel A) and an open ego-centered network (panel B). The closed network is rich in third-party relationships and closed triangles: all nodes connected to ego (yellow node) are also connected with each other. The open network is rich in brokerage opportunities and open triads: ego (orange node) acts as the broker between all contacts that would otherwise be unable to reach one another. Ego in the closed network has therefore a clustering coefficient equal to one, while ego in the open network a clustering coefficient equal to zero.

The collaboration network among organisations is shown in **Fig. 12**. While the highest-performing organisations (e.g., University of Pittsburgh and Yonsei University) achieved the largest research impact (node size) and innovation value (node colour), **Fig. 12A** suggests that the correlation between the two performance measures is far from perfect. Successful organisations (**Fig. 12C**) appear to be better connected than less successful ones (**Fig. 12B**), which are more sparsely connected.

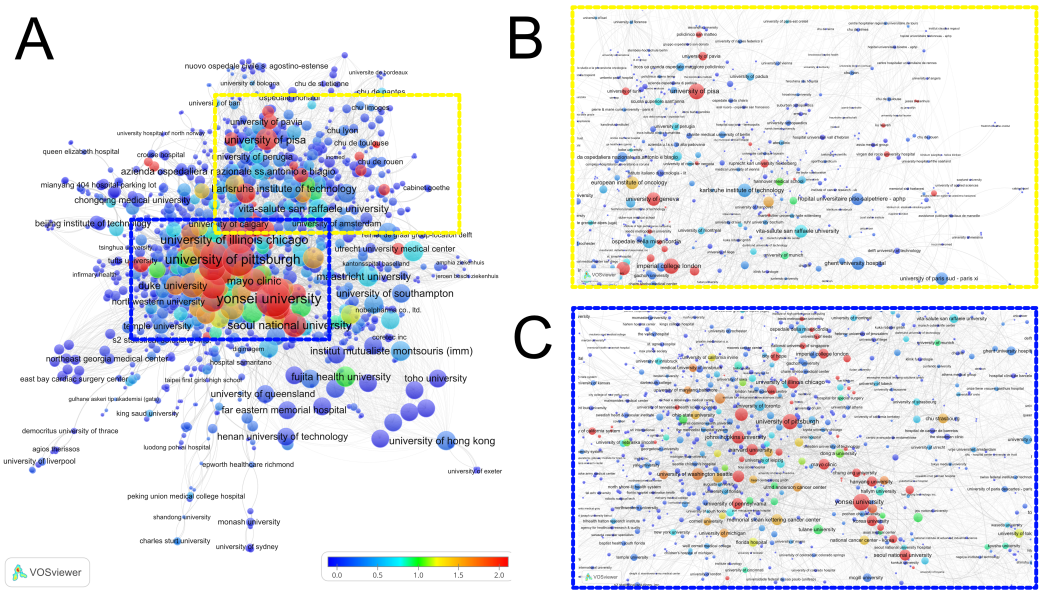


Figure 12 The global robotic surgery collaboration network. The size of each node is proportional to the average normalized citations of the corresponding organisation, while the colour is proportional the innovation index. The weight of each link (i.e., the thickness of the line connecting any two nodes) is proportional to the normalised count of collaborations between the connected pair of organisations. Panel A shows the largest connected component of the collaboration network. Panel B shows a subset of less successful organisations, more peripheral and poorly connected. Panel C shows a subset of more successful organisations, highly connected and centrally located within the global network.

Table 1 shows the maximum-likelihood estimates of the coefficients and standard errors of the two hierarchical random-intercept models of research impact and innovation. The first two estimated parameters in both models suggests that both citations and innovation value at the organisational level were statistically significantly associated with the organisation's position in the collaboration network. The local clustering coefficient at the organisational level was negatively associated with both performance measures, although only the association between clustering and innovation reached statistical significance. Both geographical entropy and industrial collaboration were positively and statistically significantly associated with both research impact and innovation. Estimates for all remaining fixed-effect and random-effect parameters are shown in Tables S3 and S5 in Appendix 2 (see also Tables S4 and S6-S8 for robustness checks).

Fig. 13 shows the topology and properties of four ego-centered networks of selected organisations that differed in terms of both innovation index and clustering coefficient. For instance, **Fig. 13A** suggests that Leiden University was characterised by a closed ego-centered network in which the collaborators tended to collaborate with one another, while Imperial College London (**Fig. 13C**) was positioned in a more open network, rich in structural holes and opportunities for brokerage between collaborators. In turn, Imperial College London was associated with a higher innovation index than Leiden University, which indicates that organisations can extract value from the structural cleavages separating their partners.

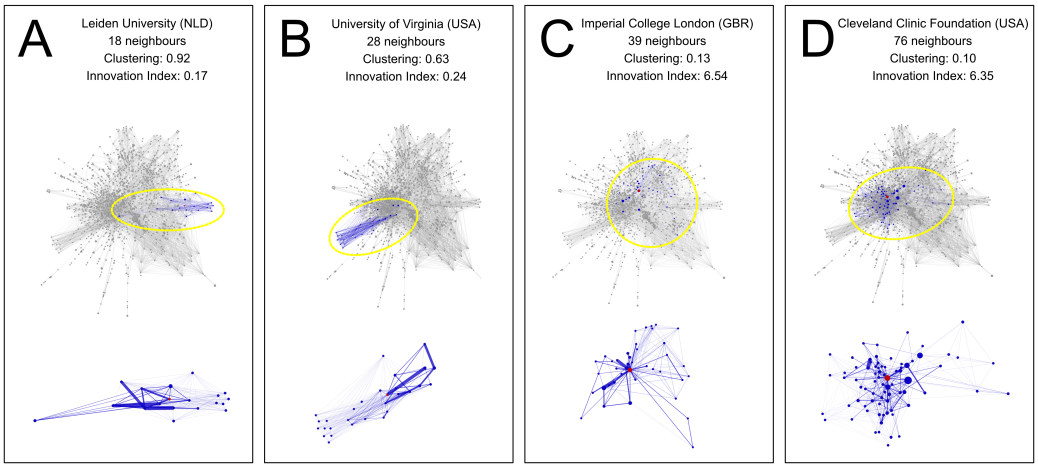


Figure 13 Ego-networks of four selected organisations, with decreasing values of clustering coefficient and increasing values of innovation index. In each panel, the ego-centered networks are identified by the yellow circles (above), and zoomed out (below).

Fig. 14 shows the association between closure of ego-centered networks (node size) and both measures of performance (node colour). **Fig. 14A** does not suggest an unambiguous relationship between network closure and research impact, as both large

and small nodes can be associated with high performance. **Fig. 14B**, however, indicates that nodes within closer structures were associated with lower values of innovation index. Organisations that produced more innovative outcomes were those that spanned structural holes between collaborators.

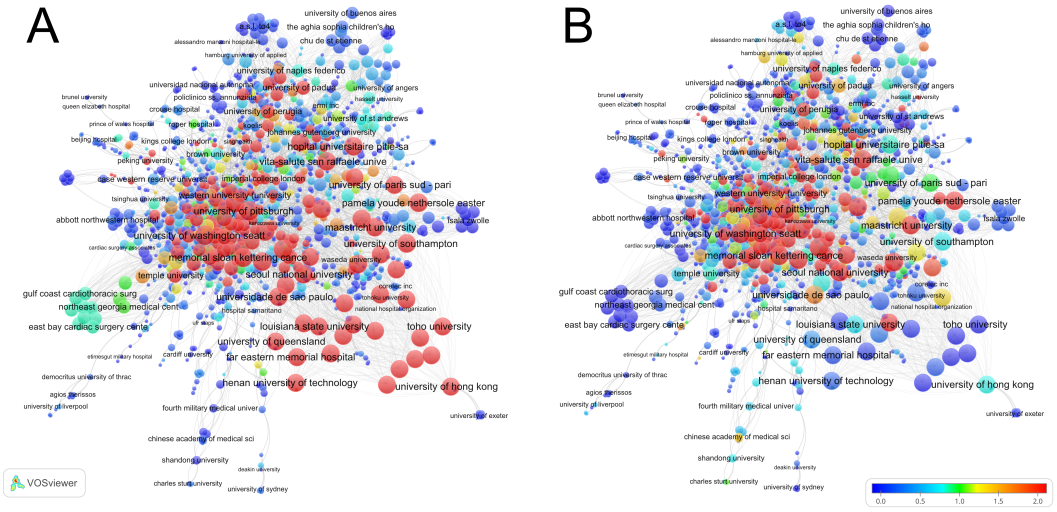


Figure 14 The association between closure of ego-centered networks (i.e., size of nodes) and both measures of performance (i.e., colour of nodes). In panel A, the colour is proportional to research impact, while in panel B to the innovation index. In both panels, the size of each node is proportional to the number of closed triangles including the node. While there is no clear-cut relationship between network closure and research impact (i.e., there are both large and small red nodes in panel A), nodes in closer structures are associated with lower values of innovation index (i.e., most large nodes tend to be the blue ones in panel B).

	Research impact (*snc*)	**Innovation value (*ii*)**
Clustering coefficient (*cc*)	- 0.4027 (0.6165)	**- 0.0820** (0.0284)
Closeness centrality (*l*)	**6.0782** (0.7253)	**0.2166** (0.0332)
Geographical entropy (*ε*)	**10.5599** (0.7564)	**0.1629** (0.0368)
Industrial collaboration (*ic*)	**5.2104** **(1.2837)**	**0.1967** (0.0686)
Number of articles in WOS (*v*)	0.0015 (0.0022)	0.0001 (0.0001)
Number of observations*	639	639

Table 1 Maximum-likelihood estimates from three-level random-intercept linear models of research impact (measured by the sum of normalised citations, snc) and innovation value (ii). Below each estimated parameter, the corresponding standard error is reported within brackets. Estimated parameters that are statistically significant at the 5% level are shown in bold.

*The regression models were estimated based on a sample of 639 organisations as these were the observation units with non-missing values across all the covariates used (see Appendix 2 for details).

4.4 Discussion

The study in this chapter was concerned with social capital in surgical research, and has uncovered structural sources of research impact and innovation. The study examined the relative benefits of two opposing structures - closed and open networks - by analysing the ego-centered networks of all organisations that published research on robotic surgery between 1988 and 2017. The association between local clustering coefficient and a two-fold measure of research performance (impact and implementation-based innovation value) was investigated. Closed structures, rich in third-party relationships, were negatively associated with both measures, although only the association with the latter was statistically significant.

It is difficult to identify the reason(s) underlying these findings. A plausible explanation (though this was not tested and would form the subject of a separate study) is that nodes in a closed structure are more likely to cite each other, thus inflating their research impact (citations), which would explain the lack of statistical significance of the negative association between clustering and research impact. At the same time, a closed structure would naturally limit the opportunities for brokerage between collaborators and in so doing compromise their ability to translate novel ideas, products, or surgical procedures they may have developed and published into truly innovative outcomes that could subsequently be implemented on a large scale. This would explain the negative and statistically significant association between closed structures (clustering) and innovation.

Closed networks may still be important to surgical innovation. They may facilitate distributed understanding, distributed ownership, and the application of complex ideas.[89] However, the present findings suggest that open networks and brokerage opportunities are vital for fostering truly innovative outcomes. Open networks enhance innovation by providing the necessary conditions and opportunities for novel combinations or rearrangements of ideas, technologies, processes, and for transforming them into well-established, widely implemented products, medical devices, or surgical procedures. This is achieved through a variety of mechanisms including access to external knowledge, data, infrastructure, and/or expertise not available 'in-house', and the sharing of otherwise prohibitive costs and risks that represent two of the greatest barriers to innovation.[91] These mechanisms are particularly salient to modern surgical research which is expensive, highly regulated, and increasingly reliant on the diverse inputs from individuals across a variety of

backgrounds (including scholars not directly related to surgery such as engineers, computer scientists, and statisticians).[4]

This study has shown a positive and statistically significant association between closeness centrality and organisational performance. This has important implications for strategy and policy as it can assist both academic surgeons and policymakers in their selection of organisations that have better access to others.[95] For instance, focusing on closeness centrality can prove crucial in a number of scenarios in surgical research such as the optimal set up of multi-centre RCTs where recruitment can be challenging and a large sample size needed.[43]

The findings have also highlighted the importance of maximising an organisation's geographical dispersion in terms of the location of research partners. Geographical entropy represents a complementary, yet distinct, dimension to structural openness. It reinforces the positive association between brokerage and performance by shedding light on the benefits that an organisation can extract from a (geographically) diverse collaboration network (see Appendix 2).

It is encouraging to see that global surgery has been recently gaining increasing attention through a number of collaborative initiatives.[85,86] While these initiatives have been shown to provide well-known benefits to surgical education and global health (through the specialised service, training and equipment provision to the developing world),[85,86] this study suggests that they can also enhance the research impact and innovative output of all partners involved. Thus, this study's findings have important implications as they can inform policymakers in their efforts to devise

effective collaborative international strategies and surgical research policies which should aim to incentivise organisations to collaborate more globally, also with partners in the developing world.

This study has demonstrated a positive association between academic-industrial partnerships on the one hand, and research impact and surgical innovation on the other. In the corporate sector, such partnerships are actively sought, as they are known to maximise profits. From the perspective of academia, however, there is more ambiguity and controversy on their advantages.[99] On average, industry contributes less than 10% of funding for academic research, and joint academic-industrial partnerships contribute to only a small fraction of the overall knowledge generated.[97] Many possible reasons for this have been put forward. The most widely accepted one relates to corporate pressures diverting academic researchers away from their scientific efforts, towards commercialisation.[97]

The positive association between collaboration with industry and research performance identified in this study may be rooted in key differences between innovation in surgery and in other fields. Surgical research and the introduction of new technology (such as robotic surgery) can be expensive.[43] This, combined with the fact that governmental funding for healthcare research allocated to surgery tends to be no higher than 5% (3% in the USA and 5% in the UK), engenders the need for seeking alternative sources of funding.[100,101]

The medical device industry, with its continuing steady growth in revenues (projected to reach $398 billions by 2023) amid periods of economic downturns across the

Western world, represents an ideal collaboration partner in some respects.[9] Not only does industry possess the scale of funds that surgical research often needs (as well as other vital resources including laboratories and human capital) but it also has a strong interest in investing in surgical research (through intellectual property acquisition and subsequent commercialisation of innovations).[9]

The internal organisational mechanisms of corporations make them highly experienced in managing various types of resources. This is likely to result in a more efficient division of labour,[97] and higher research productivity of academic surgeons, typically constrained by demanding clinical and educational commitments. Through the provision of support staff to assist with time-consuming, bureaucratic activities (such as the drafting of applications for ethical committee approval, the performance of standardised laboratory work, and patient recruitment and follow-up), industrial collaboration can further free academic surgeons to concentrate on their research.[97] It is encouraging to see that world-leading universities have already recognised the strategic salience of collaborations with industrial partners, and have set up dedicated liaison offices and even innovation hubs to act as 'incubators' for innovation.[102,103] Academic institutions involved in surgical innovation should be encouraged to follow their example.

Despite those encouraging findings, it is important to recognise that industry partnerships may not always be beneficial. Caution should be exerted, especially when corporate research funding forms part of the partnership because this will inevitably promote conflicts of interest (including those concerned with intellectual property and ownership of the innovations)[104] as the goals of industry and academia

do not often align – the former is predominantly driven by commercial incentives whilst the latter by serving the public good.[105] Problems can range from subliminal biases (e.g., surgeons opting for expensive medical devices in the absence of evidence on superiority to existing ones) all the way to research misconduct (e.g., companies suppressing the publication of trial results not favouring their sponsored products) generating ethical concerns.[106] It is however reassuring that the severity of these problems is now widely recognised and a number of measures have been taken to counteract them. Examples include the establishment of the Physician Payments Sunshine Act and initiatives from the International Committee of Medical Journals Editors (ICMJE) regarding clinical research governance that include the compulsory registration of all clinical trials and submission of conflicts of interest disclosure forms by all authors.[107]

The present study has a number of strengths and weaknesses. Its main strength lies in the novel network-based perspective on surgical research. Network science provides a comprehensive array of theories and methods for mapping and understanding collaboration patterns.[108] Despite its extensive use across the social sciences and innovation studies,[83] its application to the study of collaboration patterns in healthcare has so far remained relatively limited with only a handful of published studies.[92,95,109,110] In particular, most studies on surgical collaboration did not rely upon network metrics beyond the mere number of participating organisations and their corresponding countries.[85,87,111-114] This study took a step in this direction by computing both local (i.e., clustering coefficient) and global (i.e., closeness centrality) measures that account not only for each organisation's connections with collaborators but also for connections these collaborators have with one another (in the

organisation's ego-centered network) and with other organisations (in the global network).

This study also engaged with current debates and controversies on social capital and its structural sources. The fundamental role of networking in surgical research for both innovation and patient care is widely recognised,[87] and an increasing number of national and global surgery initiatives have been recently established.[85,111-114] However, which type of networking pattern (e.g., open vs. closed structures) matters has remained largely unexplored. The present study investigated the relative benefits of different collaboration patterns for both research impact and surgical innovation. The findings have suggested that both collaborative brokerage and geographical boundary spanning are catalysts of surgical innovation. Much of the network and innovation literature has failed to distinguish between these two mechanisms, although they remain conceptually and empirically distinct. This study contributed to disentangle them by suggesting that surgical innovation can be further enhanced when boundary-spanning leaders leverage collaborative brokerage, and brokering leaders amplify the spatial diversity of their collaborative network.

Limitations included the fact that the collaboration network was constructed based on co-authorship, and therefore did not reflect any other form of informal intellectual exchange (e.g., mentorship, discussion, informal commentary) that did not result in a publication. Fully accounting for the problem of the opaqueness of collaboration would inevitably be an arduous task, especially when conducted on a large scale, and I believe the results still remain fairly robust against possible biases that using co-authorship as proxy for collaboration might induce.[92,93] Another limitation of the

current analysis lies in its cross-sectional nature. Future work might consider investigating the evolution of the collaboration network over time and uncovering the dynamics of social capital. Collaboration was only evaluated at the meso (organisational) level. It will be interesting for future studies to evaluate structural sources of innovation and research impact at the micro (individual) and macro (country) levels as well as sources of other innovations in surgery (e.g., augmented reality for intraoperative navigation in robotic surgery and 3D printing).

4.5 Conclusion

This study can be regarded as a proof of concept suggesting how network analysis can be used in surgical research to foster innovation and thus patient care through strategic partnerships. The findings, showing that innovation is inherently a social process, have a number of implications that can potentially inform policymakers and funding bodies. Evidence was provided on the ways in which existing collaborative efforts can be adjusted and future ones strategically planned to maximise research performance.

4.6 Chapter summary

This chapter presented a novel network-based framework for the study of collaboration in surgery and demonstrated how this can be used in practice to help build and nurture collaborations that foster innovation.

Surgical innovation is a social process that originates from complex interactions among diverse participants. This has led to the emergence of numerous surgical collaboration networks. What is still needed is a rigorous investigation of these networks and of the relative benefits of various collaboration structures for research and innovation.

Network analysis of the real world innovation network in robotic surgery was conducted in this study. Hierarchical mixed-effect models were estimated to assess associations between network measures, research impact and innovation, controlling for the geographical diversity of collaborators, institutional categories, and whether collaborators belonged to industry or academia.

The network comprised of 1,700 organisations and 6,000 links. The ability to reach many others along few steps in the network (closeness centrality), forging a geographically diverse international profile (network entropy), and collaboration with industry were all shown to be positively associated with research impact and innovation. Closed structures (clustering coefficient), in which collaborators also collaborate with each other, were found to have a negative association with innovation ($p<0.05$ for all associations).

In the era of global surgery and increasing complexity of surgical innovation, this study highlights the importance of establishing open networks spanning geographical boundaries. Network analysis offers a valuable framework for assisting surgeons in their efforts to forge and sustain collaborations with the highest potential of maximising innovation and patient care.

The following chapter, Chapter 5, involves the analysis of collaboration networks in surgical innovation at the macro (country) level. This is performed to study the UK's existing research collaboration networks on a global scale, and through a number of simulation models (based on different scenarios), measure the impact that Brexit may have on the UK's global research and innovation performance in the long-term. Associated policy implications are also discussed.

5. Quantifying the implications of Brexit for research collaboration and policy through network analysis and simulation modelling

5.1 Introduction

On 23rd June 2016 the United Kingdom (UK) voted to leave the European Union (EU) through the EU referendum, a process known as 'Brexit'. Since then there has been a dramatic rise in the number of publications on Brexit and its potential impact on the UK's performance across various sectors.[115-117] Although quantitative studies have been produced on the impact that Brexit might have on the UK's economy, trade, and industries[118-122] the majority of articles in the scientific and healthcare literature are editorials and letters expressing individuals' as well as scientific and industrial bodies' concerns.[115,123-127] Even the handful of studies on the Brexit impact on healthcare, research and innovation are limited to narrative descriptions of possible scenarios, and thus fail to provide any measurement of the reported expected impact,[116,128,129] with the exception of one recent study in the field of scientometrics.[130] To preserve the UK's global leading position in healthcare innovation, measurement of impact is needed that can inspire the development of a coherent set of policies for each scenario.

As robotics features at the top of the UK Government's *Eight Great Technologies* initiative,[131] this study aims to quantify the contribution of the EU to the UK's success

as a global leader in robotic surgery.[132] Simulation-based studies have been carried out to evaluate how the UK's prominent position could be preserved (or even improved) in the event that UK-EU research collaborations should deteriorate significantly. These studies leverage network analysis to provide visual maps and a quantitative assessment of collaborative patterns and their association with innovation.[39] The findings may assist policymakers in directing the UK's international strategy post-Brexit with a view to ensuring downsides are mitigated and opportunities beyond the EU seized from 1 January 2021 when the current transition period ends.

5.2 Materials and methods

5.2.1 The dataset

All articles on robotic surgery were extracted from the *Web of Science*® platform (Clarivate Analytics, Philadelphia, PA). The search was performed on 17th of January 2017, and produced 3,866 peer-reviewed articles published between July 1988 and January 2017 (see S.1, in Appendix 3 for details), which were then used to construct the international collaboration network. Any article that did not represent original research studies on robotic surgery was excluded (e.g., review article or conference article).

5.2.2 The international collaboration network

The international collaboration network was constructed and analysed using MATLAB® (The MathWorks, Inc., Natick, MA). As co-authorship has been shown to

be a good proxy for collaboration,[92,93] the links between nodes are based on co-authorship of articles by co-authors. First, a network comprising 2,397 collaboration links connecting 772 organisations across 56 countries and 7 geographical regions was constructed (Appendix 3). Links and nodes were then aggregated to produce a collaboration network between countries. In this network, each node represents a country, and links between nodes represent collaborations between countries. This is a weighted network in which each link is associated with a value reflecting the intensity of the collaboration between the connected nodes (see S.2 in Appendix 3 for details).

5.2.3 Outcome measures

Two outcome measures were calculated at the national level:

- Research impact (snc_i): This was measured as the sum of the research performance of all organisations residing in a given country. In turn, for each organisation, academic performance was measured by computing the sum of the normalised citations received by all articles published by the organisation across the years. To obtain normalised citations, the citation count for each publication in a given year was divided by the average number of citations obtained by all articles published in the same year (see S.3.1 and S.4 in Appendix 3 for details).

- Innovation index (ii_i): This metric was specifically developed for the measurement of surgical innovation. It is based on surgical implementation stages and levels of evidence associated with articles, and was validated against real world data from the clinical setting. We measured the innovation index of a given country as the sum of the innovation indices of all organisations residing in the country. In turn, for a given organisation, the innovation index was computed as a function of the innovative value

of all articles published by authors affiliated with that organisation (see S.3.1, Figs.S2-S3, and Table S2 in Appendix 3 for details).[39]

5.2.4 Network measures

A number of measures were computed to capture the structural position that each country occupies in the global collaboration network (see S.5.1-S.5.3 in Appendix 3 for details).[39,88]

- Degree (k_i): This is defined as the number of a node's direct links to other nodes, i.e., the number of countries a focal country collaborates with. A country *i* has a link with a country *j* if there is at least one organisation in *i* that collaborates with at least one organisation residing in *j*. Thus, a link between two countries *i* and *j* implies that there are at least two scientists, one affiliated with an organisation in country *i* and the other in country *j*, that have co-authored at least one publication, thus generating a collaborative link between the corresponding countries (see S.5.1 in Appendix 3 for details). [88]

- Effective size (es_i): This measure captures the extent to which a focal node is connected to non-redundant nodes, thus reflecting brokerage opportunities between contacts. A node's links to neighbours are redundant when the neighbours are already connected to each other and can thus bypass the focal node. Following Latora *et al*. (2013),[14,88] we defined a node *i*'s effective size as a function of the node's degree k_i and its binary local clustering coefficient cc_i (defined as the ratio between the number of actual triangles centred on the node *i* and the maximum possible number of such triangles; see S.5.2 in Appendix 3 for details).[108] Thus, the more a node's neighbours are connected with each other, the more closed triangles are centred on the node, and

the less the effective size of the node is (see S.5.3 in Appendix 3 for details).

- Efficiency (e_i): This is the normalised effective size of a node i, i.e., the ratio between node i's effective size es_i and its degree k_i. To facilitate the interpretation of the estimated regression coefficients, all values of efficiency were standardised (see S.5.3 in Appendix 3 for details).

5.2.5 Geographical dispersion of collaborations

To capture the geographical dispersion of a country's collaborative activities, the *geographical entropy* ε_i of country i's collaborations was computed (see S.5.4 in Appendix 3 for details).[133] A country's geographical entropy increases as the country: (i) collaborates with (organisations located in) more countries; and (ii) devotes an equal amount of collaborative effort towards each of these countries. To facilitate the interpretation of the estimated regression parameters, all values of geographical entropy were standardised.

5.2.6 Control variables

Two additional country-level characteristics were controlled for: (i) average contribution to publication of international collaborators, $\langle p_{a_i}^{int} \rangle$, i.e., the degree to which international collaboration contributed on average to an article published by country i; and (ii) average number of publications per domestic organisation, $\langle A_{org_i} \rangle$, i.e., the ratio between the total number of articles published by country i and the total number of unique organisations in country i (see S.5.5 and Fig.S4 in Appendix 3 for details). Moreover, fixed effects for countries' geographical regions were estimated.

5.2.7 Statistical analysis

A summary of the names, symbols and interpretation of all dependent and independent variables used in the regression analysis is provided in **Table 2**. To investigate the association between brokerage and geographical entropy on the one hand and research impact and innovation on the other, efficient generalised methods-of-moments (GMM) instrumental-variables two-step estimators were computed that are robust to heteroskedasticity (see S.6. for details). For country *i*, we estimated the following two models:

$$snc_i = \beta_1 + \beta_2 \langle p_{a_i}^{int} \rangle + \beta_3 \langle A_{org_i} \rangle + \beta_4 e_i + \beta_5 \varepsilon_i + \sum_{k=6}^{11} d_{k,i} \beta_k + u_i$$

and

$$ii_i = \beta_1 + \beta_2 \langle p_{a_i}^{int} \rangle + \beta_3 \langle A_{org_i} \rangle + \beta_4 e_i + \beta_5 \varepsilon_i + \sum_{k=6}^{11} d_{k,i} \beta_k + u_i,$$

where $\langle p_{a_i}^{int} \rangle, \langle A_{org_i} \rangle, es_i,$ and ε_i are the covariates (as defined in the previous sections; see also **Table 2**), $d_{k,i}$ is the dummy variable for the *k*-th geographical region of country i, β_1 to β_{11} are fixed parameters, and u_i is the country-level error term. All models were estimated using STATA 15 (StataCorp LP, College Station, TX, USA).

Name of variable	Symbol	Variable description
Research impact	$\underline{snc_i}$	Sum of normalised citations received by all articles published by all organisations in country *i* across the years
Innovation index	$\underline{ii_i}$	Evidence-based innovative value of all articles published by country *i*
Average contribution to publication of international collaborators	$\langle p_{a_i}^{int} \rangle$	Extent to which international collaboration contributed on average to an article published by country *i*
Average number of publications per domestic organisation	$\langle A_{org_i} \rangle$	Ratio between total number of articles published by country *i* and total number of unique organisations in country *i*
Efficiency	$\underline{e_i}$	Extent to which country *i*'s connections to other countries are non-redundant
Geographical entropy	ε_i	Geographical dispersion of country *i*'s collaborations (i.e., diversity of the geographical locations of country *i*'s collaborators)

Table 2 Summary of variable names, symbols, and interpretation.

5.2.8 Simulation modelling

To understand the potential impact of Brexit on the UK's research impact and surgical innovation and to evaluate how any adverse outcomes could be mitigated through suitable strategic collaborations beyond the EU, a number of simulations were carried out based on the real world network. These simulations involved the severance of all UK-EU27 collaborations and their subsequent replacement with collaborations with different geographical clusters, with a view to investigating how the UK's performance would change.

Two simulation studies were conducted, one to assess the effect of such substitutions on research impact and the other the effect on innovation (see S.7 in Appendix 3 for details). To ensure all replacements of collaborations were fair and on the

conservative side, all UK-EU collaborations were ranked by their corresponding output (i.e., research impact or innovation) and, starting from the worst (lowest performing) one, each collaboration was progressively substituted with a UK-non-EU collaboration. Specifically, for each output, three simulations were conducted in which each UK-EU collaboration was substituted with the UK-non-EU collaboration associated with one of the following three outcome measures: (i) the 25th percentile; (ii) the median; and (iii) the maximum value of the distribution of normalised citations or innovation index of all articles resulting from all UK-non-EU collaborations. To simulate the effects of newly established UK-non-EU collaborations on performance, three geographical clusters outside the EU were identified: (i) the US (the international leader in surgical innovation); (ii) Asia (including emerging research and innovation powerhouses such as China, South Korea, and India); and (iii) the 'Rest-of-the-World' (RoW).

5.3 Results

5.3.1 Global collaboration network

Fig. 11 (Chapter 4) shows two ego-centred networks comprising links between the focal node ("ego") and its neighbours ("alters"), and links between neighbours. The figure shows that two nodes with the same degree (i.e., number of collaborators) can be characterised by different structures of their ego-centred networks. A node collaborating with others that also collaborate with each other has redundant links generating closed triangles and a closed network structure. A node collaborating with disconnected others is characterised by an open network structure, rich in structural holes and brokerage opportunities.[88,108]

Fig. 15 (left-hand panel) shows the global collaboration network in robotic surgery consisting of 2,397 links connecting 772 organisations across 56 countries, where the thickness of lines is proportional to the intensity of collaboration, and the colour and size of nodes are proportional, respectively, to the research impact and innovation index of the corresponding country. The network in the right-hand top panel only highlights UK-EU collaborations. The right-hand bottom panel shows the association between countries' research impact and innovation on the one hand, and number (i.e., degree) and geographical dispersion (i.e., entropy) of collaborations on the other.

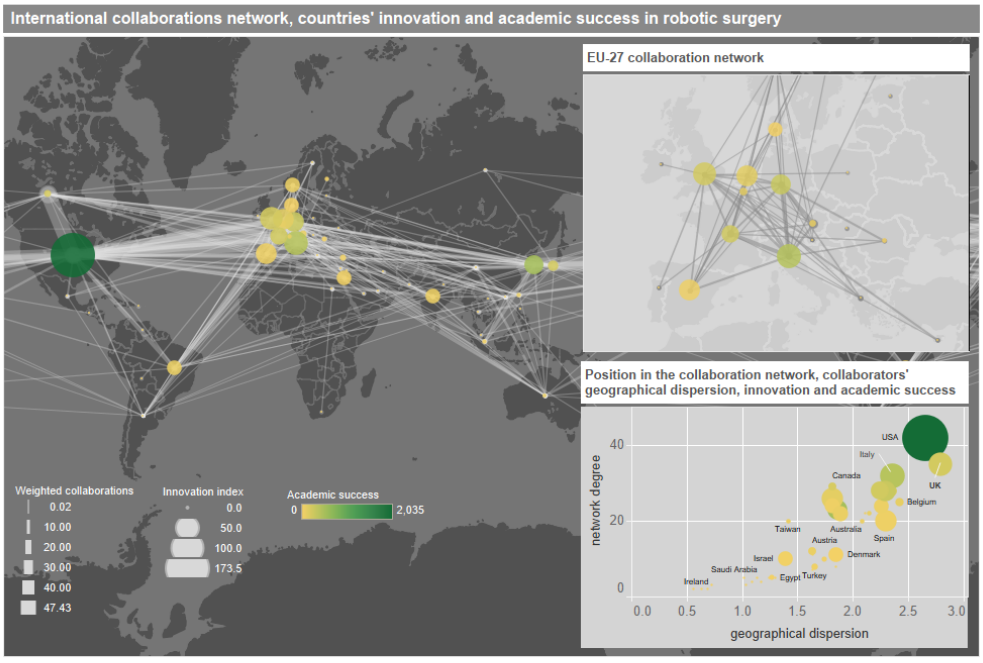


Figure 15 The international collaboration network in robotic surgery. The left-hand panel shows the global collaboration network consisting of 2,397 links connecting 772 organisations across 56 countries. The network is weighted, and the thickness of the links corresponds to the intensity of the collaboration between the connected countries. The size of each node is proportional to the corresponding country's innovation index, whereas the colour to the average normalised citations received by the country in the field of robotic surgery. The top right-hand panel shows a zoomed snapshot of the collaboration network between the UK and all EU countries. The bottom right-hand panel shows the association among countries' network degree, geographical dispersion of collaborators, research impact and innovation index.

5.3.2 Global leaders in surgical innovation and their collaboration profiles

Among all articles in robotic surgery, those involving international collaboration achieved the highest performance in terms of both citations and innovation. By contrast, articles involving no collaboration had the lowest performance. Interestingly, articles that originated solely from domestic collaborations had only a marginal superior performance to those involving no collaboration at all, both scoring significantly less than those resulting from international collaborations (in terms of both citations and innovation). The findings are summarised in **Fig. 16**.

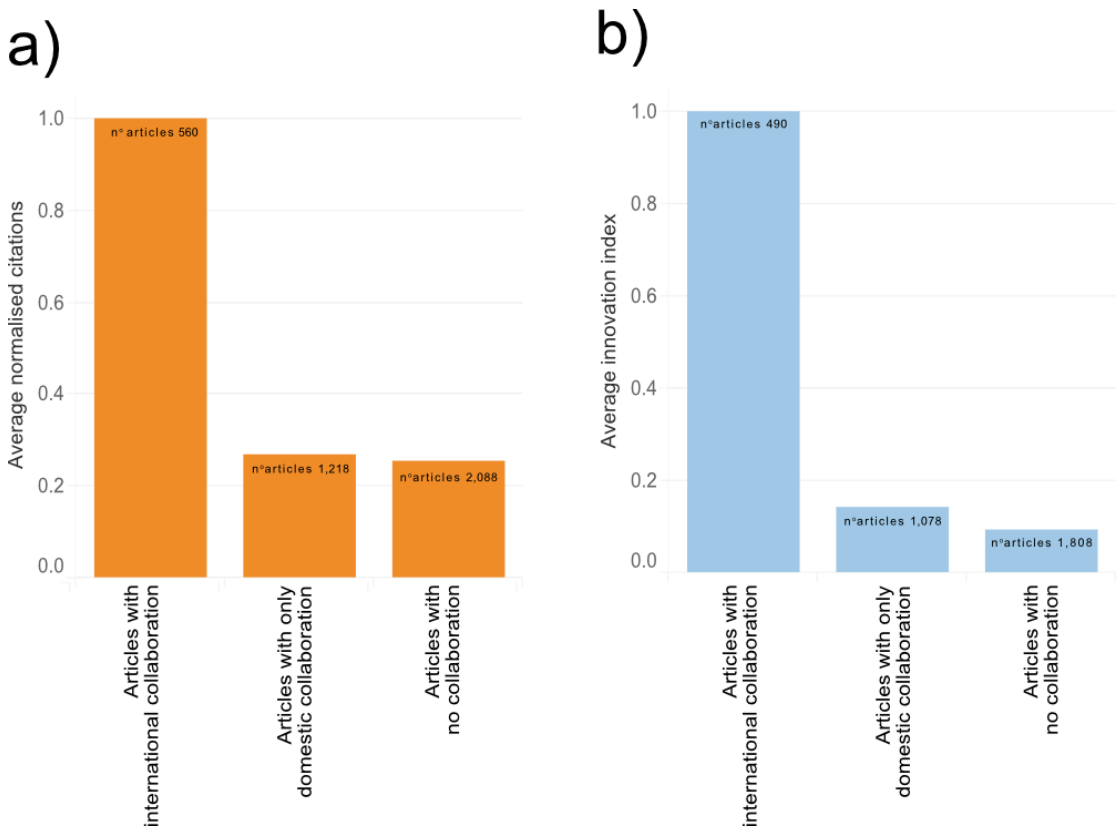


Figure 16 Average normalised citations (left-hand panel) and innovation index (right-hand panel) as a function of type of collaboration. Numbers within bars refer to the corresponding number of articles in each category. Average values of both citations and innovation index are divided by the maximum value across the corresponding three sets.

The most innovative country in the field of robotic surgery is the US, followed by Italy and then the UK. Among the top ten for surgical innovation, there are another four countries from the EU (Netherlands, Spain, Germany, France) and three from Asia (South Korea, China, and India). In terms of research impact, the ranking is similar: the US still ranks first, with the UK also appearing in the top ten (sixth), along with Canada, Japan and South Korea (ranking second). **Fig. 17** illustrates the rankings of countries by research impact and innovation. It also depicts the contribution of collaboration patterns (international; domestic; no collaboration) to each country's research impact and innovation. **Fig. 18** sheds light on the effects of international collaboration to success, by showing the contributions of specific geographical regions and countries to each individual country's research impact and innovation. It becomes apparent that for robotic surgery the UK's primary collaboration partner and largest contributor to both research impact and innovation is the EU followed by the US.

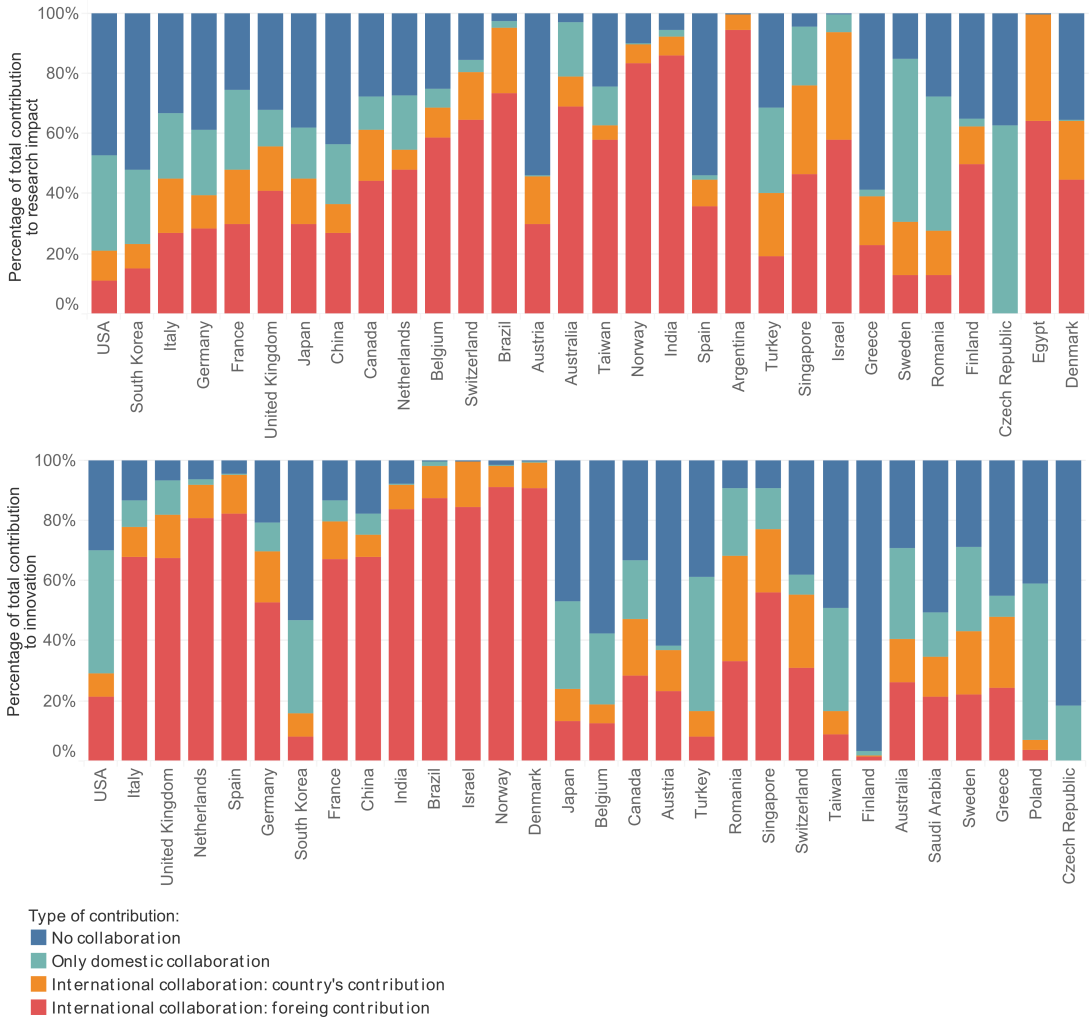


Figure 17 The top 30 countries ranked according to their research impact (top) and innovation index (bottom). Bars are colour-coded to indicate the type of collaboration that contributes to the overall output. The contribution to success of articles involving international collaboration is further split into the contribution attributable to the focal country (orange bar) and the contribution attributable to the foreign collaborators (red bar). Countries are ranked by (total) research impact (top panel) and innovation (bottom panel), starting with the top-ranked on the left.

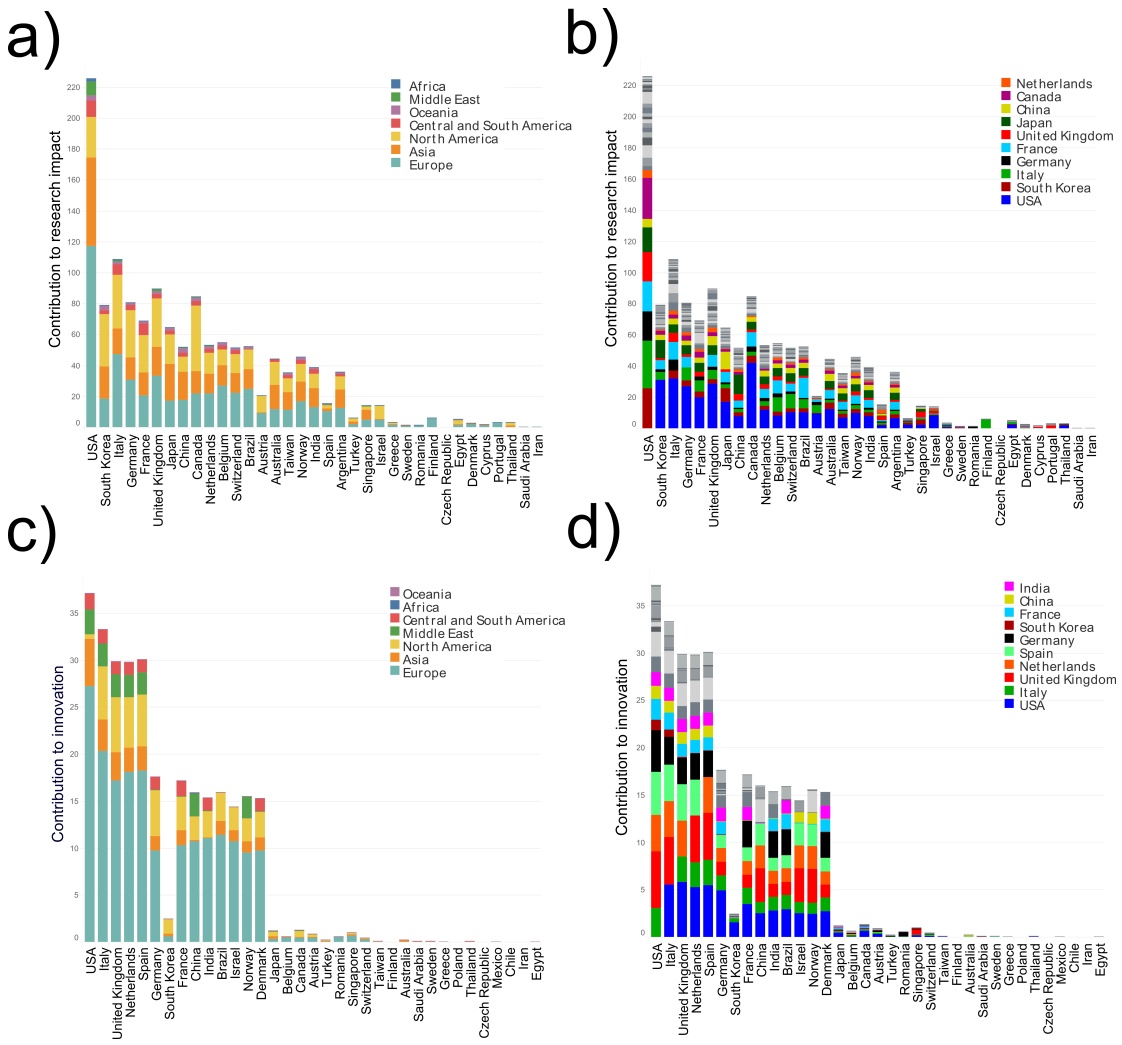


Figure 18 Contribution of geographical regions (left-hand panels) and specific countries (right-hand panels) to countries' research impact (top panels) and innovation index (bottom panels). Countries are ranked by (total) research impact (top panels) and innovation index (bottom panels), starting with the top-ranked on the left. The length of each bar corresponds to the contribution to each country's success attributable solely to international collaborators (red bars in Figure 17).

5.3.3 Regression analysis

Table 3 shows results from the estimated regression models (see Tables S3-S6 for descriptive statistics, estimates of all remaining parameters, and specification tests). There is no statistically significant difference in performance between two countries that differ in contribution of international collaborations and average number of publications per organisation, after controlling for the other covariates. By contrast, both efficiency and geographical entropy are statistically significantly associated with

both measures of performance. After controlling for the other covariates, as efficiency increases by one standard deviation, a country's normalised citations and innovation value increase by 211.320 and 18.819 units, respectively. Moreover, an increase by one standard deviation in geographical entropy is associated with an expected increase of 244.537 units in normalised citations and of 30.850 units in innovation value.

	Model (1) **Research impact**	**Model (2)** **Innovation**
Contribution of international collaborators	68.272 (158.432)	8.109 (11.384)
Number of articles per organisation	33.668 (48.883)	-2.363 (4.174)
Efficiency	**211.320** (63.625)	**18.819** (6.108)
Geographical entropy	**244.527** (55.760)	**30.850** (4.641)
Constant	74.600 (106.865)	7.116 (10.900)
No. of observations	56	56
Centred R^2 (uncentred R^2)	0.572 (0.614)	0.519 (0.587)
Root MSE ($\hat{\sigma}_e$)	184.4	17.63
F(10, 45)	**3.26**	**7.17**

Table 3 Instrumental-variables models of countries' research impact and innovation. The models are efficient GMM instrumental-variables estimators with standard errors robust to heteroskedasticity. Robust standard errors of parameter estimates are within parentheses. Reported values of the square root of the mean squared error (MSE) refer to the estimated standard deviation $\hat{\sigma}_e$ of the idiosyncratic disturbance. Estimated parameters in boldface are statistically significant at the 1 percent level (p-value<0.01).

5.3.4 Simulation results

A number of simulations were carried out to examine the potential effects of replacing EU collaborators with other international partners on the UK's research impact and innovation in robotic surgery. The US appears the most favourable collaboration partner replacing the EU in the event of a "hard Brexit" (**Fig. 19**). If that were the case, the simulation study suggests that the UK's innovation may actually improve,

especially if existing collaborations with EU countries were to be replaced with top-performing collaborations with the US. However, the effect of such substitutions on the UK's research impact would be less pronounced. In particular, the UK's research impact would ultimately suffer if an increasing percentage of UK-EU collaborations were replaced by UK-US collaborations (panel a), unless the UK managed to substitute all UK-EU collaborations with high-performing collaborations with the US (panels b,c).

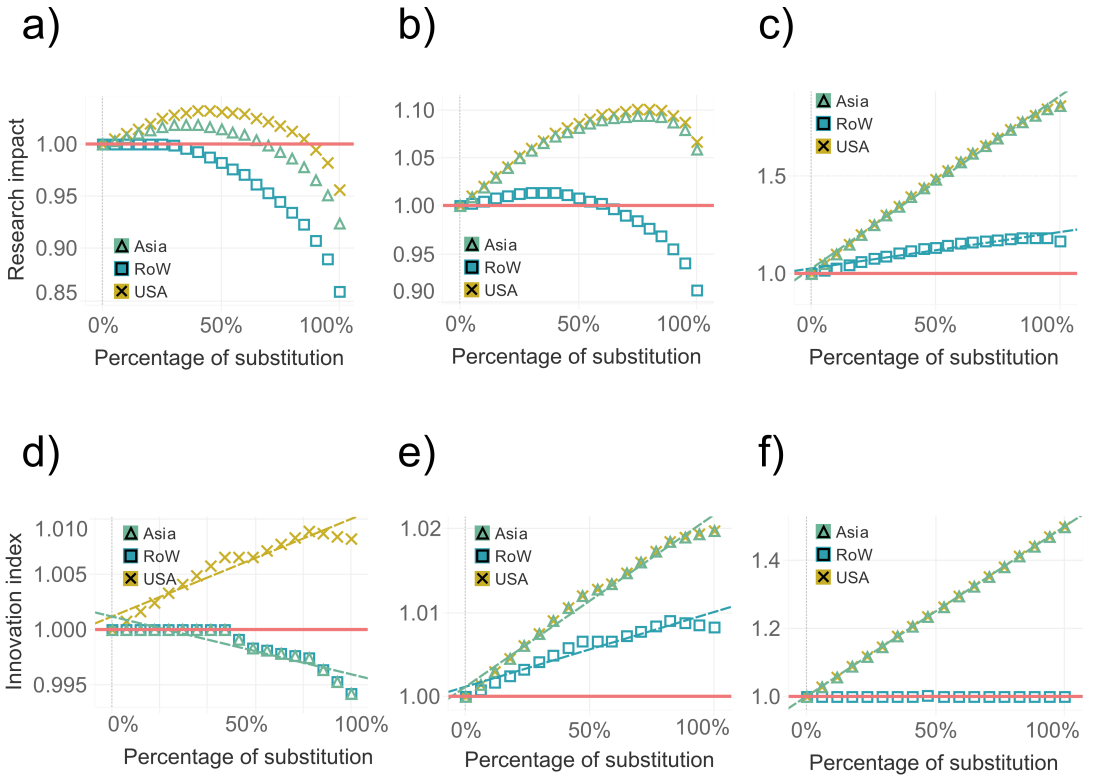


Figure 19 Simulation of the variation in the UK's research impact (top) and innovation index (bottom) as a result of substituting EU collaborators with non-EU collaborators. All articles resulting from a collaboration involving only the UK and one or more of the EU27 countries were ranked from worst to best performing by research impact (top panels) and innovation index (bottom panels). The performance of various percentages of these articles, starting from the worst performing one, was replaced by the 25th percentile (panels a,d), the median (panels b,e), and the maximum value (panels c,f) of the distributions of research impact (top panels) and innovation index (bottom panels) of the sets of articles resulting from collaborations involving only the UK and: (i) the US; or (ii) Asia; or (iii) the 'Rest-of-the-World' countries excluding the EU27 members. The red reference line corresponds to the actual value of the UK's performance, normalised to one to facilitate the visual interpretation of the simulation results. Dotted lines refer to best fitted linear trends relating research impact and innovation index to various percentages of substitution of articles. All trend lines have a p-value<0.05, with the only exception of RoW in panel (f) with p-value=0.277.

Fig. 19 also shows that leveraging collaborations with Asian countries would not be as advantageous as with the US, especially in the event the UK were unable to replace collaborations with EU countries with high-performing ones (panels a,d). The simulations also suggest that replacing UK-EU collaborations with collaborations involving RoW countries does not seem to represent an appropriate post-Brexit strategy as this would mostly result in a substantial drop in the UK's performance (except in panels c,e).

5.4 Discussion

This study evaluated research impact and innovation in robotic surgery,[131,132] and suggested that a country can benefit from brokering between many and geographically diverse collaborators. The results also illustrate the EU's crucial role in the UK's phenomenal success. Specifically, the EU was shown to represent the UK's largest collaboration partner and, more importantly, its greatest contributor to both research impact and innovation. With the future UK-EU relationship (including their collaboration on science and innovation) currently under negotiation, this represents a much-needed and timely study.

Despite wide disagreements across the country as to whether Brexit will be beneficial or harmful to the UK, the consensus within the scientific and healthcare communities is that Brexit risks undermining the UK's status as a global leader in science and innovation.[124,134] The three key mechanisms considered responsible are funding withdrawal from the EU, loss of EU researchers from UK universities, and disruption

of the UK's research and innovation networks – as a result of excluding the UK from European research collaborative agreements and the imposition of restrictions on the free movement of people (including scientists).[134] Based on our findings, but also on the wider scientific literature, it becomes apparent that for science and innovation, the UK has greatly benefited from its collaboration with the EU. This is why "it is the UK's ambition to build on its uniquely close relationship with the EU, so that collaboration on science and innovation is not only maintained, but strengthened" as explicitly stated by the UK Government in its UK-EU future partnership paper.[135]

There are a series of reasons why the UK should be prepared for weakening its current research links with EU countries as a result of Brexit. First, the EU may sever those research links to ensure a competitive advantage in other aspects of the negotiations, such as those relating to protecting its political unity and single market integrity.[136] It is also expected that the EU would not be keen to give a deserting member and research powerhouse a deal similar to what it currently holds.[137] But even if the UK managed to remain in the EU's research network, as a non-member, it would undoubtedly lose its existing influence in a number of critical issues such as defining research priorities and directing how EU research funds should be spent.[126] Furthermore, as a 'third country', the UK would likely go "at the back of the queue" when it comes to accessing EU research facilities.[134] It is for all these interconnected reasons that it has been argued that "the most effective way to support UK and European scientific research is for the UK to remain in the EU".[138]

Those supporting Brexit could argue that these are only unfounded speculations. However, factual evidence has increasingly emerged.[124,126,134,137,139] Reports from

several leading UK universities have recently documented biases against UK researchers involved in EU programmes including incidents where UK scientists already successful in gaining EU grants were asked not to participate in these programmes following the Brexit vote.[137] Other recent publications, including from the House of Commons Science and Technology Committee, have suggested that a number of UK researchers were coaxed to withdraw from their leading role in Horizon 2020 programmes,[117] some research consortia including UK scientists were dismantled following the Brexit vote,[117] and some researchers were refused to submit applications for EU grants that included collaborators from the UK.[117]

The UK must therefore be prepared for the formation of new strategic partnerships beyond the EU, and not simply await the outcome of the negotiations that started following the UK's departure from the EU on 31 January 2020 and are expected to last until the end of the year, according to the agreed framework for the future relationship between the EU and the UK (importantly the joint Political Declaration published on 19 October 2019 is not legally binding and is "subject to ratification").[140-142] From this perspective, the simulation studies provide insights into the likely effects of such new strategic partnerships on the UK's scientific performance, and can thus prove valuable in directing the UK's future international strategy. Our study suggests that, within the field of robotic surgery, the best future partner for the UK in the event of a 'hard' (or a 'failed', i.e., no-deal) Brexit[128] would be the US, currently its second largest research partner. However, our findings also suggest that the UK could secure an increase in research impact only if collaborations with the EU could always be replaced with top-performing collaborations with the US, something that would be very difficult and costly to achieve in practice, at least in

the short term. On the other hand, the UK's scientific performance may suffer if collaborations with the EU could only be replaced with collaborations with less successful countries or geographical regions (e.g., with non-Asian countries) or with less successful organisations in top-performing countries (e.g., in Asia).

Thus, even if there were a 'hard' Brexit, the UK's best strategy in the short term might be to try and maintain its academic links to the EU. This represents a more pragmatic approach and is based on the findings that the EU represents not only the UK's largest collaboration partner but also its greatest contributor to both research impact and innovation (see **Fig. 18a,c**). In addition to maintaining existing collaborations with the US, another effective strategy, at least in the short term, would be to reinforce existing (and successful) partnerships with Asian countries.[143,144] This, combined with the catalytic role of geographical entropy in boosting both citations and innovation, is likely to help the UK to sustain and reinforce its role as an "open, inclusive and outward-facing" global country. Moreover, in the longer term, the establishment of new and strategic relationships with successful research powerhouses such as the US, China and India will be crucial for maintaining the UK's global leadership in science and innovation.[135,145]

Before concluding, it is important to consider the limitations of this study. The analysis focused on a very specific field, and caution is required in generalising the results to other domains. Additional studies would be needed to uncover the role of collaborative networks in different scientific fields. Moreover, the potential loss of EU funding (the UK received from the EU €8.8 billion for research and innovation in the 2007-2013 period as part of the Horizon 2020 programme grants)[146] and/or of

human capital (almost one in five academics in UK universities are from the EU)[147] as a result of Brexit have not been accounted for. This makes the simulation models conservative, as those losses would only further deteriorate the UK's performance. Finally, another methodological limitation is chiefly concerned with the use of citation-based metrics as indicators of scientific merit.[148]

Despite these limitations, this study has a number of strengths. It is the first of its kind in surgical research,[132] using network analysis and simulation modelling, to examine the impact of Brexit on a key area in which, according to the Government's *Industrial Strategy*, the UK is set to be a global leader.[131] This study also comes at a very critical time point, when the level of participation of the UK in future EU programmes on science and innovation remains under negotiation. By studying the UK's real world international collaboration network, the findings of this study are likely to be of practical value to the UK Government, its negotiators, and policymakers when it comes to setting priorities, making decisions, and devising international strategies to be implemented from 1 January 2021 following the completion of the current transition period.

5.5 Conclusion

As the future UK-EU relationship is still to be determined, there is no doubt that the decisions to be taken over the coming months will be of historical importance and will affect the UK's (and EU's) future for many decades to come. This study on the impact Brexit might have on the UK's scientific research confirms what has long been argued, namely that the UK-EU research partnership has been mutually beneficial and that its continuation represents the best possible outcome for both negotiating parties.

Importantly, this study also provides new insights into the value of opportunities for research collaboration beyond the EU that the UK should be seeking to establish. Finally, the described network- and simulation-based analysis can be seen as a proof of concept for conducting similar studies of other domains that are equally important for the UK's prosperity, such as international trade and financial services, and thus for inspiring policy and directing strategy towards ensuring that the UK remains a global leader in the post-Brexit era.

5.6 Chapter summary

The objective of this study was to evaluate the role of the European Union (EU) as a research collaborator in the United Kingdom (UK)'s success as a global leader in healthcare research and innovation and quantify the impact that Brexit may have.

Methods included network and regression analysis of research collaboration networks, followed by simulation models based on possible Brexit scenarios. The international real world collaboration network of all countries involved in robotic surgical research and innovation were studied. These included 772 organisations from industry and academia nested within 56 countries and connected through 2,397 collaboration links.

The main outcome measures employed were research impact measured through citations, innovation measured through the innovation index, and an array of attributes of social networks to measure brokerage and geographical entropy at national and international levels.

It was shown that, globally, the UK ranks third in robotic surgical innovation, and the EU constitutes its prime collaborator. Brokerage between collaborators and their geographical diversity are associated with a country's research impact (c=211.320 and 244.527, respectively; p-value<0·01) and innovation (c=18.819 and 30.850, respectively; p-value<0·01). Replacing EU collaborators with US ones is the only strategy that could benefit the UK, but on the condition that US collaborators are chosen among the top-performing ones, which is likely to be very difficult and costly, at least in the short term.

This study suggests what has long been argued, namely that the UK-EU research partnership has been mutually beneficial and that its continuation represents the best possible outcome for both negotiating parties. However, the uncertainties raised by Brexit necessitate looking beyond the EU for potential research partners. In the short-term, the UK's best strategy might be to try and maintain its academic links with the EU. In the longer-term, strategic relationships with research powerhouses including the US, China and India are likely to be crucial for the UK to remain a global innovation leader.

The next chapter, Chapter 6, examines innovation in healthcare from a different perspective. Using the example of Transcatheter aortic valve implantation (TAVI), a healthcare innovation associated with an unparalleled diffusion, it introduces the concept of 'disruptive innovation' as this applies to healthcare and examines the factors associated with this, with a particular attention to the evidence base and marketing strategies employed.

6. Evaluation of marketing effects and disruptive innovation in healthcare: Network analysis of TAVI

6.1 Introduction

Aortic stenosis (AS) represents the commonest valvular heart disease. Its increasing incidence has been attributed to a progressively ageing population where AS remains a major cause of cardiovascular mortality.[149] The gold-standard treatment for AS is aortic valve replacement (aVR). However, the procedure involves major surgery that includes a sternotomy (ministernotomy or minithoracotomy for the minimally-invasive approach), aortic cross clamping, and cardiopulmonary bypass. This limits the role of aVR in high-risk patient groups such as the elderly, the group where symptomatic AS is most commonly encountered.[150]

Transcatheter Aortic Valve Implantation (TAVI) represents a recent innovation, initially developed for the treatment of AS in patients where aVR is considered prohibitive.[150] This is most commonly due to the high-risk profile of patients (primarily the elderly but also the frail, and those with significant co-morbidities) but can also be a result of anatomical features posing significant intra-operative challenges (such as a small aortic annulus or 'porcelain' aorta).[151]

Following the first live human case report in 2002, the uptake of TAVI has been unprecedented.[152] In less than a decade, TAVI evolved into a standard therapy for AS

creating its own new 'market' (those AS patients deemed inoperable or high-risk for aVR).[153] Its exponential uptake combined with the fact that it addresses a previously unmet clinical need have led to TAVI being described as a 'disruptive innovation'.[154] However, market disruption is highly uncommon in modern healthcare, especially when it comes to invasive procedures where typically innovations are incremental in nature.[155] This is primarily the result of intense regulatory frameworks established to ensure patient safety at all stages of the innovation process.[39]

Using the example of TAVI, this study introduces the concept of disruptive innovation and examines the factors driving diffusion of innovations in healthcare, with a particular attention to the evidence base and marketing strategies employed.

6.2 Materials and Methods

6.2.1 Dataset

The *Web of Science*® (WOS®) database (Clarivate Analytics, Philadelphia, PA) was searched on 7th December 2017 using the keywords 'aortic valve stenosis' and 'aortic valve regurgitation'. All pre-clinical and clinical original research studies were included to study the evolution of the innovation process at all stages from idea generation to multi-centre RCTs, international registry reports, and the formulation of guidelines. Publications not reporting original research studies (e.g. review articles, book chapters, editorials, letters, patents) were excluded through filters (limiting to 'article' in *Document Types*).

Of the original research articles, exclusion criteria involved studies on paediatric patients (age < 18 years) and those where additional procedures to the aortic valve intervention were performed in the same setting (e.g. TAVI with concurrent coronary angioplasty). In constructing the citation network, articles that did not receive or make at least one citation from/to another article respectively within the dataset were also excluded. The flow diagram illustrating the search strategy is shown in **Fig. 20**.

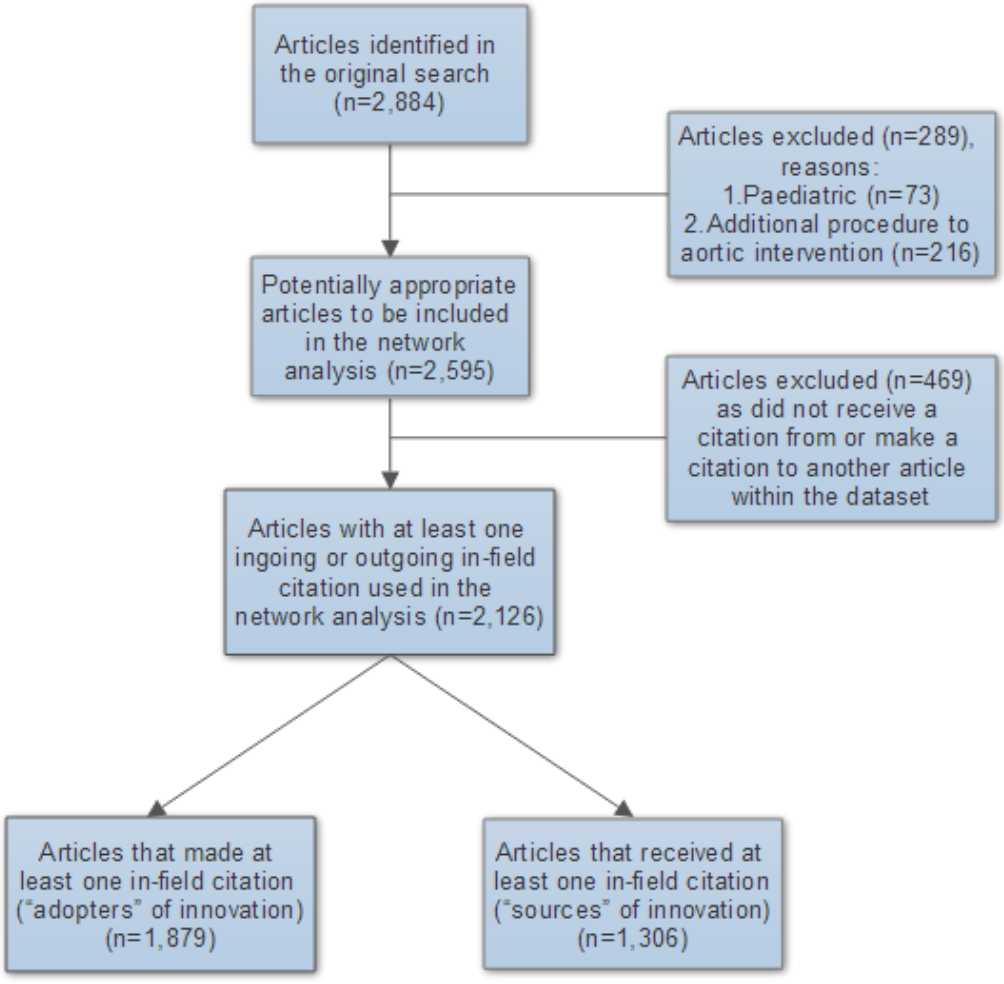


Figure 20 Flow diagram illustrating the search strategy used for generating the citation networks.

6.2.2 Data classification

Another researcher and I critically appraised and classified independently the data for each study: level of evidence, country of study origin (based on corresponding author's affiliation), type of intervention (e.g. TAVI, aVR), manufacturer (e.g. Medtronic, Edwards Lifesciences), type of approach (e.g. transfemoral for TAVI or minimally-invasive for aVR), and for each intervention arm: number of patients, mean age (and standard deviation), number of patient deaths and follow-up time at which mortality was reported, and number of patients developing disabling cerebrovascular accident (CVA), major vascular complications, and moderate-severe paravalvular leak (for the last three outcomes 30-day rates were extracted). Any discrepancies were resolved by consensus.

6.2.3 Citation network and diffusion cascades

In the citation network, the nodes represent the articles addressing the topic of "aortic valve stenosis" or "aortic valve replacement", and a directed link is established from each citing article to all other articles cited in its bibliography addressing the same topic.[156] Through the citation network (n=2,126), the diffusion cascade along which information spreads and adoption of innovation propagates was mapped. The citation-based chain of adoption permits the tracing back to the seed node representing the original article in which a given idea was introduced (e.g. first description of a new approach for TAVI) or initially trialed (e.g. the first human trial of a novel device, e.g. balloon-expandable TAVI). The structure of these cascades can therefore shed light on the diffusion process through which innovations, once proposed, built up momentum over time.[35]

6.2.4 Outcome measures

- Structural Virality (sv_{C_s}): a measure of diffusion of innovation calculated by

$$sv_{C_s} = \frac{1}{l}\sum_{i\in C_s}\sum_{j\in C_s} d_{\overrightarrow{i,j}}, i \neq j,$$

where for $n>1$ articles, for each seed article s, structural virality sv_{C_s} of cascade C_S is defined as the average length of the shortest directed paths between pairs of articles in C_s. $d_{\overrightarrow{i,j}}$ denotes the length of the directed shortest path from article i to article j, C_s is the set of nodes belonging to the cascade originating from article s, and l is the number of directed paths connecting pairs of articles in C_s. So constructed, sv_{C_s} becomes larger as the adopters of the original idea proposed by seed article s are farther apart from one another and from s, thus producing a multi-generational cascade. Structural virality represents a metric specifically developed for the measurement of diffusion of innovation in healthcare and validated against big data from the clinical setting (see Chapter 3).[39]

- Innovation index (ii_t): a measure of surgical innovation output calculated by

$$ii_t = \frac{\sum_{t\in S}\frac{1}{c}\times\frac{p_t^{org}}{P_t}}{\max_{t\in G}\left(\sum_{c\in S}\frac{1}{t}\times\frac{p_t^{org}}{P_t}\right)} \times 100,$$

where for a technological intervention $t, \in [1, \dots, S]$ is an integer value labeling the ordinal category associated with the implementation stage according to the level of evidence, p_s^t is the count of publications on this technological intervention appearing in stage s, and $P_s = \sum_c p_s^t$ is the total number of publications appearing in stage s across all technologies at a global scale (G). Similar to structural virality, the

innovation index represents a metric developed specifically for the measurement of innovation in healthcare, also validated against big data from the clinical setting (see Chapter 3).[39]

- Mortality-per-person-year: a measure of the number of events (deaths) divided by the amount of person-time (year) observed calculated by

$$m_{py} = \frac{E}{NT}$$

where for a study (article), the mortality-per-person-year m_{py} is the incidence rate or density (not a proportion) of event (death) E reported over person-time (year) (i.e. the product of persons N multiplied by time T).[157] Mortality rate using person-time (year) was chosen because individual patient data were not available and follow-up duration was variable between clinical studies. Mortality-per-person-year has been widely used in the epidemiological and meta-analytic literature when the patient follow-up duration varies (as was the case in this study).[158] Moreover, mortality (in general) constitutes the most commonly utilised 'hard' outcome metric for benchmarking and assessing cardiac interventions.[159,160]

- Altmetric score: a metric of 'visibility' used to track the attention that research output and datasets receive online. It pools data from a variety of online platforms that include social media (e.g. Twitter, Facebook and Google+), traditional media – both mainstream (e.g. The Times, BBC) and field-specific (e.g. New Scientist), blogs – both of major organisations (e.g. Cancer Research UK) but also of individual researchers, and online reference managers (e.g. Mendeley).[161] The altmetric score

cleans up and normalises the sources to quantify the "digital attention" an article receives, thus providing an alternative measure of online visibility.[162]

6.2.5 Statistical analysis

Mortality-per-person year was compared for different interventions (i.e. TAVI vs. aVR). An age-stratified analysis of distributions' differences by intervention was also performed for distinct age groups (categorised in three groups: <60, 61-80, >80 years) in view of the inherently different age of patients in the two groups. In view of the distributions not being normal as indicated by histograms (see Appendix 4), non-parametric statistical tests (Kruskal-Wallis and Wilcoxon signed-rank tests) were employed for comparisons. Moreover, a linear regression model has been fitted to the data to best represent the time trend of mortality-per-person-year by intervention.

To assess the nature of the relationship between mortality-per-person-year and structural virality on the one hand and altmetric score on the other, best-fitting models were used. Specifically, in the assessment of the relationship between mortality-per-person-year and structural virality, a polynomial model of degree 2 was employed, and for assessing the relationship between mortality per person-year and altmetric score a logarithmic model was employed (see Appendix 4 for details of model selection).

6.3 Results

The search revealed 2,884 articles published between 1965 (when the first animal study on percutaneous valve implantation for aortic valve disease was published)[163]

and December 2017. Of these, 289 articles were excluded as they evaluated aortic interventions in paediatric patients (n=73) or involved an additional procedure at the same time as the aortic intervention (n=216). Of the remaining articles evaluating different interventions for aortic stenosis and/or regurgitation (n=2,595), 469 articles were excluded as they did not receive a citation from or make a citation to another article within the dataset. The resulting 2,126 articles with at least one ingoing or outgoing in-field citation were used to construct the network (Fig. S1, Appendix 4). Of these, 1,879 articles made at least one in-field citation ("adopters" of innovation) and 1,306 articles received at least one in-field citation ("sources" of innovation).

6.3.1 Diffusion of innovations and innovation index

The number of TAVI cases performed worldwide (TAVI uptake) was found to follow an exponential growth curve starting from 2009 (**Fig. 21**). The innovation index for TAVI was significantly higher to that of aVR (see Appendix 4).

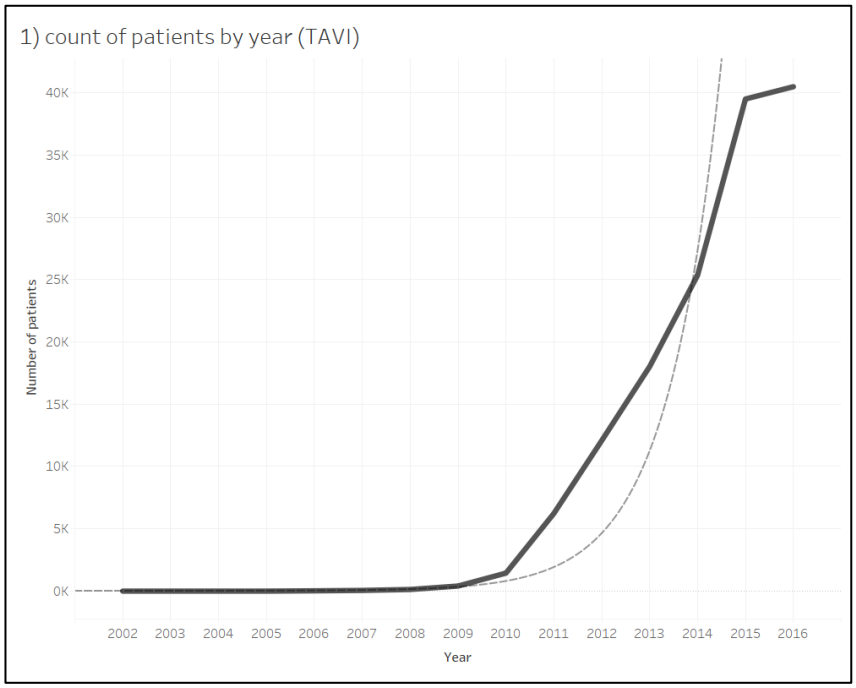


Figure 21Number of patients per year reported to undergo TAVI on a global scale based on peer-reviewed publications.

6.3.2 Mortality

The median mortality-per-person-year was significantly higher for TAVI compared to aVR (25.1 vs. 4.6, $p<0.05$). The mortality difference favouring aVR over TAVI remained statistically significant even when patients were stratified by age ($p<0.05$ for all age groups, Table S3 in Appendix 4). Mortality for TAVI decreased in a linear fashion until 2011 when it reached a plateau, albeit at a higher level to that for aVR, which remained stable throughout the same period (**Fig. 22**).

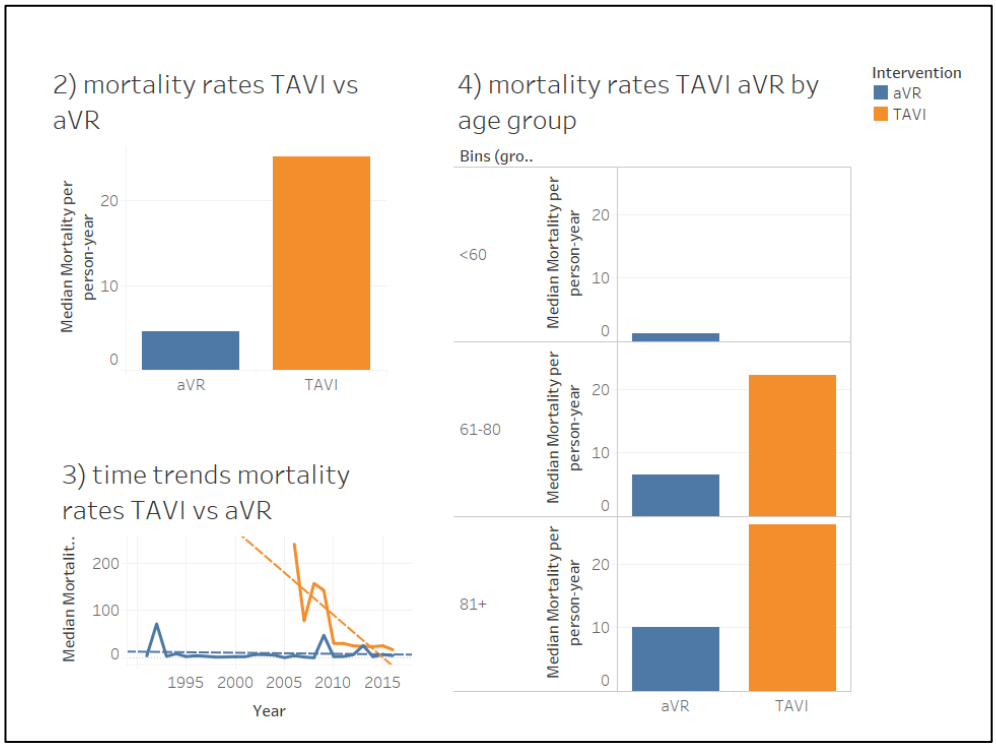


Figure 22 Mortality rates of TAVI vs. aVR (top left) and sub-group analysis by age (right).

6.3.3 Virality and its relationship to mortality

The virality of TAVI was significantly higher to that of aVR ($p<0.05$). When plotting virality against mortality, a U-shaped curve results for TAVI indicating that as its uptake became increasingly viral, following an initial drop and short plateau, mortality started to increase. On the other hand, virality did not demonstrate any relation with mortality for aVR, which remained constant and low throughout the study period. All findings were statistically significant ($p<0.05$) and are graphically presented in **Fig. 23** (the statistical tests used and exact p-values can be found in Appendix 4).

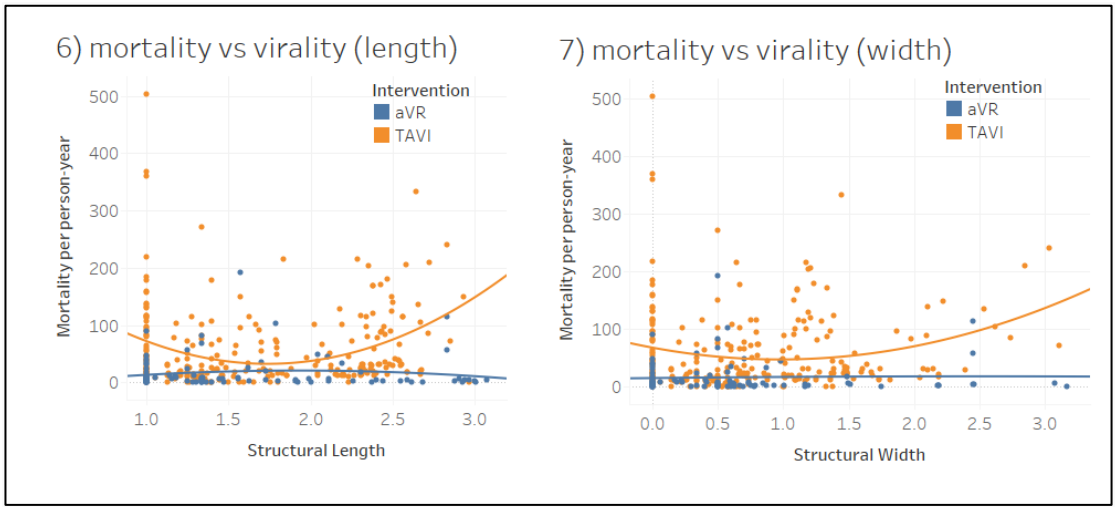


Figure 23 The relationship between virality and mortality for TAVI (orange) and aVR (blue).

6.3.4 Altmetric score and its relationship to mortality

The median altmetric score was significantly higher for TAVI than for aVR ($p<0.05$) indicating a significantly higher 'visibility' for TAVI. When plotting altmetric scores against mortality, a linear decreasing trend was observed for TAVI (i.e. the lower the mortality for TAVI reported by a paper, the more 'visible' that paper was in mainstream and online media) whilst for aVR there did not appear to be any relationship between the two (both findings were statistically significant, $p<0.05$). The above findings are presented in **Fig. 24** (the statistical tests used and exact p-values can be found in Appendix 4).

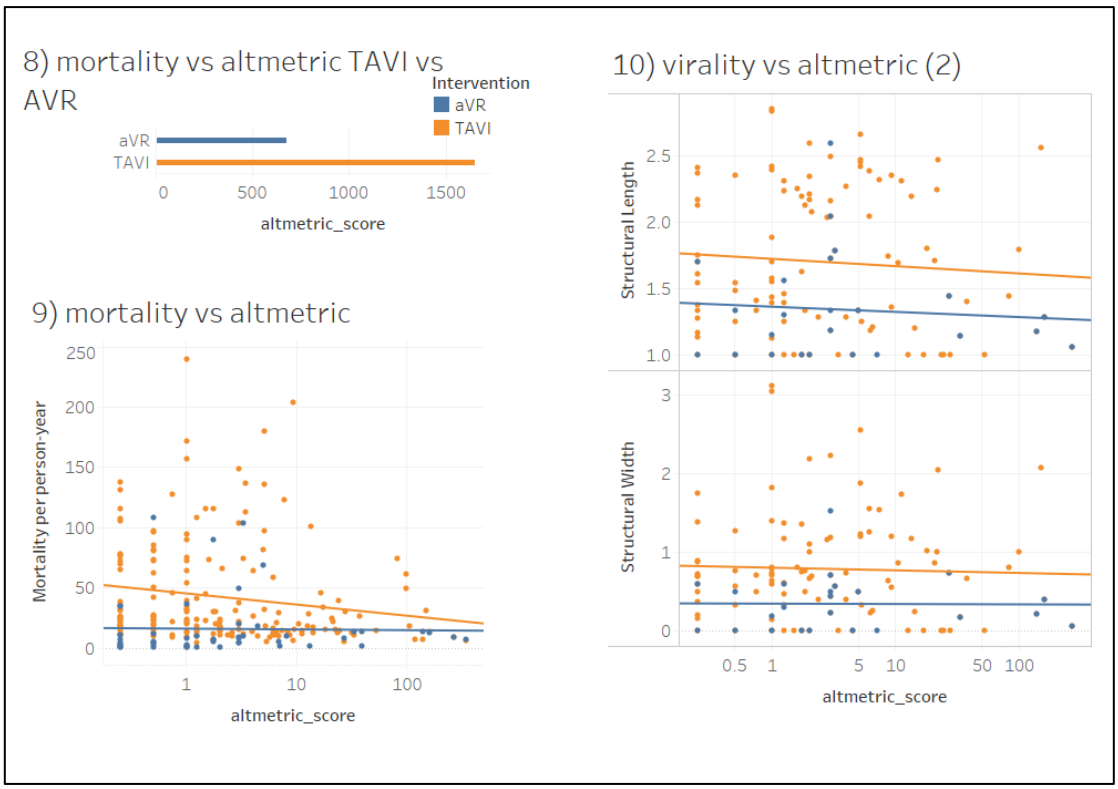


Figure 24 Altmetric scores for TAVI (blue) and aVR (orange, top left), the relationship between altmetric score and mortality (bottom left) and the relationship between virality and altmetric score (no relationship found).

6.4 Discussion

This study reports a number of important findings that illustrate the power of marketing in enhancing the diffusion of innovations in healthcare and ensuring translation to market success. Using the example of TAVI, a topical and highly successful innovation in terms of clinical uptake and sales, it also demonstrates that marketing can drive diffusion independent of the evidence.

TAVI was specifically chosen as the index innovation precisely because of the conflicting (and vast) literature regarding whether or not it truly represents a disruptive innovation. Though TAVI has undoubtedly attracted a great interest in the

cardiovascular field generating numerous multi-centered RCTs, there remains a lack of consensus around its indications and comparative effectiveness. This is in complete contrast to its unparalleled clinical uptake and is what makes TAVI a unique case study meriting an in-depth analysis to understand how "practice has gone beyond the evidence".[164]

Proponents of TAVI attribute its unprecedented clinical uptake to it being a "breakthrough technology" and a "disruptive innovation".[154,165] However, this view is not widely shared since TAVI has been associated with both higher complication rates and a lack of cost-effectiveness over aVR, the incumbent market leader.[164]

To understand whether TAVI truly represents a disruptive innovation, it is imperative to first consider the definition of the term. Contrary to most innovations that are incremental in nature, disruptive innovations represent "simpler, more convenient, and less costly alternatives initially designed to appeal to a small or insignificant market" (**Table 4**).[166] At first, disruptive products/technologies may appear unattractive but they eventually displace established competitors in a bottom up manner. Market disruption ensues through the provision of a *low-cost alternative* of *acceptable performance* becoming the new market leader.[167,168]

The characteristics of disruptive innovation
Simpler, cheaper, and lower performing
Promise lower margins, not higher profits
Often initially shunned by leading customers who can't use or don't want them
First commercialised in small or insignificant markets

Table 4 The characteristics of disruptive innovation as defined by Christensen.[167,168]

Looking at how TAVI may fit into this definition, TAVI constitutes a simpler and more convenient alternative to conventional aVR in view of its percutaneous nature. However, it is neither less costly nor was it designed to appeal to a small or insignificant market. On the contrary, TAVI has consistently been shown to be less cost-effective than aVR, even for high-risk patients.[169] This group, which along with 'inoperable' patients constitute the main target groups for TAVI, represent over a third of all patients with symptomatic aortic stenosis[170] making the TAVI market highly significant.[165] Finally, the similar (but not superior) survival rates between TAVI and aVR, coupled with the higher rates of disabling stroke, major vascular complications, and moderate or severe aortic regurgitation (for TAVI) as indicated by large, multi-centre RCTs,[171,172] make it difficult to describe its performance as 'acceptable' in the presence of an established, safer, and cheaper alternative (aVR).

It thus becomes apparent that TAVI does not constitute a disruptive innovation. What it represents is an incremental innovation to pre-existing percutaneous cardiac interventions (e.g. angioplasty, stenting), which themselves were disruptive as they significantly altered the degree of invasiveness for coronary artery revascularisation whilst offering the same degree of protection against death, CVA and myocardial

infarction to coronary artery bypass grafting (CABG) at a lower cost ('simpler and cheaper' than CABG).[173] These facts, widely supported by level I evidence, justify describing the performance of percutaneous angioplasty and stenting as 'acceptable' despite the greater need for repeat revascularisation compared to CABG ('lower performing' than CABG).[173]

Despite not constituting a disruptive innovation, TAVI has exhibited a unique diffusion rate in a field (healthcare) where innovations are known to diffuse very slowly. This is due to regulatory, financial, and ethical barriers established to safeguard patient safety and control costs at the various stages of the innovation process.[43,52,174] As shown in **Fig. 21**, the clinical uptake of TAVI has been exponential and continues unopposed to date. This is despite serious concerns having been raised over the years, especially in relation to its complication profile and costs.[164,175] Therefore, clinical factors cannot (solely) account for the clinical uptake of TAVI.

In the present era of evidence-based medicine, evidence (with RCTs at the top of the hierarchy) is used to guide clinical practice and, in conjunction with cost considerations, inform policy.[43] This does not appear to be the case with TAVI. As shown in this and other studies,[176,177] its exponential uptake started in 2009, preceding the publication of the first multi-centre RCTs supporting its use (PARTNER trial cohorts A and B).[170,172] Even as early as 2004, TAVI already occupied a prominent market share that reached up to 20% of all aortic valve procedures in Western Europe.[165]

In addition to the time lag (of years) between evidence and clinical implementation (diffusion), the 'landmark' RCTs (PARTNER cohorts A and B) commonly cited to justify the wider implementation of TAVI have come under heavy criticism. Reasons include publication bias, lack of transparency, unbalanced patient characteristics (favouring TAVI), and failure to completely declare conflicts of interest by the principal investigators.[164] It is also imperative to note that these RCTs were all designed to assess non-inferiority as opposed to superiority.[172]

Most importantly, the survival advantage reported in favour of TAVI in the 'landmark' RCTs only related to the subgroup of patients deemed inoperable (n=358), i.e. where TAVI was compared to conservative treatment (medical management and/or balloon aortic valvuloplasty).[170,172] In the original cohort (n=699) where TAVI was compared 'head-to-head' against aVR, the survival rates were found to be similar both at 1- and 5-years, not superior.[171,172] At the same time, complication rates relating to disabling CVA and major vascular complications were consistently higher in the TAVI group compared to both conservative treatment and aVR.[170,172]

The phenomenon of choosing suboptimal comparators (known as 'straw man comparators') when evaluating a new technology/treatment, is widely prevalent in industry-sponsored RCTs,[178] which represent virtually all high-quality RCTs evaluating TAVI.[164] Naturally, such comparisons increase the likelihood of obtaining favourable results (for the new product/intervention) and are thus preferred by sponsors over 'head-to-head' comparisons against established treatments (the 'gold standard').[178] TAVI is no exception to this (a survival benefit was only demonstrated against conservative management and not aVR).[170,172]

At present, the major concern with TAVI is that "indications are slipping".[179] Large registry data from both sides of the Atlantic reveal that TAVI is now widely performed in intermediate-risk and younger patients in the absence of reliable data regarding long-term valve durability (when it is known from surgical valves' research that a minimum follow-up of 10 years would be required).[180] This off-label performance of TAVI in lower-risk patients has formed a significant proportion of clinical practice for years; long before the SURTAVI trial (that compared TAVI with aVR in intermediate-risk patients) was published.[176,181] Similar to PARTNER, SURTAVI was also designed as non-inferiority and sponsored by a leading TAVI manufacturer (Medtronic, Inc.), albeit different to that of PARNER (Edwards Lifesciences Corp.)[181]

The apparent implementation-evidence mismatch advocates the involvement of non-clinical factors (i.e. marketing) in the diffusion of TAVI.[164,180] The negative and statistically significant association identified between the mortality reported for TAVI per study and its altmetric score further reinforces this hypothesis, especially as no similar relationship was found to exist for aVR. As altmetric score represents an increasingly recognised tool for measuring the real-time reach and influence of articles,[162] this negative association denotes a bias towards selectively promoting those studies favouring TAVI across both mainstream and online media. This is also indicative of a shift in the manufacturers' marketing strategy where in addition to targeting physicians and other healthcare providers involved in the purchasing of their products (*business-to-business model*), they also target directly patients and the public (*business-to-consumer model*).[182,183]

The healthcare sector with the strongest marketing component remains the pharmaceutical industry where one of the most powerful predictors of new drug uptake has been shown to be a high exposure to marketing.[184] This study's findings indicate that marketing is now playing an increasingly important role in the medical devices sector too, in this case in promoting the diffusion of TAVI. The new marketing strategies aimed at the patients and public directly are likely to be crucial in generating the increasing demand seen, with a recent study showing that patient perception is a key factor when it comes to opting for TAVI over other treatments.[185]

Diffusion represents the final phase of the innovation value chain and, from the manufacturers' perspective, constitutes a key metric of success (both sales and profits are directly linked to diffusion).[35,39] However, in healthcare, for innovation to generate value, its diffusion must also be shown to improve patient outcomes and/or reduce costs.[32] As TAVI lacks cost-effectiveness over aVR,[169] the relationship between diffusion and mortality was evaluated, to see whether the diffusion of TAVI has been associated with a reduction in mortality rates.

As illustrated in **Fig. 23**, initially, the diffusion of TAVI was found to be associated with a reduction in mortality. This likely reflects the fact that the introduction of TAVI addressed a previously unmet clinical need (for inoperable patients with aortic valve disease).[154] However, following this initial period, as the diffusion of TAVI continued unopposed, mortality started to increase. This second phase corresponds to the 'market disruption' period where TAVI (transfemoral and transapical) started taking over ('disrupting') the market of aVR through its increasing popularity

resulting in lower-risk and younger patients (suitable for aVR) to opt for TAVI despite no demonstrable survival benefit or cost-effectiveness over aVR.[179]

Another technology that resembles TAVI in terms of the imbalance between its diffusion and evidence base is robotic surgery, also commonly (and erroneously) labeled as a 'disruptive innovation' when it actually constitutes an incremental innovation based on laparoscopic surgery,[27] a truly disruptive innovation, which forever changed the surgical landscape.[155] Not only is robotic surgery significantly more expensive than laparoscopic surgery but more importantly, robotic-assistance has failed to translate to superior patient outcomes over the laparoscopic approach (with the possible exception of robotic prostatectomy).[6,27] Yet, robotic surgery is experiencing a marked diffusion across multiple surgical specialties raising again the issue about marketing (rather than evidence) driving its diffusion.[6]

The final finding that merits mentioning relates to the mortality of TAVI, which contrary to published randomised data reporting it to be similar to that of aVR (non-inferior),[171,186] was shown to be significantly higher in this study even after controlling for age. This discrepancy is likely to be the result of pooling all published data (trials and registry data), as opposed to confining the analysis to RCTs only. By looking at the entire evidence spectrum, the findings presented here are likely to be more representative of the 'real world'.[187]

Before concluding, it is important to consider the strengths and limitations of this study. The two key strengths relate to the type of data pooled and the analysis employed. Data originating from a variety of heterogeneous sources were extracted

and linked. These extended beyond the clinical setting to alternative metrics in view of the existing imbalance between evidence and implementation. Moreover, this is the first study to employ network analysis to provide an in-depth quantitative assessment of the diffusion of TAVI beyond simply evaluating trends in terms of numbers of cases performed per annum globally (which was also done). Network-based diffusion metrics (structural virality) and visibility metrics (altmetric score) were incorporated in the analysis and their relationship to clinical outcomes evaluated, something not previously performed in this context.

Limitations include the fact that market diffusion metrics such as sales and profits were not taken into consideration, and similarly clinical outcomes beyond mortality were not considered. However, in view of the novelty of the study in terms of the data breadth, their linkage, and subsequent network analysis, it was felt that incorporating more metrics would substantially complicate the study without adding a further dimension to those already studied (diffusion, visibility, and patient outcomes). Another limitation relates to the fact that the pooling of data was at the study (or registry) level (as reported in each publication), and although we acknowledge that individual patient data would have significantly strengthened the findings, the shear volume of the data used (exceeding 2,000 publications reporting on more than 160,000 patients) make the findings robust.

Finally, there are limitations associated with the choice of TAVI as the index innovation. These relate to the fact that TAVI was considered as a single intervention when actually a variety of valves exist with potentially distinct characteristics, depending on multiple factors including their manufacturer and generation. In

addition, different routes can be used for valve implantation, each with its distinct indications and risk profile. Though this is true, the aim of this study was to evaluate the diffusion of TAVI as a healthcare innovation overall rather than perform specific comparisons between different devices and manufacturers. These limitations, which relate to product heterogeneity, are by no means unique to TAVI but apply to the evaluation of most novel healthcare interventions, especially when it comes to innovative medical devices.[188,189] Importantly, despite the large number of existing manufacturers and different implantation routes available, TAVI represents a relatively 'homogeneous' intervention, with only a handful of manufacturers dominating the market and the transfemoral route constituting the commonest route employed by far.[190,191]

6.5 Conclusion

This study has shown that the unparalleled diffusion of TAVI appears to be the result of highly successful marketing strategies rather than the evidence base. Based on the findings, it would be advisable to halt the continuing expansion of TAVI indications to lower-risk and younger patients until longer follow-up results of current trials and international registries become available. In addition, the widespread use of TAVI in high-risk operable patients should be scrutinised more thoroughly since the evidence to date has failed to demonstrate clinical superiority or cost-effectiveness over aVR.

6.6 Chapter summary

The objective of this study was to examine the factors associated with the unparalleled diffusion of TAVI, with a particular attention to the evidence base and marketing strategies employed.

Network and regression analyses were used. All pre-clinical and clinical original research studies (n=2,595) reporting on the evolution of the innovation process in the treatment of 'aortic valve stenosis' and 'aortic valve regurgitation' in the *Web of Science*® database were included. These encompassed 160,892 patients undergoing treatment for aortic valve disease in the period between 2002 (when the first live human TAVI case was performed) and December 2017 (time of data extraction).

TAVI constituted the primary focus as the index 'disruptive' innovation. Aortic valve replacement (aVR), the current gold standard, was used as the control intervention. The main outcome measures included the number of cases performed per annum globally, structural virality (a validated metric of diffusion of innovation in healthcare), innovation index, mortality-per-person-year, and altmetric score (a metric of research 'visibility' in mainstream and online media).

It was shown that the number of TAVI cases performed worldwide is following an exponential growth curve that started in 2009 preceding the publication of the first multi-centre RCTs supporting its use. The median mortality-per-person-year is significantly higher for TAVI compared to aVR (25.1 vs. 4.6, $p<0.05$) even when patients are stratified by age. Both the virality and media visibility of TAVI are

significantly higher to aVR ($p<0.05$). The relationship between virality and mortality for TAVI follows a U-shaped curve (no relationship is evident for aVR). A linear decreasing trend is observed between mortality and visibility for TAVI (no relationship is evident for aVR).

In summary, the unparalleled diffusion of TAVI and continuous expansion of its indications to progressively lower-risk and younger patients appear to be the result of highly successful marketing strategies targeted directly at patients and the public rather than the evidence base.

The next chapter, Chapter 7, represents the final study comprising this thesis. It examines innovation from a regulatory perspective. This is crucial as the balance that regulation aims to achieve between ensuring patient safety and promoting innovation can be especially fine. Using TAVI as an example of high-risk innovation in healthcare, different regulatory frameworks across the world are evaluated and solutions proposed towards the development of a regulatory ecosystem promoting medical device innovation without compromising on the safety of patients and the public. To ensure the full patient safety profile was evaluated, a number of complications beyond mortality were also assessed. These included 30-day rates for disabling CVA, major vascular complications, and the development of post-TAVI paravalvular leak.

7. The innovation-regulation conundrum: A network perspective

7.1 Introduction

The principal motivation behind the regulation of healthcare innovation relates to ensuring the safety of patients and the public. In their attempt to do so, the various regulatory frameworks and accompanying legislation across the world share two common objectives: i) safeguarding public health and ii) ensuring innovative technologies reach patients promptly.[192] However, this can prove exceptionally challenging in practice as the two can often be in direct contradiction, especially when it comes to expediting market access for very promising but high-risk innovations, such as implantable medical devices.

The fast pace and large numbers of innovations arising and evolving on a continuous basis across the wide breadth of the various healthcare sectors can (and commonly does) outpace the ability of regulatory bodies to establish their safety and effectiveness in a timely manner (i.e. prior to market entry). This poses a significant problem in need of urgent attention, as the risk is that innovations may become outdated before they even reach patients. Not only can this diminish the potential value that innovations could offer to patients (and the wider healthcare systems) but it can also act as a strong deterrent to existing and aspiring innovators. The latter is a particularly serious concern as in the long-term it can adversely impact on the economy and society as a whole. [193]

Innovation not only provides the foundation for progress in healthcare,[13] it also constitutes the core ingredient for prosperity and growth at all societal levels (ranging from that of the individual all the way to national level), both core determinants of global health.[194] It is thus imperative that innovation is enhanced to the maximum. Investing more (not only financially, but also in terms of time and human capital) is crucial, but so is recognising and systematically addressing existing barriers. One of the key barriers to innovation relates to certain aspect of regulation.[195] This is especially true in healthcare where regulatory requirements may on occasion impede the timely access of patients to novel diagnostics and treatments.[196]

Using the timely example of TAVI, the most dominant innovation in the field cardiovascular medicine and surgery on a global scale, the final study of this thesis aims to evaluate the effect of regulation on both innovation and patient safety. Innovation metrics and clinical outcomes for geographical regions with distinct regulatory frameworks were measured and compared in an attempt to elicit the barriers and catalysts to innovation and understand how these can impact on patient safety. This information is likely be of distinct value to all stakeholders - including governments, regulatory bodies, policy makers, clinicians, and patients - in helping to inform future policymaking and point to areas where legislative changes are most likely to optimise the (fine) balance between driving innovation and ensuring the safety of patients and the public.

7.2 Materials and Methods

7.2.1 Dataset

The *Web of Science*® (WOS®) database (Clarivate Analytics, Philadelphia, PA) was used. The search was performed on 2nd June 2017 and all publications in the *Topics* "aortic valve stenosis" and "aortic valve regurgitation" were retrieved. The search was then refined to a subset of categories including "cardiac cardiovascular systems" and "surgery" (see Appendix 5 for details) and focused on peer-reviewed articles only.

Both pre-clinical and clinical original research studies were included to study the evolution of the innovation process at all stages from idea generation all the way to multi-centre RCTs, international registry reports, and the formulation of guidelines. Publications not reporting original research studies (e.g. review articles, book chapters, editorials, letters, patents) were excluded through filters (limiting to 'article' in *Document Types*). Exclusion criteria also involved studies on paediatric patients (<18 years) and studies where additional procedures to the aortic valve treatment were performed in the same setting (e.g. TAVI with concurrent coronary angioplasty).

7.2.2 Data classification

Another researcher and I critically appraised and classified independently the data for each study: level of evidence, country of study origin (see below), type of intervention (e.g. TAVI, aVR), and for each intervention arm: number of patients, mean age (and standard deviation), number of patient deaths and follow-up time at which mortality was reported, and number of patients developing a disabling cerebrovascular accident (CVA), major vascular complications, and moderate-severe paravalvular leak based

on echocardiography (for the last three outcomes 30-day rates were extracted). Any discrepancies were resolved by consensus.

In terms of the geographical classification, this was based on the geographical origin of the corresponding author's affiliated institution. For each article, a unique geographical region was attributed among the following six: Africa, Asia, Central and South America, Europe, Middle East, North America, and Oceania (see Appendix 5).

7.2.3 Citation network and diffusion cascades

In the citation network, the nodes represent the articles, and a directed link is established from each citing article to all other articles cited in its bibliography addressing the same topic.[156] Through the citation network, it is possible to map the diffusion cascade along which information spreads and adoption of innovation propagates. Moreover, the citation-based chain of adoption permits the tracing back to the seed node representing the original article in which a given idea was introduced (e.g. first description of a new approach for TAVI) or initially trialed (e.g. the first human trial of a novel device, e.g. balloon-expandable TAVI). The structure of these cascades can therefore shed light on the diffusion process through which innovations, once proposed, built momentum over time (both in terms of diffusion but also implementation based on the sequential innovation stages reached – see innovation metrics section below).[35,39] Only articles that made or received at least one citation to/from another article in the dataset were used to construct the network (since citations are represented by the links between nodes).

7.2.4 Clinical outcome measures

Four clinical outcome measures were used:

-Mortality-per-person-year: a measure of the number of events (deaths) divided by the amount of person-time (year) observed calculated by

$$m_{py} = \frac{E}{NT}$$

where for a study (article), the mortality-per-person-year m_{py} is the incidence rate or density (not a proportion) of event (death) E reported over person-time (year) (i.e. the product of persons N multiplied by time T).[157] Mortality rate using person-time (year) was chosen because individual patient data were not available and follow-up duration varied between studies. Mortality-per-person-year has been widely used in the epidemiological and meta-analytic literature when follow-up duration varies among studies.[158] Moreover, mortality (in general) constitutes the most commonly utilised 'hard' outcome for benchmarking and assessing cardiothoracic interventions.[159,160]

- Disabling stroke: percentage of patients developing a disabling CVA at 30 days

- Major vascular complications: percentage of patients developing major vascular complications at 30 days

- Moderate or severe paravalvular leak: percentage of patients developing moderate or severe paravalvular leak (as evidenced on echocardiographic evaluation) at 30 days

7.2.5 Other performance measures

Additional performance measures were employed to capture innovation, academic performance, and media visibility for each research article. These were:

- Innovation metrics

- Innovation index (ii_t): a measure of surgical innovation output calculated by

$$ii_t = \frac{\sum_{t \in S} \frac{1}{c} \times \frac{p_t^{org}}{P_t}}{\max_{t \in G} \left(\sum_{c \in S} \frac{1}{t} \times \frac{p_t^{org}}{P_t} \right)} \times 100,$$

where for a technological intervention t, $\in [1, \ldots, S]$ is an integer value labeling the ordinal category associated with the implementation stage according to level of evidence, p_s^t is the count of publications on this technological intervention appearing in stage s, and $P_s = \sum_c p_s^t$ is the total number of publications appearing in stage s across all technologies on a global scale (G). The innovation index represents a metric specifically developed for the measurement of innovation in healthcare and has been validated against big data from the clinical setting (see Chapter 3).[39]

- Structural Virality (sv_{C_s}): a network-based measure of diffusion of innovation calculated by

$$sv_{C_s} = \frac{1}{l} \sum_{i \in C_s} \sum_{j \in C_s} d_{\vec{i,j}}, i \neq j,$$

where for $n>1$ articles, for each seed article s, structural virality sv_{C_s} of

cascade C_S is defined as the average length of the shortest directed paths between pairs of articles in C_s. $d_{\overrightarrow{i,j}}$ denotes the length of the directed shortest path from article *i* to article *j*, C_s is the set of nodes belonging to the cascade originating from article *s*, and *l* is the number of directed paths connecting pairs of articles in C_s. So constructed, sv_{C_s} becomes larger as the adopters of the original idea proposed by seed article *s* are farther apart from one another and from *s*, thus producing a multi-generational cascade. Similar to the innovation index, structural virality represents a metric specifically developed for the measurement of diffusion of innovation in healthcare also validated against big data from the clinical setting (see Chapter 3).[39]

- Academic success

  - Citation numbers: the traditional measure of academic success measured over the citation network.[197] This corresponds to the in-degree of an article, i.e. the number of incoming links (citations) from the other articles in the network. See Appendix 5 for a more formal description of this measure.

- Media visibility

  - Altmetric score: a metric of 'visibility' used to track the attention that research output and datasets receive online. It pools data from a variety of online platforms that include social media (e.g. Twitter, Facebook and Google+), traditional media – both mainstream (e.g. The Times, BBC) and field-specific (e.g. New Scientist), blogs – both of major organisations (e.g. Cancer Research UK) but also of individual researchers, and online

reference managers (e.g. Mendeley).[161] The altmetric score cleans up and normalises the sources to quantify the "digital attention" an article receives, thus providing a measure of visibility across mainstream and online media.[162]

7.2.6 Statistical analysis

Following the ranking of geographical regions in a decreasing order for each of the clinical, innovation, academic, and media visibility metrics, the one-sided Wilcoxon signed-rank test was used to perform pair-wise comparisons (non-parametric) in view of their skewed distributions (see histograms in Appendix 5). Given the scarcity of observations for Asia, Central and South America, Middle East, and Oceania (see Results), no meaningful comparisons could be performed for these four regions, and these four regions were thus excluded from further analysis. Hence, the pair-wise comparisons focused between Europe and North America, which represented the two groups with the largest number of observations. Statistical analyses were carried out with R version 3.5.1 (Bell Laboratories, Berkeley Heights, NJ).

7.3 Results

The search revealed 2,884 articles. Of these, 289 articles were excluded as they evaluated aortic interventions in paediatric patients (n=73) or involved an additional procedure to the aortic intervention in the same setting (n=216). Of the remaining articles evaluating different interventions for aortic stenosis and/or regurgitation (n=2,595), 469 articles were further excluded as they did not receive a citation from or make a citation to another article within the dataset. The resulting 2,126 articles with at least one ingoing or outgoing in-field citation were used to construct the

network (Appendix 5). Of these, 1,879 articles made at least one in-field citation ("adopters" of innovation) and 1,306 articles received at least one in-field citation ("sources" of innovation). The search strategy is illustrated in **Fig. 25**.

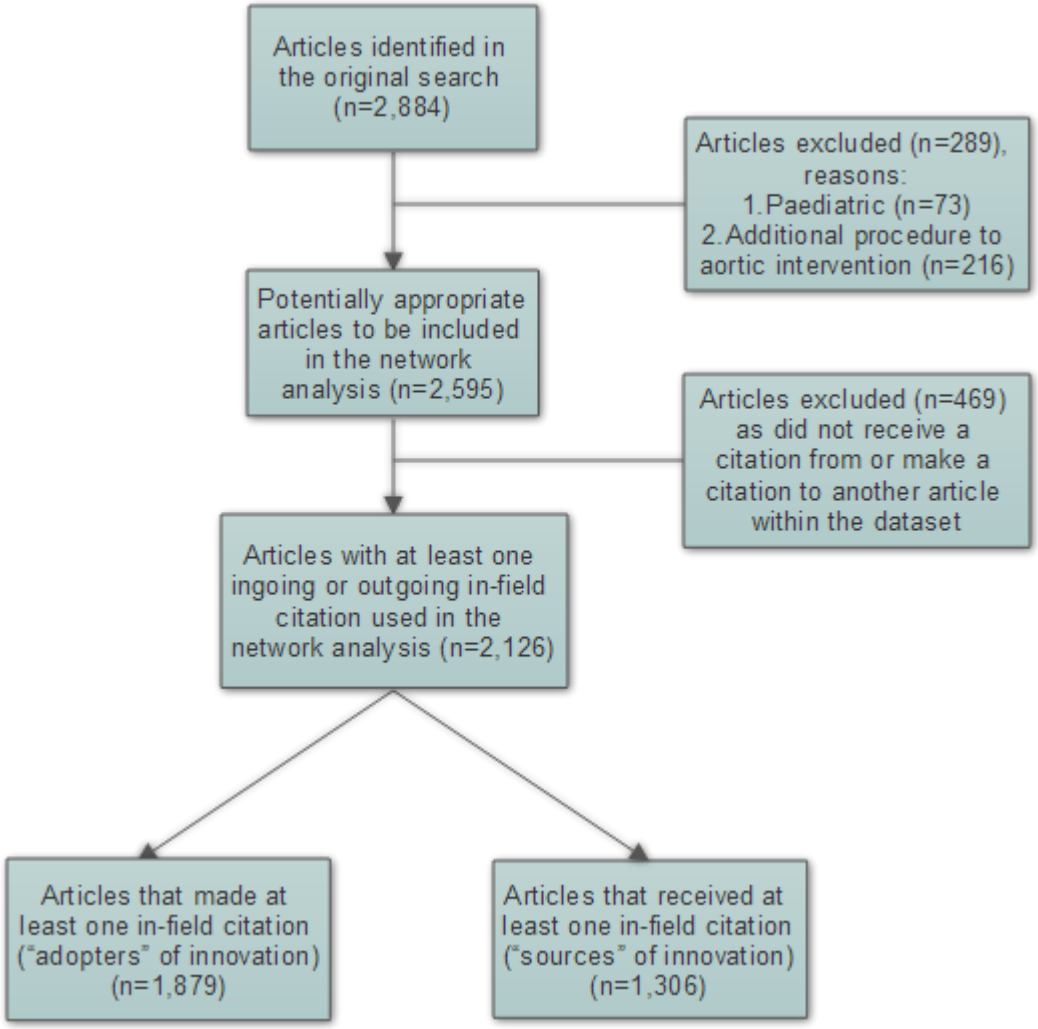


Figure 25 Flow diagram illustrating the search strategy used for generating the citation network.

The network consisting of 2,126 nodes (articles) joined by 10,491 links (citations) is depicted in **Fig. 26**. In it, 728 articles reporting studies on TAVI are included, 473 of which originate from Europe, 186 from North America, 28 from Asia, 15 from the Middle East and 13 from Central and South America.

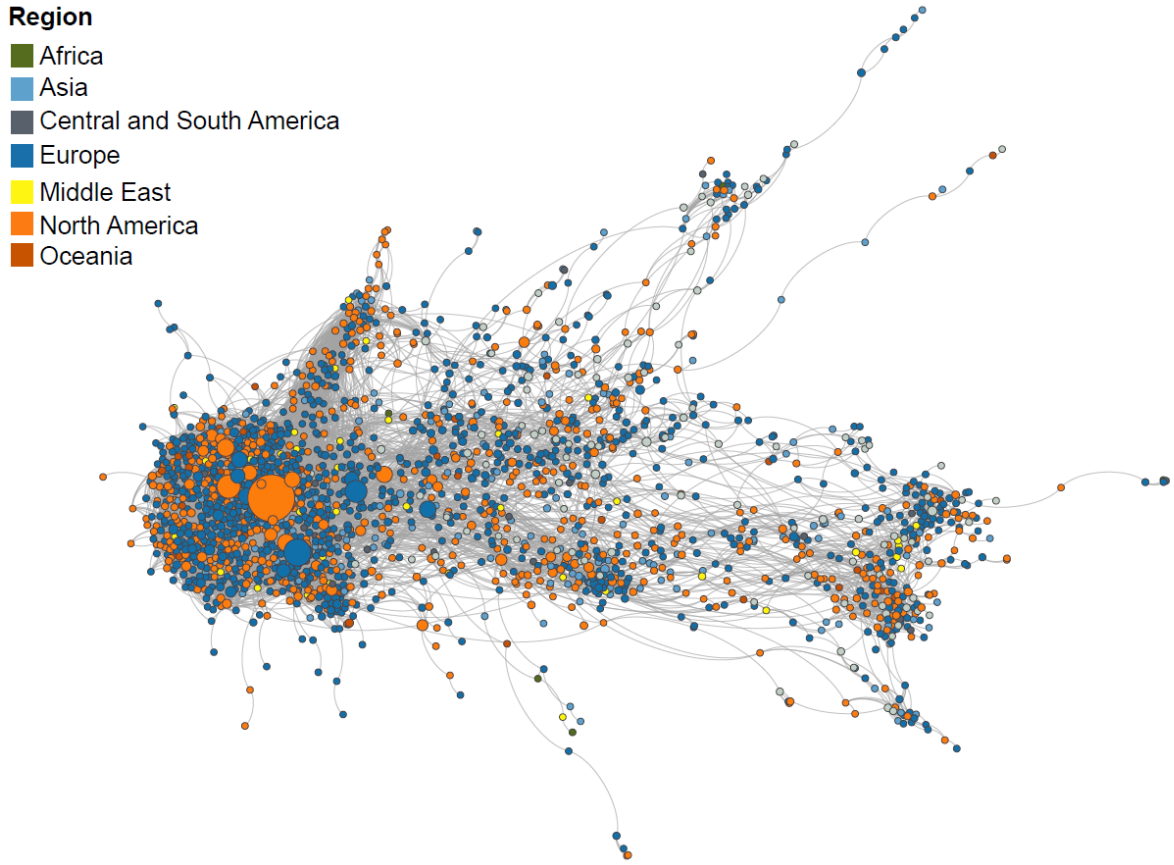


Figure 26 The giant component (2,006 articles, 10,491 citation links) of the citation network comprising articles concerning "aortic valve stenosis" and "aortic valve regurgitation". Nodes are coloured according to assigned region of the corresponding author's affiliation (see top left panel for geographical colour coding) and sized according to the number of citations received.

Rankings for all outcomes are illustrated in **Fig. 27**. Europe and North America consistently perform superiorly in terms of all clinical outcomes (i.e. they exhibit the lowest mortality and complication rates) whereas the Middle East ranks first for mortality and all other complications (i.e. highest reported mortality and complication rates) except for disabling CVA where Oceania ranks first. However, as explained in the Materials and Methods section, the scarcity of observations for Asia, Central and South America, Middle East, and Oceania, did not permit meaningful comparisons for these four regions. When focusing on the pair-wise comparison between Europe and North America (**Fig. 28**), Europe ranks higher for mortality than North America

though this difference was not statistically significant (p=0.470). On the other hand, in terms of all three other complications evaluated, North America ranks higher than Europe, though the only difference reaching statistical significance was that relating to disabling CVA (p=0.02 for disabling CVA, p=0.149 for major vascular complications, and p=0.134 for moderate or severe paravalvular leak).

When looking at academic success and media visibility metrics - as measured by citations and altmetric scores, respectively - North America ranks significantly better than Europe for both (p=0.021 and p=0.004 respectively). However, in terms of the actual diffusion of innovation (TAVI), Europe exhibits superior performance (than North America) based on both the volume of patients undergoing TAVI as well as its structural virality (though the latter finding was not statistically significant; p=0.342).

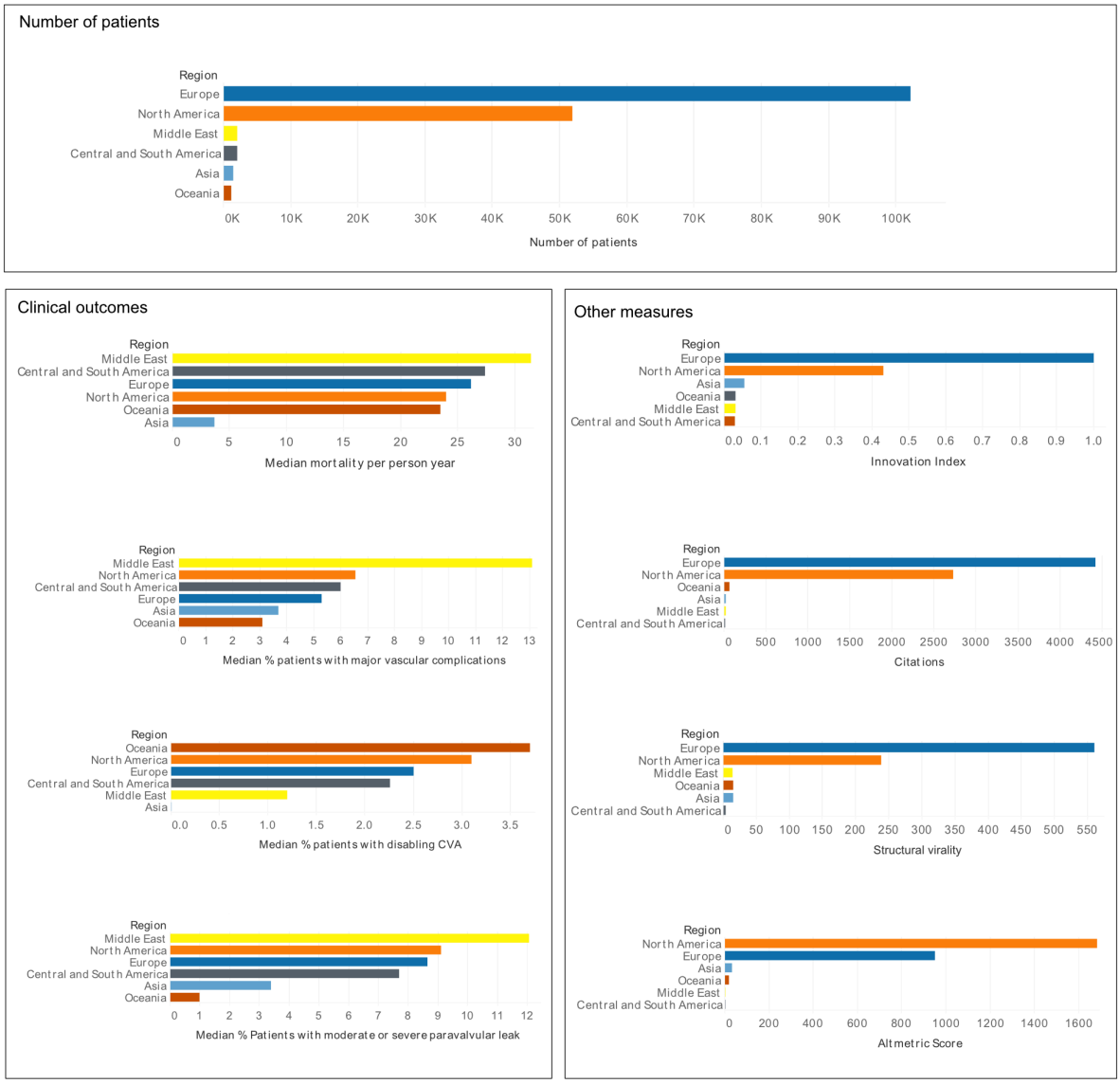


Figure 27 Bar charts indicating the total number of patients per region (top panel), the median values of clinical outcome measures including mortality-per-person-year, and percentage of patients developing major vascular complication, disabling cerebrovascular accident (CVA) and moderate or severe paravalvular leak at 30 days post-TAVI (left-hand panel), and other measures including the innovation index, and the sum of citations, structural virality and altmetric scores by region (right-hand panel).

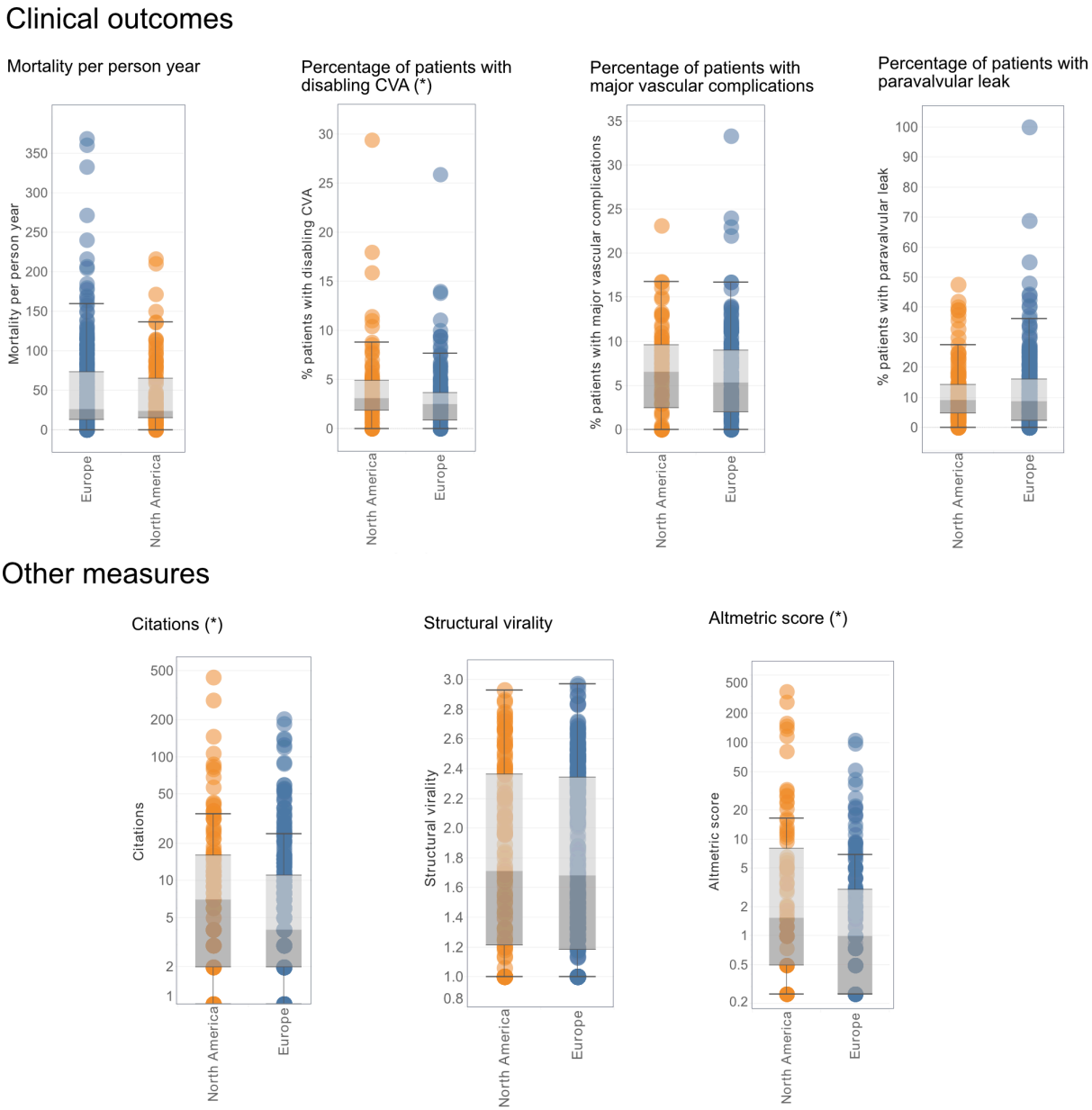


Figure 28 Boxplots of distributions of clinical outcome and other performance measures for North America and Europe. Shaded areas represent the interquartile range whilst the darker areas represent the distance between the 25th percentile and median values. Asterisks denote significance (p<0.05).

7.4 Discussion

This study evaluating healthcare innovation and patient safety in the context of TAVI across different geographical regions - each with its own regulatory framework - reports a number of important findings. Prior to discussing these, it is pertinent to briefly consider some of the key elements of the world's oldest and most established

regulatory frameworks for medical devices, namely those of the EU and US. This will provide the context for interpreting the findings and recognising existing deficiencies in need of modification. This study comes at a critical time point as both the EU and US are in the process of reviewing and updating their existing regulatory frameworks for medical devices.

In the EU, for a medical device to gain market access, it must be granted a CE (Conformité Européenne) mark. This can only be issued by Notified Bodies, which are independent commercial organisations dispersed across the EU and monitored by the Competent Authority of the member state in which they reside. In the UK, the Competent Authority is the Medicines and Healthcare Products Regulatory Agency (MHRA).[198] Once a CE mark has been granted, the medical device can be marketed across all EU member states.[199]

Medical devices are categorised by classes (and sub-classes) according to their risk profile, with Class I representing those associated with the lowest risk and Class III those with the highest risk (e.g. stethoscopes and TAVI, respectively).[199] The two key aspects that a Notified Body is looking for when it comes to assessing a medical device's suitability for CE marking, are its *safety* and *performance* (i.e. that it performs as intended) – but not necessarily its *effectiveness* (i.e. demonstration of clinical benefit).[192,193] Even for moderate- (Class IIb) and high-risk implantable (Class III) medical devices where clinical (in addition to pre-clinical) data are mandatory, these do not need to be randomised. The exact requirements remain elusive as they are not available in the public domain.[192]

On the contrary, in the US, regulation is centralised through the Food and Drug Administration (FDA) and the evidence requirements are clearly defined for each medical device class. For medium-risk (Class II) devices and those shown to be 'substantially equivalent' to pre-authorised devices already available in the market, the 510(k) route of *substantial equivalence* is most commonly followed.[193] For innovative high-risk (Class III) medical devices, the FDA expects to be presented with RCT data that demonstrate not only their safety but also their effectiveness (for the intended use).[164] This involves a rigorous process known as Pre-Market Authorisation (PMA).[192]

As a result, medical devices (including innovative, high-risk, and/or implantable ones) commonly gain their CE mark (and thus enter the EU market), long before they are granted FDA approval - with TAVI representing a typical example of this. Both the SAPIEN valve (Edwards Lifesciences Corporation, Nyon, Switzerland) and the Core Valve (Medtronic, Inc., Minneapolis, MN) gained their CE marks in 2007 in the absence of any RCT. In the US, it took another four years for SAPIEN (2011) and five years for Core Valve (2012) to be granted approval by the FDA (as a result of the FDA's more rigorous evidence requirements).[164]

In view of the above, the EU regulatory framework for medical devices has come under intense criticism with opponents arguing that the existing system is inadequate, lacks rigor (compared to FDA) and is thus putting patients at risk.[200] One of the most commonly cited examples to support these claims relates to the transvaginal mesh implant (Ethicon Gynecare TVT Sling, Johnson & Johnson, New Brunswick, NJ) scandal where following market approval, these implantable devices were found to be

associated with unacceptably high complication rates (including deaths), which in turn resulted in an unprecedented rise in lawsuits and subsequent need for explantation in thousands of women.[201]

Another widely publicised tragedy that is also commonly cited is that involving the PIP breast implants (Poly Implant Prothèse, La Seyne-sur-Mer, France).[202] These scandals have understandably led to major concerns regarding the regulation of medical devices and its impact on patient safety. Using the topical example of TAVI, one of the principal innovations in the medical devices sector at present,[154] this study aimed to evaluate the impact of different regulatory frameworks (and in particular the world's two most established ones) on patient safety but also on the diffusion (adoption) of innovations and on innovativeness (i.e. whether they enhance or hinder innovation in healthcare). This is an important question in need of a timely answer as both the EU and US regulatory bodies are currently in the process of introducing reforms to address existing shortcomings in an attempt to prevent such scandals from ever happening again.[198,202,203]

As already mentioned, at least two TAVI models from different manufacturers (one European and one American) received their CE marks and were widely implanted in patients across Europe for almost 5 years before they were granted FDA approval.[164] Critics would argue that this is yet another example where the EU medical device regulator has been 'lenient', potentially exposing large numbers of patients to unknown risks. However, this is based on the assumption that such innovative, high-risk implantable devices carry substantial risks for patients that are unknown at the time of approval and that a 'more rigorous' approach mandating pre-market RCTs will expose these.[204]

This brings us to the fundamental issue that this study aimed to explore, which was to understand how different regulatory frameworks perform not only in terms of patient (and public) safety but also innovation - not forgetting that innovation (especially in healthcare) is vital not only for enhancing patient outcomes but also to ensure wider societal prosperity in the longer term (through public wellbeing).[39,205] Based on this study's findings as well as the literature, recommendations for stakeholders and policymakers are provided on how the existing regulatory frameworks can be improved through striking a (finer) balance between promoting patient (and public) safety and smothering innovation.[206]

As evidenced by the findings, though mortality was found to be higher in Europe compared to North America, the difference was not statistically significant (p=0.470). In view of the dataset size (exceeding 2,000 publications reporting on more than 160,000 patients), this lack of statistical significance makes this finding highly reliable. This is also supported by the literature on TAVI registries across both the EU and US reporting similar mortality rates on both continents.[207-209]

To ensure the full patient safety profile was evaluated, a number of complications beyond mortality were also assessed. These included 30-day rates for disabling CVA, major vascular complications, and the development of post-TAVI paravalvular leak. Again, no statistically significant difference between the EU and US was found for major vascular complications and paravalvular leak rates, whilst for disabling CVA, the rate was higher in the US (compared to the EU), a difference that was also found to be statistically significant (p=0.002). On a global scale, both the EU and US

consistently ranked low for mortality and all other complications with the Middle East topping the rankings except for disabling CVA where Oceania exhibited the highest complication rate.

These findings reiterate the vital role that medical device regulatory frameworks play in ensuring the safety of patients and the public (with the EU and US where the most robust regulatory eco-systems feature consistently outperforming all other geographical areas). At the same time, what becomes apparent is that 'over-regulation' (in the form of compulsory randomised data at the pre-market evaluation stage) does not necessarily translate to enhanced patient safety nor does it translate to superior clinical outcomes (with the EU and US performing evenly across the clinical spectrum). Another important finding relates to the role of over-regulation as a potential barrier to innovation in view of the marked differences seen in terms of adoption and diffusion of innovation (favouring the EU over the US). The policy implications of these findings are significant, especially as over-regulation has also been shown to substantially increase associated healthcare costs in a number of studies.[193,210]

Based on this, more rigid data requirements demanding RCTs for all high-risk novel medical devices prior to market approval may need reconsideration. Though RCTs are undoubtedly among the most reliable sources of evidence when it comes to comparing healthcare interventions,[43] "absence of evidence is not evidence of abscence".[213] Moreover, RCTs - especially when it comes to comparing minimally invasive with open treatments - have been shown to suffer a number of limitations.[63] Finally, as adequately powered multi-centre RCTs commonly take years to complete,

making randomised data mandatory for initial approval, will inevitably delay patient access to innovative treatments. Thus, where an innovation has been shown to address an unmet clinical need for a common and life-threatening illness (as was the case with TAVI for inoperable aortic stenosis),[154,165] it may be justifiable to authorise its initial market approval (under strict conditions – see below) without having to wait for RCTs.

Prior to discussing novel regulatory approaches aimed at enhancing both patient safety and innovation (a fine balance that is proving increasingly difficult to attain), it is important to consider what lessons can also be learned from the aforementioned scandals involving other high-risk implantable devices. In the case of TVT implants, both the FDA and EU regulators failed to protect patients from their unknown (at the time) risks and, importantly, when these risks were eventually identified, this was not through RCTs.[214] Similarly, in the case of metal-on-metal hip replacements, the medical device (DePuy ASR™ XL Acetabular Cup System, Johnson & Johnson, New Brunswick, NJ) had received approval throughout the world including by the FDA for the US market (through the 510(k) process). Instead, concerns about these implants were first raised much later, initially in Australia in 2008 and subsequently in the UK in 2010, through their respective National Joint Registries' audits that revealed excessively high revision rates. This ultimately led to the device being recalled (including by the FDA), leading (once again) to thousands of medicolegal claims around the world and the subsequent need for mass explantations.[215]

These findings match closely those of the TAVI case in this study where, once more, neither the EU nor US proved superior in terms of ensuring the safety of their patients

and public despite the significant differences in their regulatory requirements. Those failures in protecting patients and the public – the indisputable number one priority of all regulatory systems - call for a radical rethink on how medical device regulation can be best improved on a global level.

It is thus encouraging to see that a number of initiatives are currently being introduced (or at least being considered) in an attempt to optimise both patient and public safety but also promote innovation. A prime example includes the establishment of the Medical Device Epidemiology Network (MDEpiNET), a global public-private partnership initiated by the FDA. MDEpiNET is aimed at promoting public health and informing stakeholders (including surgeons, patients, regulators, and policy makers) through the use of existing resources, thus presenting a lower-cost "national medical device surveillance system".[217]

MDEpiNET signifies a paradigm shift towards a much-needed medical device ecosystem that will contribute to overcoming many of the existing regulatory barriers through collaboration among all stakeholders. The need for inter-disciplinary collaboration that will see the surgical community, industry, regulatory bodies, and patients working closely together is now widely recognised by the FDA. Such a collaborative approach will streamline regulatory evaluation to expedite the delivery of innovative and effective medical devices to patients in the safest possible way.[189]

A notable example of how this can be achieved involves the reconfiguration of certain regulatory barriers from PMA to Post-Market Surveillance (PMS). Through the establishment of registries and prototype pre-market studies embedded within these,

data can be continuously collected and analysed, generating up-to-date, real world evidence to support regulatory decision-making.[217] This can only be possible if regulatory bodies work very closely with surgeons and the industry across all stages of surgical innovation.[39] Though promising, implementation is likely to involve a prolonged and demanding process for a number of reasons.

Some of the key obstacles are likely to relate to accessing, linking, and analysing real world data in an automated and timely fashion.[187] In addition to the practical challenges posed by the overwhelming volume and speed at which real world data are continuously being generated – which to an extent can be addressed through advanced analytical technologies developed for big data (see Chapter 2) – the legal barriers may prove insurmountable. At a time where cybersecurity scandals involving personal data breaches and misuse are becoming increasingly prevalent and, in response, data governance is getting progressively arduous, accessing highly sensitive data involving patients but also commercial products (medical devices) that can generate vast profits is likely to prove challenging.[218] Moreover, regulatory barriers will not stop even if/when data access is granted. On the contrary, they are likely to intensify when it comes to how these data can (and cannot) be used. Finally, for the data to generate evidence used to inform policy, linkage between different databases will be needed, something which will require decryption, thus raising (once more) concerns regarding data privacy and their potential for misuse.[219]

Beyond the legal and analytical challenges posed, PMS requires every single medical device to be traceable throughout its Total Product Life Cycle (TPLC) to the individual patient in whom it has been used/implanted. Promising steps are already

being taken in this direction by both the FDA and European Commission including not only on labeling requirements through the creation of Unique Device Identification (UDI), but also regarding legislation concerning data submission and the formulation and maintenance of databases (which include detailed guidance for data storage, encryption, and access to maximise online safety and ensure transparency).[220,221]

The case for change in the medical device ecosystem is compelling.[222] Whilst the aforementioned PMS interventions are being implemented, more needs doing in the interim to expedite the approval and adoption of innovative medical devices so that patients can benefit from these in a timely manner. Once more, collaboration is the key with a notable example being the Accelerated Access Pathway (AAP), a unique collaborative between the NHS and industry aimed at "improving patient access to breakthrough technologies and treatments in a cost-effective model".[222] The aim of the AAP is to identify highly transformative innovations at an early stage and streamline their route to the market so as to maximise the benefit for patients and improve the efficiency of the existing system. Similar to PMS, this will be achieved by minimising the (many) barriers to innovation uptake by generating real world evidence to supplement clinical trials.[223]

Prior to concluding, it is important to acknowledge the strengths and limitations of this study. A key strength relates to the dataset, which involved over 2,000 publications reporting on more than 160,000 patients. Aside of volume, which makes the findings robust, the dataset is multi-dimensional consisting of clinical, bibliometric, visibility, and innovation data. This broad spectrum, combined with the

diverse analytical tools employed (including 'traditional' statistical techniques, established epidemiological methods, and network analysis), make this the first study of its kind.

On the other hand, a key limitation relates to the fact that the analysis focused on a specific topic. Despite this, it would be safe to assume that the findings are generalisable to other medical devices since i) TAVI, as a Class III implantable cardiovascular device, carries the highest risk profile for patients and ii) the current findings on regulatory and safety aspects of TAVI are in alignment with those previously reported for other implantable devices as discussed above. A further limitation relates to the scarcity of observations outside the EU and the US, which made it impossible to make specific policy recommendations for the RoW countries aside of the need for developing more structured regulatory frameworks based on those already in place in the EU and the US (in view of their demonstrable consistently superior outcomes in terms of both patient safety and innovation).

7.5 Conclusion

The medical device regulatory ecosystem is in the process of undergoing major transformation from the traditional pre-market authorisation model to post-market surveillance in an attempt to strike the right balance between promoting patient safety and enhancing innovation. The findings presented here, combined with lessons learned from global scandals involving other high-risk implantable medical devices, reiterate the need for a novel collaborative approach that will see the surgical community, industry, regulatory bodies, and patients working closely together to streamline the regulatory evaluation for innovative medical devices. This exclusive

partnership will greatly enhance the ability to gather data intelligence towards the generation of real world evidence that can supplement clinical trials, support regulatory decision-making, and inform policy. This study provides important evidence for stakeholders and policymakers that justifies the need for transitioning towards a 'smart' regulatory ecosystem for all medical devices, including those considered to be high-risk.

7.6 Chapter summary

The objective of this chapter's study was to understand how distinct medical device regulatory systems differ across the world in terms of their patient safety and innovation profiles using network and regression analysis.

Pre-clinical and clinical original research studies (n=2,595) reporting on TAVI were extracted from the *Web of Science*® database and categorised based on geography. These reported on 160,892 patients from all over the world undergoing treatment for aortic valve disease in the period between 2002 (when the first live human TAVI case was performed) and December 2017 (time of data extraction).

TAVI constituted the innovative, high-risk intervention studied with the main outcome measures being geographical differences in terms of patient volume, structural virality, innovation index, mortality-per-person-year, and 30-day rates for disabling cerebrovascular accident (CVA), major vascular complications, and moderate-severe paravalvular leak.

The world's most established regulatory frameworks (EU and US) were shown to consistently perform superiorly in terms of all clinical outcomes compare to the Rest of World (RoW). When directly comparing the EU and US, no statistically significant differences were found in terms of mortality or complications rates (except for disabling CVA favouring the EU), despite significant differences in their regulatory requirements. In terms of innovation, the EU exhibited superior performance based on the number of patients treated, their timely access to TAVI, and its structural virality.

At a time when the medical device regulatory ecosystem is transitioning from the traditional pre-market authorisation model to an active post-market surveillance, this study provides important evidence for stakeholders and policymakers that justifies the need for close collaboration between the surgical community, industry, regulatory bodies, and patients to streamline the regulatory pathway and expedite the delivery of innovative and effective medical devices to patients in the safest possible way.

8. Conclusions and Future Work

This thesis examined surgical innovation from a novel perspective. Network analysis was implemented for the first time in surgical innovation to map diffusion and measure the evidence-based implementation value generated using big data spanning across the several dimensions of innovation in surgery. The network-based surgical innovation metrics developed and validated as part of this thesis are already proving to be key in evaluating surgical innovations and shaping the future of surgery.[224] This is through the provision of new intelligence crucial for informing healthcare policy, optimising medical device regulation, and directing surgical research strategy on a global scale.

Following an overview of what innovation in surgery entails, the original concept of the *surgical innovation funnel* was presented. This, combined with the novel surgical innovation metrics that were developed and validated, exemplifies the first complete and quantitative surgical innovation framework. Previous frameworks considered the first stage of surgical innovation to be "the first use of a new procedure in a patient".[50] However, surgical innovation commonly involves a number of preceding steps.[31] The *surgical innovation funnel* not only accounts for those pre-clinical steps, but also permits the measurement of surgical innovation for the first time (the unifying limitation of all pre-existing frameworks being their qualitative nature).

The ability to measure surgical innovation is crucial. The five original research studies included in this thesis illustrate how novel surgical innovation metrics were developed and validated and how they can be used in practice to inform healthcare

policy and optimise surgical decision-making. Moreover, the power of network analysis of real world surgical collaboration networks was explored to uncover the barriers and catalysts to surgical innovation, a key aspect for promoting high-quality surgical research and enhancing patient care globally. Finally, additional aspects of surgical innovation were also explored including the influence of marketing on diffusion of innovations in the global healthcare setting and how different regulatory frameworks can impact on both healthcare innovation and patient safety.

The first study illustrated how network analysis can be used to examine the structural foundations of innovation diffusion in surgery. Based on citation networks, diffusion was mapped out as a time-varying cascade of adoptions propagating over potentially many generations of adopters. The network-based metric of *structural virality* was presented enabling the creation of rankings from the most 'viral' innovations to those that only generate short-term surges in popularity and then quickly die out. A second metric, the *innovation index* was also developed aimed at capturing the intrinsic value of a given surgical innovation based on its evidence-based implementation. Both innovation metrics were validated using big data (including from the real world) that exceeded seven million hospital stays per year (NIS® database).

The second study involved the construction and analysis of a different type of surgical innovation network. Based on the fact that surgical innovation is a social process originating from complex interactions among diverse participants, the real world collaboration network for robotic surgical research was built and analysed. The importance of establishing open networks, forging a geographically diverse international profile and actively collaborating with industry in enhancing both

research impact and innovation were illustrated. This study can be regarded as proof of concept signifying how network analysis can be used in surgical research to foster innovation through strategic partnerships. Evidence was also provided on the ways in which existing collaborative efforts can be modified and future ones strategically planned to maximise research performance and enhance patient care.

The third study depicted the application of network analysis combined with simulation modelling to examine the impact of Brexit on a key area in healthcare in which the UK is set to be a global leader. With the UK setting its independent path on the global stage outside the EU following its exit from the block on 31 January 2020, this represents a much-needed study, especially as the future UK-EU framework still remains undetermined. It is the first Brexit impact study to employ network analysis and thus can be considered as a prototype for the conduction of similar studies in other important domains for the UK's prosperity, such as international trade and financial services. Both sectors are fundamental to the UK economy and ideally suited for network analysis that can be used to map and measure relationships and flows among the represented agents, e.g. nodes can denote countries or financial institutions and links the flow of goods or services, respectively.[225]

The fourth study provided an in-depth examination of disruptive innovation in healthcare. Big data originating from a variety of heterogeneous sources (including clinical, academic, and social) were extracted and linked to study TAVI, a healthcare innovation that has disrupted the modern cardiovascular medical device market.[9] Network analysis and multivariate regression were used to investigate the factors driving its unparalleled diffusion, with a particular attention to the underlying

evidence base and marketing strategies employed by manufacturers. A number of important findings illustrating the power of marketing in enhancing the diffusion of innovations in healthcare and ensuring their translation to market success were reported. The demonstrated implementation-evidence mismatch advocates the power of non-clinical factors (i.e. marketing) in the diffusion of certain (heavily promoted) innovations in healthcare and should act as a wake-up call for both clinicians and regulatory bodies when it comes to introducing novel medical devices in the market.

The fifth and final study examined different regulatory frameworks across the world in terms of not only innovation but also - crucially - patient safety. At a time when the regulation of the medical device industry has come under heavy criticism and calls for further intensifying it are increasing, this represents a principal and timely study. Using a network-based approach it elucidated the need for a collaborative approach that will see the surgical community, industry, regulatory bodies, and patients working closely together to streamline regulatory evaluation to expedite the delivery of innovative and effective medical devices to patients in the safest possible way. This will ensure greater transparency and facilitate the transition from the traditional pre-market authorisation model to an active post-market surveillance facilitated by 'smart' technology. These findings provide important evidence for stakeholders and policymakers.

Before closing, it is important to acknowledge the strengths and limitations of this thesis. The key strengths relate to the type of data pooled and the original methods employed to analyse these. This is the first time that network analysis has been used in the study of surgical innovation, and its practical value in mapping and measuring

surgical innovation was extensively demonstrated through the extraction, curation, and linkage of big data from a variety of heterogeneous sources. These extended beyond the (traditional) clinical setting to also include bibliometric databases, the real world, as well as alternative metrics (altmetrics). In doing so, healthcare innovation was examined across its several dimensions. Furthermore, a number of different networks were constructed and analysed (citation, collaboration).

In terms of limitations, these include the fact that innovation was measured only at the meso (institutional) and macro (national) levels. Moreover, only few examples of healthcare innovations were studied (robotic surgery, TAVI). Although this will naturally restrict the generalisability of the results, the methodology has now been clearly laid out for future studies that can be performed at any level – including micro (surgeon-scientist) level - and for any innovation (e.g. 3D printing, artificial intelligence, telemedicine). Of note, this limitation is inherent to innovation studies more widely, since each distinct innovation, whether technological or other, would need to be evaluated on an individual (case-by-case) basis.

Another limitation is that the ethical aspects of surgical innovation were not addressed. This however represents such an important and voluminous topic that would in itself form the subject of a separate thesis. What is crucial to reiterate is that innovation should be patient-centered and not driven by commercial interests and/or surgeon research interests. For this to be achieved, a collaborative approach similar to that described for regulation is needed, where patients, surgeons, and the industry all work together in a close partnership. This is imperative because innovation must add value, and what patients value may markedly differ to what surgeons or the industry

value.[226] In the current environment where costs and metrics are becoming increasingly important, the work conducted in this thesis can serve as a building block for future studies on surgical ethics and, in particular, on the challenges these can pose on surgical innovation and how they can be overcome.

To conclude, this thesis offers unique new opportunities for understanding, modelling, and measuring surgical innovation and ultimately for enhancing this through the assessment, comparison and ranking by generative value. Importantly, the methodology described in this thesis can be used to evaluate any innovation in surgery (and beyond) and at any level required. In doing so, it lays the foundations for future studies across the entire spectrum of surgical research that will foster innovation and enhance patient care. The novel surgical innovation metrics developed may prove of great value especially for guiding policy makers, funding bodies, surgeons, and healthcare providers in the current climate of competing national priorities for investment. Indeed, they already form part of the *Future of Surgery* report recently published by The Royal College of Surgeons of England.

References

1. Riskin DJ, Longaker MT, Gertner M, Krummel TM. Innovation in surgery: a historical perspective. *Ann Surg.* 2006;244(5):686-693.
2. Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet.* 2009;374(9695):1097-1104.
3. Hughes-Hallett A, Mayer EK, Marcus HJ, et al. Quantifying innovation in surgery. *Ann Surg.* 2014;260(2):205-211.
4. O'Sullivan GC. Advancing surgical research in a sea of complexity. *Ann Surg.* 2010;252(5):711-714.
5. McCulloch P, Feinberg J, Philippou Y, et al. Progress in clinical research in surgery and IDEAL. *Lancet.* 2018;392(10141):88-94.
6. Wright JD. Robotic-Assisted Surgery: Balancing Evidence and Implementation. *JAMA.* 2017;318(16):1545-1547.
7. Ricciotti HA, Armstrong W, Yaari G, Campion S, Pollard M, Golen TH. Lessons from Google and Apple: creating an open workplace in an academic medical department to foster innovation and collaboration. *Acad Med.* 2014;89(9):1235-1238.
8. Innovation Point from inspiration to impact. The Complete Guide to Innovation Metrics - How to Measure Innovation for Business Growth. 2017; http://www.innovation-point.com/innovationmetrics.htm.
9. Wong KA, Hodgson L, Garas G, et al. How can cardiothoracic and vascular medical devices stay in the market? *Interact Cardiovasc Thorac Surg.* 2016;23(6):940-948.

10. Economou JS. Engines of Discovery and Innovation. *Ann Surg.* 2016;264(3):405-412.

11. Rizzo A, Porfiri M. Innovation diffusion on time-varying activity driven networks. *The European Physical Journal B.* 2016;89(20):1-8.

12. OECD. *Oslo Manual: Guidelines for Collecting and Intrerpreting Innovation Data* 3rd Edition ed. Paris, France: OECD Publishing; 2005.

13. Hutchison K, Rogers W, Eyers A, Lotz M. Getting Clearer About Surgical Innovation: A New Definition and a New Tool to Support Responsible Practice. *Ann Surg.* 2015;262(6):949-954.

14. Liu TC, Bozic KJ, Teisberg EO. Value-based Healthcare: Person-centered Measurement: Focusing on the Three C's. *Clin Orthop Relat Res.* 2017;475(2):315-317.

15. Brook RH. The end of the quality improvement movement: long live improving value. *JAMA.* 2010;304(16):1831-1832.

16. Scott TW, Tiessen P. Performance measurement and managerial teams. *Accounting, Organizations and Society.* 1999;24:263-285.

17. Doran GT. There 's a S.M.A.R.T. way to write management's goals and objectives. *Management Review AMA FORUM.* 1981;70(11):35-36.

18. Apuzzo ML, Elder JB, Faccio R, Liu CY. The alchemy of ideas. *Neurosurgery.* 2008;63(6):1035-1044; discussion 1044.

19. Ashrafian H, Rao C, Darzi A, Athanasiou T. Benchmarking in surgical research. *Lancet.* 2009;374(9695):1045-1047.

20. Malik HT, Marti J, Darzi A, Mossialos E. Savings from reducing low-value general surgical interventions. *Br J Surg.* 2018;105(1):13-25.

21. Moris D, Felekouras E, Linos D. Global surgery initiative in Greece: more than an essential initiative. *Lancet.* 2016;388(10048):957.

22. Fryatt RJ, Bhuwanee K. Financing health systems to achieve the health Sustainable Development Goals. *Lancet Glob Health.* 2017;5(9):e841-e842.

23. Frankel S. Health needs, health-care requirements, and the myth of infinite demand. *Lancet.* 1991;337(8757):1588-1590.

24. Mueller HF, Uphoff WH, Zoellner H. Medical services--demand versus need: a comment. *Am J Public Health.* 1974;64(1):54-57.

25. Wright JD. Robotic-Assisted Surgery Balancing Evidence and Implementation. *JAMA.* 2017;318(16):1545-1547.

26. Ashrafian H, Clancy O, Grover V, Darzi A. The evolution of robotic surgery: surgical and anaesthetic aspects. *Br J Anaesth.* 2017;119(suppl_1):i72-i84.

27. Tan A, Ashrafian H, Scott AJ, et al. Robotic surgery: disruptive innovation or unfulfilled promise? A systematic review and meta-analysis of the first 30 years. *Surg Endosc.* 2016;30(10):4330-4352.

28. Wright JD, Tergas AI, Hou JY, et al. Effect of Regional Hospital Competition and Hospital Financial Status on the Use of Robotic-Assisted Surgery. *JAMA Surg.* 2016;151(7):612-620.

29. Barbash GI, Glied SA. New technology and health care costs--the case of robot-assisted surgery. *N Engl J Med.* 2010;363(8):701-704.

30. Ibrahim SA. Decision Aids and Elective Joint Replacement - How Knowledge Affects Utilization. *N Engl J Med.* 2017;376(26):2509-2511.

31. Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Beyond IDEAL: the importance of surgical innovation metrics. *Lancet.* 2019;393(10169):315.

32. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.

33. Garas G, Okabayashi K, Ashrafian H, et al. Which hemostatic device in thyroid surgery? A network meta-analysis of surgical technologies. *Thyroid.* 2013;23(9):1138-1150.

34. Staiger DO, Dimick JB, Baser O, Fan Z, Birkmeyer JD. Empirically derived composite measures of surgical performance. *Med Care.* 2009;47(2):226-233.

35. Hansen MT, Birkinshaw J. The innovation value chain. *Harv Bus Rev.* 2007;85(6):121-130, 142.

36. Dawkins R. *The Selfish Gene.* Oxford: Oxford University Press; 1999.

37. Dunphy SM, Herbig PR, Howes ME. The Innovation Funnel. *Technological Forecasting and Social Change.* 1996;53(3):279-292.

38. Sagar SP, Law PW, Shaul RZ, Heon E, Langer JC, Wright JG. Hey, I just did a new operation!: Introducing innovative procedures and devices within an academic health center. *Ann Surg.* 2015;261(1):30-31.

39. Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery. *PLoS One.* 2017;12(8):e0183332.

40. Bush RW. Reducing waste in US health care systems. *JAMA.* 2007;297(8):871-874.

41. Chan JK, Shalhoub J, Gardiner MD, Suleman-Verjee L, Nanchahal J. Strategies to secure surgical research funding: fellowships and grants. *JRSM Open.* 2014;5(1):2042533313505512.

42. Dzau VJ, McClellan MB, McGinnis JM, et al. Vital Directions for Health and Health Care: Priorities From a National Academy of Medicine Initiative. *JAMA*. 2017;317(14):1461-1470.

43. Garas G, Ibrahim A, Ashrafian H, et al. Evidence-based surgery: barriers, solutions, and the role of evidence synthesis. *World J Surg.* 2012;36(8):1723-1731.

44. Kirsner S. What Big Companies Get Wrong About Innovation Metrics. *Harvard Business Review*. 2015:1-5.

45. Barlow J. Why do we need to understand healthcare innovation? *Managing Innovation in Healthcare*. Singapore: World Scientific Publishing Europe Ltd.; 2017.

46. McAfee A, Brynjolfsson E. Big data: the management revolution. *Harv Bus Rev*. 2012;90(10):60-66, 68, 128.

47. Schneeweiss S. Learning from big health care data. *N Engl J Med.* 2014;370(23):2161-2163.

48. Weiser TG, Makary MA, Haynes AB, et al. Standardised metrics for global surgical surveillance. *Lancet.* 2009;374(9695):1113-1117.

49. Barkun JS, Aronson JK, Feldman LS, et al. Evaluation and stages of surgical innovations. *Lancet.* 2009;374(9695):1089-1096.

50. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet.* 2009;374(9695):1105-1112.

51. Lamont J. Innovation: What are the real metrics? *KMWorld Content, Document and Knowledge Management.* Vol 242015:1-2.

52. Vecht JA, von Segesser LK, Ashrafian H, et al. Translation to success of surgical innovation. *Eur J Cardiothorac Surg.* 2010;37(3):613-625.

53. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg.* 2015;102(2):e93-e101.

54. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med.* 2017;376(26):2507-2509.

55. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *Journal of Big Data.* 2015;2(1):1-21.

56. Bordes A, Glorot X, Weston J, Bengio Y. Joint learning of words and meaning representations for open-text semantic parsing. *International Conference on Artificial Intelligence and Statistics JMLRorg.* 2012:127-135.

57. Davenport TH, Ronanki R. Artificial Intelligence for the Real World. *Harvard Business Review.* 2018;96(1-2):108-116.

58. Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big Data and Machine Learning in Plastic Surgery: A New Frontier in Surgical Innovation. *Plast Reconstr Surg.* 2016;137(5):890e-897e.

59. IBM Watson for Oncology©. 2020; https://www.ibm.com/watson/health/oncology-and-genomics/oncology/.

60. Curioni-Fontecedro A. A new era of oncology through artificial intelligence. *ESMO Open.* 2017;2(2):e000198.

61. Hagen T. IBM's Watson Achieves High Concordance in Tumor Board Test. 2016; http://www.onclive.com/conferencecoverage/sabcs-2016/ibms-watson-achieves-high-concordance-intumor-board-test.

62. van Aalst HF. Networking in Society, Organisations and Education. In: OECD, ed. *Networks of Innovation Towards New Models for Managing Schools and Systems*. Paris, France: OECD Publishing; 2003.

63. Garas G, Markar SR, Malietzis G, et al. Induced Bias Due to Crossover Within Randomized Controlled Trials in Surgical Oncology: A Meta-regression Analysis of Minimally Invasive versus Open Surgery for the Treatment of Gastrointestinal Cancer. *Ann Surg Oncol.* 2018;25(1):221-230.

64. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med.* 2016;375(23):2293-2297.

65. Scarlat MM, Mavrogenis AF, Pecina M, Niculescu M. Impact and alternative metrics for medical publishing: our experience with International Orthopaedics. *Int Orthop.* 2015;39(8):1459-1464.

66. Iron Mountain® Secure Storage. 2020; http://www.ironmountain.com/industries/healthcare-services/secure-storage.

67. IBM© Blockchain in Healthcare. 2020; https://www.ibm.com/blogs/blockchain/category/blockchain-in-healthcare/.

68. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK, Ideal Group. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ.* 2013;346:f3012.

69. Wiener H. Structural determination of paraffin boiling points. *J Am Chem Soc.* 1947;69(1):17-20.

70. Pastor-Satorras R, Vespignani A. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters.* 2001;86(14):3200-3203.

71. Elsevier®. SciVerse Scopus® Amsterdam, The Netherlands. 2020.

72. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* 2003:137-146.

73. Ciotti V, Bonaventura M, Nicosia V, Panzarasa P, Latora V. Homophily and missing links in citation networks. *EPJ Data Science.* 2016;5(7):1-14.

74. Newman MEJ. *Networks An Introduction.* New York: Oxford University Press; 2010.

75. Goel S, Anderson A, Hofman J, Watts DJ. The Structural Virality of Online Diffusion. *Management Science.* 2016;61(1):180-196.

76. Dow PA, Adamic LA, Friggeri A. The Anatomy of Large Facebook Cascades. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* 2013:145-154.

77. Dunphy SM, Herbig PR, Howes ME. The Innovation Funnel *Technological Forecasting and Social Change.* 1996;53:279-292.

78. Dawkins R. *The Selfish Gene.* Oxford: Oxford University Press; 1976.

79. National (Nationwide) Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). USA 2012.

80. Arora A, Garas G, Sharma S, et al. Comparing transaxillary robotic thyroidectomy with conventional surgery in a UK population: A case control study. *Int J Surg.* 2016;27:110-117.

81. Trehan A, Dunn TJ. The robotic surgery monopoly is a poor deal. *BMJ.* 2013;347:f7470.

82. Department of Health. *High Quality Care For All NHS Next Stage Review Final Report.* London: United Kingdom;2009.

83. Arena M, Cross R, Sims J, Uhl-Bien M. How to Catalyze Innovation in Your Organization. *MIT Sloan Management Review.* 2017;58(4):39-47.

84. Davis JP, Eisenhardt KM. Rotating Leadership and Collaborative Innovation: Recombination Processes in Symbiotic Relationships. *Administrative Science Quarterly.* 2011;56(2):159-201.

85. Chao TE, Riesel JN, Anderson GA, et al. Building a global surgery initiative through evaluation, collaboration, and training: the Massachusetts General Hospital experience. *J Surg Educ.* 2015;72(4):e21-28.

86. Davies JI, Meara JG. Global surgery--going beyond the Lancet Commission. *Lancet.* 2015;386(9993):507-509.

87. Soreide K, Alderson D, Bergenfelz A, et al. Strategies to improve clinical research in surgery through international collaboration. *Lancet.* 2013;382(9898):1140-1151.

88. Latora V, Nicosia V, Panzarasa P. Social Cohesion, Structural Holes, and a Tale of Two Measures. *J Stat Phys.* 2013;151:745-764.

89. Fleming L, Mingo S, Chen D. Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly.* 2007;52:443-475.

90. Social Capital Research. Definition of Social Capital. 2020; https://www.socialcapitalresearch.com/literature/definition.html.

91. Ye J, Kankanhalli A. Exploring innovation through open networks: A review and initial research questions. *IIMB Management Review.* 2013;25:69-82.

92. Ortega JL, Aguillo IF. Institutional and coutry collaboration in an online service of scientific profiles: Google Scholar Citations. *Journal of Informetrics.* 2013;7:394-403.

93. Smith M. The trend toward multiple authorship in psychology. *American Psychologist.* 1958;13:596-599.

94. Opsahl T, Panzarasa P. Clustering in weighed networks. *Social Networks.* 2009;31:155-163.

95. Su Y, Long C, Yu Q, Zhang J, Wu D, Duan Z. Global scientific collaboration in COPD research. *Int J Chron Obstruct Pulmon Dis.* 2017;12:215-225.

96. Opsahl T, Agneessens F, Skvoretz J. Node Centrality in Weighed Networks: Generalalizing Degree and Shortest Paths. *Social Networks.* 2010;32:245-251.

97. Savage N. Industry links boost research output. *Nature.* 2017;552(7683):S11-S13.

98. Web of Science®. *InCites Indicators Handbook.* Philadelphia, PA: Clarivate Analytics;2020.

99. Bjerregaard T. Universities-industry collaboration strategies: a micro-level perspective. *European Journal of Innovation Management.* 2009;12(2):161-176.

100. The Royal College of Surgeons of England. Funding Surgical Research. 2020; https://www.rcseng.ac.uk/about-the-rcs/support-our-work/funding-surgical-research/.

101. Mann M, Tendulkar A, Birger N, Howard C, Ratcliffe MB. National institutes of health funding for surgical research. *Ann Surg.* 2008;247(2):217-221.

102. Imperial College London Faculty of Medicine. Corporate Pantnerships. London, United Kingdom: 2020; http://www.imperial.ac.uk/medicine/partnership/corporate-partnerships/.

103. Johns Hopkins School of Medicine. Industrial Liaison Office. Baltimore, MD: 2020; https://www.bme.jhu.edu/about/careers/industrial-liaison/.

104. Behrens TR, Gray DO. Unintended consequences of cooperative research: impact of industry sponsorship on climate for academic freedom and other graduate student outcome. *Research Policy.* 2001;30(2):179-199.

105. Welsh R, Glenna L, Lacy W, Biscotti D. Close enough but not too far: Assessing the effects of university–industry research relationships and the rise of academic capitalism. *Research Policy.* 2008;37(10):1854-1864.

106. Baird P. Getting it right: industry sponsorship and medical research. *CMAJ.* 2003;168(10):1267-1269.

107. Camilleri M, Tremaine WJ. Governance of clinical research. *Am J Gastroenterol.* 2012;107(3):336-338.

108. Newman MEJ. *Networks An Introduction.* Oxford, United Kingdom: Oxford University Press; 2010.

109. Chen K, Yao Q, Sun J, He ZF, Yao L, Liu ZY. International publication trends and collaboration performance of China in healthcare science and services research. *Isr J Health Policy Res.* 2016;5:1.

110. Gonzalez-Alcaide G, Park J, Huamani C, Belinchon I, Ramos JM. Evolution of Cooperation Patterns in Psoriasis Research: Co-Authorship Network Analysis of Papers in Medline (1942-2013). *PLoS One.* 2015;10(12):e0144837.

111. Campbell B, Patrick H, Register Group. International collaboration in the use of registries for new devices and procedures. *Br J Surg.* 2012;99(6):744-745.

112. Jamjoom AA, Phan PN, Hutchinson PJ, Kolias AG. Surgical trainee research collaboratives in the UK: an observational study of research activity and publication productivity. *BMJ Open.* 2016;6(2):e010374.

113. Massenburg BB, Raykar NP, Pawaskar A, Gnanaraj J, Roy N. Collaboration and innovation in rural surgery. *Int Health.* 2016;8(6):367-368.

114. Nepogodiev D, Chapman SJ, Kolias AG, et al. The effect of trainee research collaboratives in the UK. *Lancet Gastroenterol Hepatol.* 2017;2(4):247-248.

115. Tijssen R, Yegros A. Brexit: UK universities and European industry. *Nature.* 2017;544(7648):35.

116. Kazzazi F, Pollard C, Tern P, Ayuso-Garcia A, Gillespie J, Thomsen I. Evaluating the impact of Brexit on the pharmaceutical industry. *J Pharm Policy Pract.* 2017;10:32.

117. House of Commons Science and Technology Committee. *Leaving the EU: implications and opportunities for science and research.* London, United Kingdom; November 2016.

118. HM Government. *EU Exit Long-term economic analysis* London, United Kingdom; November 2018.

119. HM Government. *EU Exit Long-term economic analysis Technical Reference Paper.* London, United Kingdom; November 2018.

120. HM Government. HM Treasury analysis: the immediate economic impact of leaving the EU. London, United Kingdom; May 2016.

121. HM Government. HM Treasury analysis: the long-term economic impact of EU membership and the alternatives. London, United Kingdom; April 2016.

122. Ge J, Polhill JG, Matthews KB, Miller DG, Spencer M. Not one Brexit: How local context and social processes influence policy analysis. *PLoS One.* 2018;13(12):e0208451.

123. Reijntjes S, Albayaty M, Bush J, et al. The Association for Human Pharmacology in the Pharmaceutical Industry London Meeting 2018: Brexit

and Other Challenges in Early Phase Drug Development. *Front Pharmacol.* 2018;9:1301.

124. Halliday A, Hacker J. Brexit and European science. *Science.* 2017;358(6361):279.

125. Breckenridge A, Feldschreiber P. Impact of Brexit on UK and EU Drug Regulation and Patient Access. *Clin Pharmacol Ther.* 2018.

126. Siva N. UK researchers digest the fallout from Brexit. *Lancet.* 2016;388(10040):115-116.

127. Shapey IM, Summers AM, Simkin IJ, Augustine T, van Dellen D. When politics meets science: What impact might Brexit have on organ donation and transplantation in the United Kingdom? *Clin Transplant.* 2018;32(8):e13299.

128. Fahy N, Hervey T, Greer S, et al. How will Brexit affect health and health services in the UK? Evaluating three possible scenarios. *Lancet.* 2017;390(10107):2110-2118.

129. Simpkin VL, Mossialos E. Brexit and the NHS: Challenges, uncertainties and opportunities. *Health Policy.* 2017;121(5):477-480.

130. Hu Z, Lin G, Sun T, Wang X. An EU without the UK: mapping the UK's changing roles in the EU scientific research. *Scientometrics.* 2018;115(3):1185-1198.

131. HM Government. *Industrial Strategy Eight Great Technologies*. London, United Kingdom; January 2017.

132. Garas G, Cingolani I, Patel V, Panzarasa P, Darzi A, Athanasiou, T. Evaluating the implications of Brexit for research collaboration and policy: A network analysis and simulation study. *BMJ Open.* 2019;9(9):e025025.

133. Garas G, Cingolani I, Patel V, et al. Surgical Innovation in the Era of Global Surgery: A Network Analysis. *Ann Surg.* 2018. [Epub ahead of print]

134. Cressey D. Academics across Europe join 'Brexit' debate. *Nature.* 2016;530(7588):15.

135. HM Government. *Collaboration on science and innovation A future partnership paper*. Department for Exiting the European Union. London, United Kingdom; 2017.

136. Wilsdon J. UK science, post-Brexit. *Science.* 2017;355(6331):1243.

137. Stokstad E. EUROPE. Uncertainty reigns in Brexit Britain. *Science.* 2016;353(6298):437.

138. The Lancet. Wellcome seeks Brexit carve-out for UK research. *Lancet.* 2018;391(10123):814.

139. Hawkes N. UK scientists would lose money and influence under "Brexit," says Lords report. *BMJ.* 2016;353:i2274.

140. HM Government. *The future relationship between the United Kingdom and the European Union.* London, United Kingdom; July 2018.

141. HM Government. *Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and the European Atomic Energy Community.* London, United Kingdom; October 2019.

142. HM Government. *Political Declaration setting out the framework for the future relationship between the European Union and the United Kingdom.* London, United Kingdom; October 2019.

143. Zhou P, Leydesdorff L. The emergence of China as a leading nation in science. *Research Policy.* 2006;35(1):83-104.

144. The Royal Society. *Knowledge, networks and nations: Global scientific collaboration in the 21st century.* London, United Kingdom; March 2011.

145. House of Commons Foreign Affairs Committee. *Global Britain.* London, United Kingdom; March 2018.

146. The Lancet Infectious Diseases. Looking beyond Brexit: union without Union. *Lancet Infect Dis.* 2016;16(11):1203.

147. The Royal Society. *UK research and the European Union The role of the EU in international research collaboration and researcher mobility.* London, United Kingdom; 2016.

148. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: toward an objective measure of scientific impact. *Proc Natl Acad Sci U S A.* 2008;105(45):17268-17272.

149. Ramakrishna H, Patel PA, Gutsche JT, et al. Transcatheter Aortic Valve Replacement: Clinical Update on Access Approaches in the Contemporary Era. *J Cardiothorac Vasc Anesth.* 2016;30(5):1425-1429.

150. Malaisrie SC, Iddriss A, Flaherty JD, Churyla A. Transcatheter Aortic Valve Implantation. *Curr Atheroscler Rep.* 2016;18(5):27.

151. Al-Adhami A, Al-Attar N. Recent advances in aortic valve replacement for aortic stenosis. *F1000Res.* 2016;5.

152. Cribier A, Eltchaninoff H, Bash A, et al. Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: first human case description. *Circulation.* 2002;106(24):3006-3008.

153. Hamm CW, Arsalan M, Mack MJ. The future of transcatheter aortic valve implantation. *Eur Heart J.* 2016;37(10):803-810.

154. Leon MB. Transcatheter aortic valve replacement: a breakthrough medical therapy! The 20-year odyssey, and now, a 10-year anniversary. *Arch Cardiovasc Dis.* 2012;105(3):129-131.

155. Soleimani F, Zenios S. Disrupting incrementalism in health care innovation. *Ann Surg.* 2011;254(2):203-208.

156. Newman MEJ. Citation Networks. *Networks An Introduction*. New York: Oxford University Press; 2010:67-72.

157. CTSPedia. Incidence Rate Using Person-Time. 2009; https://www.ctspedia.org/do/view/CTSpedia/StudyIncidencePersonTime.

158. Guevara JP, Berlin JA, Wolf FM. Meta-analytic methods for pooling rates when follow-up duration varies: a case study. *BMC Med Res Methodol.* 2004;4:17.

159. D'Agostino RS, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2017 Update on Outcomes and Quality. *Ann Thorac Surg.* 2017;103(1):18-24.

160. Kappetein AP, Head SJ, Genereux P, et al. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. *J Thorac Cardiovasc Surg.* 2013;145(1):6-23.

161. About Altmetric and the Altmetric Attention Score. 2020; https://help.altmetric.com/support/solutions/articles/6000059309-about-altmetric-and-the-altmetric-attention-score.

162. Warren HR, Raison N, Dasgupta P. The Rise of Altmetrics. *JAMA.* 2017;317(2):131-132.

163. Davies H, Lessof MH, Roberts CI, Ross DN. Catheter-mounted valve for temporary relief of aortic insufficiency. *Lancet.* 1965;285(7379):250.

164. Van Brabandt H, Neyt M, Hulstaert F. Transcatheter aortic valve implantation (TAVI): risky and costly. *BMJ.* 2012;345:e4710.

165. Berlin DB, Davidson MJ, Schoen FJ. The power of disruptive technological innovation: Transcatheter aortic valve implantation. *J Biomed Mater Res B Appl Biomater.* 2015;103(8):1709-1715.

166. Christensen CM. What is disruptive innovation? *Harvard Business Review.* 2015;93(10):44-53.

167. Christensen CM. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* Boston, MA: Harvard Business School Press; 1997.

168. Christensen CM, Grossman JH, Hwang J. *The Innovator's Pre- scription: A Disruptive Solution for Health Care.* New York, NY: McGraw-Hill; 2009.

169. Manolis AS. Transcatheter aortic valve implantation economics: a grisly reality. *Ann Cardiothorac Surg.* 2017;6(5):516-523.

170. Leon MB, Smith CR, Mack M, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med.* 2010;363(17):1597-1607.

171. Mack MJ, Leon MB, Smith CR, et al. 5-year outcomes of transcatheter aortic valve replacement or surgical aortic valve replacement for high surgical risk patients with aortic stenosis (PARTNER 1): a randomised controlled trial. *Lancet.* 2015;385(9986):2477-2484.

172. Smith CR, Leon MB, Mack MJ, et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med.* 2011;364(23):2187-2198.

173. Serruys PW, Unger F, Sousa JE, et al. Comparison of coronary-artery bypass surgery and stenting for the treatment of multivessel disease. *N Engl J Med.* 2001;344(15):1117-1124.

174. Berwick DM. Disseminating innovations in health care. *JAMA.* 2003;289(15):1969-1975.

175. Fish RD. Percutaneous heart valve replacement: enthusiasm tempered. *Circulation.* 2004;110(14):1876-1878.

176. Abdelghani M, Serruys PW. Transcatheter Aortic Valve Implantation in Lower-Risk Patients With Aortic Stenosis: Is It Justified to Be the Preferred Treatment? *Circ Cardiovasc Interv.* 2016;9(4):e002944.

177. Mylotte D, Osnabrugge RLJ, Windecker S, et al. Transcatheter aortic valve replacement in Europe: adoption trends and factors influencing device utilization. *J Am Coll Cardiol.* 2013;62(3):210-219.

178. Lathyris DN, Patsopoulos NA, Salanti G, Ioannidis JP. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur J Clin Invest.* 2010;40(2):172-182.

179. O'Riordan M. STACCATO; transapical TAVI in surgery-eligible patients stopped due to adverse events. *Heartwire.* 2011.

180. Arsalan M, Walther T. Durability of prostheses for transcatheter aortic valve implantation. *Nat Rev Cardiol.* 2016;13(6):360-367.

181. Reardon MJ, Van Mieghem NM, Popma JJ, et al. Surgical or Transcatheter Aortic-Valve Replacement in Intermediate-Risk Patients. *N Engl J Med.* 2017;376(14):1321-1331.

182. Teece DJ. Business Models, Business Strategy and Innovation. *Long Range Planning* 2010;43:172-194.

183. Turley LW, Kelley SW. A Comparison of Advertising Content: Business to Business versus Consumer Services. *Journal of Advertising.* 2013;26(4):39-48.

184. Lubloy A. Factors affecting the uptake of new medicines: a systematic literature review. *BMC Health Serv Res.* 2014;14:469.

185. Lauck SB, Baumbusch J, Achtem L, et al. Factors influencing the decision of older adults to be assessed for transcatheter aortic valve implantation: An exploratory study. *Eur J Cardiovasc Nurs.* 2016;15(7):486-494.

186. Kapadia SR, Leon MB, Makkar RR, et al. 5-year outcomes of transcatheter aortic valve replacement compared with standard treatment for patients with inoperable aortic stenosis (PARTNER 1): a randomised controlled trial. *Lancet.* 2015;385(9986):2485-2491.

187. Resnic FS, Matheny ME. Medical Devices in the Real World. *N Engl J Med.* 2018;378(7):595-597.

188. Egbosimba D. Medical devices industry: the problem of clinical evaluation reports. *Ann Biol Clin (Paris).* 2019;77(5):514-516.

189. Guerra-Bretana RM, Florez-Rendon AL. Impact of regulations on innovation in the field of medical devices. *Res Biomed Eng.* 2018.

190. Allied Analytics LLP. *Transcatheter Aortic Valve Implantation Market by Procedure - Global Opportunity Analysis and Industry Forecast, 2018-2025.* 2018. 4603645.

191. Overtchouk P, Modine T. A comparison of alternative access routes for transcatheter aortic valve implantation. *Expert Rev Cardiovasc Ther.* 2018;16(10):749-756.

192. Sorenson C, Drummond M. Improving medical device regulation: the United States and Europe in perspective. *Milbank Q.* 2014;92(1):114-150.

193. Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. *N Engl J Med.* 2012;366(9):848-855.

194. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep.* 2014;129 Suppl 2:19-31.

195. European Commission. *Assessing the Impacts of EU Regulatory Barriers on Innovation Final Report.* Luxembourg: Publications Office of the European Union: European Union; 2017.

196. Herzlinger RE. Why innovation in health care is so hard. *Harv Bus Rev.* 2006;84(5):58-66, 156.

197. Patel VM, Ashrafian H, Ahmed K, et al. How has healthcare research performance been assessed?: a systematic review. *J R Soc Med.* 2011;104(6):251-261.

198. Kesselheim AS, Rajan PV. Regulating incremental innovation in medical devices. *BMJ.* 2014;349:g5303.

199. European Commission. *Guidance document - Classification of Medical Devices - MEDDEV 2.4/1 rev.9.* Brussels, European Union; 2010.

200. Freemantle N. Commentary: Evaluating and regulating device therapy. *BMJ.* 2011;342:d2839.

201. Lee D, Bacsu C, Zimmern PE. Meshology: a fast-growing field involving mesh and/or tape removal procedures and their outcomes. *Expert Rev Med Devices.* 2015;12(2):201-216.

202. Zuckerman D, Booker N, Nagda S. Public health implications of differences in U.S. and European Union regulatory policies for breast implants. *Reprod Health Matters.* 2012;20(40):102-111.

203. Heneghan C. The saga of Poly Implant Prosthese breast implants. *BMJ.* 2012;344:e306.

204. Hwang TJ, Sokolov E, Franklin JM, Kesselheim AS. Comparison of rates of safety issues and reporting of trial outcomes for medical devices approved in the European Union and United States: cohort study. *BMJ.* 2016;353:i3323.

205. HM Government. Building our Industrial Strategy *Green Paper*. London, United Kingdom; January 2017.

206. Iacobucci G. Babylon app will be properly regulated to ensure safety, government insists. *BMJ.* 2018;362:k3215.

207. Eggebrecht H, Bestehorn M, Haude M, et al. Outcomes of transfemoral transcatheter aortic valve implantation at hospitals with and without on-site cardiac surgery department: insights from the prospective German aortic valve replacement quality assurance registry (AQUA) in 17 919 patients. *Eur Heart J.* 2016;37(28):2240-2248.

208. Gilard M, Eltchaninoff H, Iung B, et al. Registry of transcatheter aortic-valve implantation in high-risk patients. *N Engl J Med.* 2012;366(18):1705-1715.

209. Suradi HS, Hijazi ZM. TAVR update: Contemporary data from the UK TAVI and US TVT registries. *Glob Cardiol Sci Pract.* 2015;2015(2):21.

210. PricewaterhouseCoopers. *Medical Technology Innovation Scorecard The race for global leadership.* 2011.

211. Thielmann M, Eggebrecht H, Wendt D, et al. New techniques for the treatment of valvular aortic stenosis--transcatheter aortic valve implantation

with the SAPIEN heart valve. *Minim Invasive Ther Allied Technol.* 2009;18(3):131-141.

212. Zahn R, Gerckens U, Grube E, et al. Transcatheter aortic valve implantation: first results from a multi-centre real-world registry. *Eur Heart J.* 2011;32(2):198-204.

213. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ.* 1995;311(7003):485.

214. Heneghan CJ, Goldacre B, Onakpoya I, et al. Trials of transvaginal mesh devices for pelvic organ prolapse: a systematic database review of the US FDA approval process. *BMJ Open.* 2017;7(12):e017125.

215. Ardaugh BM, Graves SE, Redberg RF. The 510(k) ancestry of a metal-on-metal hip implant. *N Engl J Med.* 2013;368(2):97-100.

216. De Maria C, Di Pietro L, Diaz Lantada A, et al. Safe innovation: On medical device legislation in Europe and Africa *Health Policy and Technology.* 2018;7:156-165.

217. Krucoff MW, Sedrakyan A, Normand SL. Bridging Unmet Medical Device Ecosystem Needs With Strategically Coordinated Registries Networks. *JAMA.* 2015;314(16):1691-1692.

218. Martin G, Martin P, Hankin C, Darzi A, Kinross J. Cybersecurity and healthcare: how safe are we? *BMJ.* 2017;358:j3179.

219. Lawler M, Morris AD, Sullivan R, et al. A roadmap for restoring trust in Big Data. *Lancet Oncol.* 2018;19(8):1014-1015.

220. Food and Drug Administration. *Unique Device Identification (UDI) for Medical Devices.* Silver Spring, MD; 2012.

221. European Commission. *The EU Regulation on Medical Devices 2017/745*. Brussels, European Union; 2017.

222. Department of Health. *Making a reality of the Accelarates Access Review*. London, United Kingdom; 2017.

223. National Institute for Health and Clinical Excellence (NICE). Accelarated Access Collaborative. London, United Kingdom. 2018; https://www.nice.org.uk/aac - pathway.

224. The Royal College of Surgeon of England. *Future of Surgery*. London, United Kingdom; December 2018.

225. Espinosa-Vega MA, Sole J. Introduction to the Network Analysis Approach to Stress Testing. In: Ong LL, ed. *A Guide to IMF Stress Testing Methods and Models*. Washington, DC: International Monetary Fund; 2014:205-208.

226. Angelos P. Surgical ethics and the challenge of surgical innovation. *Am J Surg*. 2014;208(6):881-885.

Appendices

Appendix 1

S1. Data

S1.1 The construction of the citation networks

To construct the network of citations among articles concerning robotic surgery, we searched in Elsevier's Scopus database for all those articles related to robotic surgery. In total, we retrieved 14,775 articles that were assigned to the major subject area "Medicine", including "robotic" in the title or in the abstract or in the keywords, and were published from 1974 until the 19th of December 2015.

In Scopus, articles can be assigned to one or more major subject areas according to the classification of the Journal where they have been published. Journals are typically assigned to one or more subject areas of different granularity. In total, there are 27 major subject areas, and 313 specific subject categories according to the Scopus® Subject Areas and Subject Categories. The scientific journals that belong to the major subject area "Medicine" represent 34% of all journals included in Scopus (last update in October 2016). Notice that all journals included in the MEDLINE database are also included in Scopus. Keywords, when available, included those selected directly by the authors of articles and those assigned manually by a Scopus team of professional indexers, according to the MeSH vocabulary and the EMTREE medical terms. Typically, Scopus adds index terms to about 80% of articles included in the Scopus database.

Since the Scopus database only included part of the articles published in 2015, we decided to filter out from our data set the 1,504 articles published in that year. Moreover, our original data set included 31 articles expected to be published in 2016 that we decided not to include. Once it was cleaned and filtered as described above, the data set included a total of 13,240 articles.

Given these 13,240 articles of interest, we collected the unique electronic identificators (EIDs) used by Scopus database to uniquely identify them. We then used the EIDs to retrieve the information needed to build an acyclic directed network of citations. For each of the 13,240 articles that received at least one citation, we retrieved the EIDs of the articles that made a citation to the focal article during the period starting from the year of the publication of the article up to the 31st of December 2014. In particular, we considered only citations originating from articles pertaining to robotics in Medicine and published by the end of 2014 (i.e., belonging to our set of 13,240 articles). In this way, we restricted our focus to the citation network among articles only within robotics in Medicine. Indeed, for example, if any given article among the set of 13,240 ones had received citations only from articles not belonging to this set of 13,240 articles or from articles pertaining to robotics in Medicine but published after the 31st of December 2014, then this article would be associated with a zero count of citations. On the one hand, among the 13,240 articles here considered, 5,961 received at least one citation from an article belonging to the same set (i.e., 5,961 articles have an in-degree equal to or higher than one). On the other, 8,158 articles among the 13,240 ones have made at least one citation to another article belonging to the same set (i.e., 8,158 article have an out-degree equal to or

higher than one). A total of 9,423 articles out of the 13,240 ones are connected to the others through at least one incoming or outgoing citation.

In total, our data set includes 4,860 articles. Among these, there are 2,101 articles across ten surgical specialties (see section S2.2) that received at least one citation. Each of these 2,101 articles represents the seed of a diffusion cascade. In addition, the data set includes 1,049 articles across the ten specialties that did not receive any citations, but are part of chains of citations leading to articles in those specialties. Finally, the data set includes 1,710 articles that do not belong to any of the ten specialties, but are part of chains of citations leading to articles in those specialties.

S1.2 Visualisation of the citation network

Figure 8 in the main text shows the network of citations among the 9,423 articles that are connected to other articles through at least one incoming or outgoing citation. For unweighted networks, it is common practice to attribute unitary weight to the links. The algorithm used to visualize the networks is called ForceAtlas2, which is the default layout algorithm.[223] It implements an attraction-repulsion model aimed at simulating a physical system: unconnected nodes pull each other further apart, while links between connected nodes push the nodes closer to each other. The process of physically locating nodes depends only on connections between nodes, and the final result varies according to the initial state. Possible exogenous or endogenous attributes of nodes are not taken into account in locating them. The position of each node cannot be interpreted on its own, but needs to be compared with the position of

the other nodes, depending on the whole structure of connection. The energy model that underpins the ForceAtlas2 algorithm is based on two simple rules: (a) the attraction between two nodes depends linearly on the distance between them; and (b) the repulsion force between any two nodes that is inversely proportional to the distance between them, and is directly proportional to the product of their degrees. In this way, repulsion is weaker between a very connected node and a poorly connected node than between two very connected nodes. Given that in our specific application we are considering binary networks with unitary weights, the distance between any two nodes is set to be equal to one, and is constant across all connections in each network. Unlike other algorithms, and as a result of the interplay between the attraction and repulsion forces implemented, ForceAtlas2 produces densely populated spatial regions that identify structural communities. Moreover, ForceAtlas2 comes with a set of optional settings that can be used by users to optimize the visualization of networks. One of the settings used to depict networks is the "scaling" constant that is applied to stretch homogenously all the dyadic repulsion forces between nodes in order to expand the network distances without affecting the relative positioning of the nodes in the layout. We set the "scaling" constant at a value equal to two (default is one). In this way, the image becomes more readable, thus enabling an interpretation of the networks. We also use the "Gravity" option, and set the parameter at 70 (default is one). This prevents disconnected components from being pulled apart from each other. Moreover, it enables nodes within these components to be placed toward the center of the visualized network. Finally, we used the "LinLog Mode" option that applies a logarithmic function to the attraction force. This creates a visualization of the network that maximises the modularity of the community structure.

S2. Methods

S2.2 Partition of articles into groups related to surgical specialties

We clustered the articles into 17 groups. Among these, 16 correspond to 16 surgical specialties/procedures: prostatectomy, cardiac surgery, nephrectomy, hysterectomy, cystectomy, thyroidectomy, colectomy, pulmonary lobectomy, TORS, hepatectomy, oesophagectomy, gastrectomy, cholecystectomy, skull base surgery, transplantation and microvascular surgery. In addition, we created a group of articles that did not fall into any of the above 16 groups. In order to partition the articles into these groups, we searched for those articles containing words in their titles that uniquely identify the various surgical operations. For example, for robotic prostatectomy we identified 1,551 articles within the broad field of robotic surgery that contain "prostatectomy" or "prostate" in their titles.

We then selected ten of the 16 surgical specialties/procedures corresponding to those with a number of publications higher than 100 out of the original 13,240 articles that belong to the major subject area "Medicine" and are concerned with robotics. In particular, the following six specialties were filtered out: cholecystectomy, oesophagectomy, gastrectomy, transplantation, skull base surgery, and microvascular surgery. In total, our study focused on the following ten surgical specialties/procedures: prostatectomy, cardiac surgery, nephrectomy, hysterectomy, cystectomy, thyroidectomy, colectomy, pulmonary lobectomy, transoral robotic surgery (TORS), and hepatectomy. Figure 7 in the main text shows that these ten

groups of articles related to the robotic-assisted surgical specialties evolve over time in different ways.

We applied a pattern matching technique (*Regex*) for allocating each article among the 5,961 that received at least one citation to one of the main 16 surgical specialties, i.e., cardiac surgery, cholecystectomy, colectomy, cystectomy, oesophagectomy, gastrectomy, hepatectomy, hysterectomy, microvascular surgery, nephrectomy, prostatectomy, pulmonary lobectomy, skull base surgery, TORS, transplantation, and thyroidectomy.

Specifically, we identified all those articles containing in their title words or combination(s) of words related to the above mentioned 16 surgical specialties as detailed in Table S1. *Regex* (or *regular expression*) is a language for logical pattern matching. It is chiefly used to produce text patterns matching a given text. For example, using the regular expression language when we searched for all the titles containing either "prostatectomy" or "prostate" followed by or preceded by any text, we searched for a pattern written as "*prostat\w**". The expression "*\w**" after "*prostat*" (the common text pattern between the two words "prostatectomy" and "prostate") is meant to represent all the possible combinations of characters following the common string pattern. The symbol "|" in a regular expression stands for the logical conjunction "or", meaning that, for example, in "cystectomy| bladder" we searched for all titles containing either the word "cystectomy" or the word "bladder". Moreover, as occurred for example in the case of cardiac surgery, we have identified a set of words uniquely associated with all those articles pertaining to this specialty. Among them, there is also a combination of two words such as "coronary" and

"*bypass*" that can be also combined with "artery" to form "coronary artery bypass". In order to search for both "coronary bypass*"* and "coronary artery bypass" we used the string pattern "*coronary.*bypass*" that matches all those patterns within the title starting with the word "coronary*"*, ending with the word "bypass*"*, and (although not necessarily) containing some other object (e.g., word, space, hyphen) between those two.

Finally, to address possible differences in language, as for example in the case of oesophagectomy, depending on whether the authors use American English or British English, we can distinguish between titles containing "*oesopha\w**" or "*esopha\w**" as a common string pattern followed by any string of characters. In this case, the regular expression to use is "*o?esopha\w**".

Surgical specialty	*Regular expressions*
cholecystectomy	" cholecystectomy| gallbladder"
cardiac surgery	" heart| coronary| coronary.*bypass| cardiac| valve| atrial septal defect| atrial fibrilation| ablation| pacemaker leads| mitral| aortic"
colectomy	" colectomy| colorectal| colonic| ca?ec\w*| hemicolectomy| right| transverse| left| sigmoid\w*| rectum| anterior resection"
cystectomy	" cystectomy| bladder"
gastrectomy	" gastrectomy| stomach| sleeve"
hepatectomy	" hepatectomy| liver| hepatic"
hysterectomy	" hysterectomy| salpingo-oopherectomy| uterus| fibroid| endometri\w* "
microvascular surgery	" free.*flap| microvascular"
nephrectomy	" nephrectomy| renal"
oesophagectomy	" o?esopha\w*
prostatectomy	" prostat\w*"
pulmonary lobectomy	" pulmonary.*lobectomy| lung.* lobectomy| bronch\w*.*lobectomy| lung| pulmonary| bronch\w*"
skull base surgery	" skull.*base| pituitary"
thyroidectomy	" thyroid\w*"
TORS	" transoral.*robot| TORS"
transplantation	" transplant\w*"

Table S1. Regular expressions used to cluster articles into groups related to surgical specialties.

S2.2 Analytical definition of measures

S2.2.1 Citations

In the citation network, the nodes are the articles, while a directed link originates from one article to another when the former makes a citation to the latter. As citations can only point backward in time, from newer articles to older ones, there are no closed loops of directed links in citation networks. As such, these networks are acyclic.[73]

Each article in the network is characterized by the count of citations received from other articles belonging to the network that includes all published articles pertaining to robotic surgery. So constructed, the number of citations received by an article corresponds to the in-degree of the corresponding node in the citation network, i.e., the number of links pointing to that node. Let us define $\boldsymbol{A} = [a_{ij}]$ as the adjacency matrix describing the connections among all the articles (nodes) in the network, such that the entry a_{ij} is equal to one if there is a directed link (citation) from i to j, and zero otherwise. Thus, the count of citations cit_i received by an article i or, equivalently its in-degree k_i^{in} in the network of citations, can be defined as follows

$$cit_i = k_i^{in} = \sum_j a_{ji} .$$

S2.2.2 Cascades, structural depth and structural width

Any cited article can be considered a seed node s from which a cascade C_s originates. We define a cascade C_s as the set of nodes including the seed node s and all nodes connected to node s through direct or indirect links.

To calculate structural depth, for each seed node s, we computed all shortest directed paths originating from each pair of nodes belonging to the cascade C_s. Given two nodes i and j belonging to C_s, the shortest directed path from node i to node j is given by the length of the shortest sequence of directed links that separate node i from node j.[2] We denote with $d_{\overrightarrow{i,j}}$ the length of the directed shortest path from node i to node j within a given cascade, and with l the number of all directed shortest paths connecting all pairs of nodes in the cascade. Note that, given the directed nature of our network, the existence of a shortest directed path from node i to node j does not necessarily imply that a shortest path between node j and node i also exists. Therefore, because in a citation network paths are always directed backwards in time, and thus closed loops are not allowed, each pair of articles i and j contributes only once to the value of sd_{C_s} to the extent that a directed path exists from article i to article j, or vice versa.

Our measure of structural depth sd_{C_s} draws on, and extend, a classical graph property, the *Wiener index*, that was originally proposed in mathematical chemistry.[3] The application of the Wiener index to online diffusion cascades was recently introduced by Goel and colleagues.[4] Our definition of structural depth builds upon this analytical tradition. Like the measure of structural virality proposed by Goel and colleagues,[4] our measure takes into account all shortest paths; however, unlike the Wiener index and Goel *et al*.'s structural virality, our measure of structural depth explicitly accounts for the direction of links. In particular, given the acyclic directed nature of the

network of citations, we chose to focus only on directed shortest paths between pairs of articles.

An important property of structural depth is that it varies as a function of the size of the cascade only under certain conditions. Specifically, in the case of a stylized tree-like cascade in which each article is cited by at most k other articles (i.e., a k-ary rooted tree), sd_{C_s} will increase with the total number of articles N in the cascade only when $k << N$. Indeed, when $k << N$ is kept unchanged, an increase in N will amplify the number of generations in the cascade, thus increasing sd_{C_s}. In particular, in a complete k-ary tree-like cascade (i.e., a cascade in which each article that is not a leaf is cited by k other articles, and all articles that are leaves have the same length), sd_{C_s} will be proportional to the height of the cascade. Conversely, when $k \approx N$, an increase in N will only produce larger broadcasts, with virtually no effect upon sd_{C_s} (approaching the value of one in the limiting case of the star graph).

To calculate the structural width sw_{C_s} of cascade C_s, for each article i in C_s (including seed article s), we computed cit_i^n, namely the number of citations received by all articles citing article i, and then calculated the average of such values over all articles in C_s. To obtain cit_{C_s}, for each article i in C_s (including seed article s), we calculated cit_i, namely the number of citations received by i, and then calculated the average of such values over all articles in C_s.

Figure 9 (panels a and b) in the main text shows two simple cases that differ in terms of their cascade structure. Even though the two seed nodes have the same number of citations, they are associated with different values of structural depth and width. Table

S2 illustrates how the lengths of the shortest paths were calculated on each of the two cases. Table S3 illustrates how citations and second-step citations were calculated for each node in each of the cascades. Tables S4 shows how citations, structural depth and structural width were computed in each of the two examples.

Case a)		Case b)	
length of shortest path	**count**	**length of shortest path**	**count**
1	3	1	22
2	0	2	19
3	0	3	14
4	0	4	8

Table S2. Count of shortest paths by length in the cases in Figure 4 (panels a,b; main text).

Case a)			Case b)		
node	**citations**	**second-step citations**	**node**	**citations**	**second-step citations**
s	3	0	s	3	5
1	0	0	1	2	2
2	0	0	2	1	2
3	0	0	3	2	2
			4	2	2
			5	0	0
			6	2	4
			7	0	0
			8	2	2
			9	0	0
			10	2	0
			11	2	0
			12	2	0
			13	2	0
			14	0	0
			15	0	0
			16	0	0
			17	0	0
			18	0	0
			19	0	0
			20	0	0
			21	0	0
			22	0	0

Table S3. Count of citations and second-step citations in the examples in Figure 4 (panels a,b; main text).

Measure	Case a)	Case b)
Citations	3	3
Structural depth	3/3=1	(1×22+2×19+3×14+4×8)/63=2·127
Structural width	[(0×4)/4]/[(0×3+3×1)/4] = (0/4)/(3/4) = 0	[(0×16+2×5+4×1+5×1)/23]/[(0×12+1×1+2×9+3×1)/23] = (19/23)/(22/23)=0·86

Table S4. Citations, structural depth, and structural width in two examples in Figure 4 (panels a,b; main text).

Below we report values of structural depth for each of the examples in Figure 9 (panels a,b; main text) calculated using the formula introduced by Goel *et al*.,[4] and compare them with the ones calculated with our measure (structural depth).

Measure	Case a)	Case b)
Structural virality	(2×6+1×6)/12 = 1·5	(1×44+2×62+3×48+4×56+5×80+6×88+7×80+8×48)/506=4·76
Structural depth	3/3=1	(1×22+2×19+3×14+4×8)/63=2·13

Table S5. Structural virality based on Goel *et al*.,[4] and structural depth.

S2.3 Measuring innovation

S2.3.1 The innovation funnel and the innovation value chain

Innovation can hardly be defined unequivocally across domains. In particular, in the context of surgery, innovation may relate to manufacturing and advancement of medical devices, improvement in surgical technique or approach, change in service delivery process, or organizational reconfiguration that must be "altogether new, new to anatomic location, and new to the category of patient".[5]

Our study has benefitted from the use of two concepts originally applied to industrial innovation: the *innovation funnel* and the *innovation value chain*.[79,224] The *surgical innovation funnel* (SIF) depicts the process along which new ideas, once introduced (*idea generation*), progress until only those that are successful in being materialized can reach the stage of *idea conversion* or *materialization* (e.g., manufacturing of a new medical device, performance of novel first-in-human surgical technique, or introduction of a new service delivery model in surgical practice). As the funnel further narrows, many of these 'materialized' innovations will die out while a select minority will continue until they reach the final *implementation* stage (i.e., usage of a

new medical device or surgical technique in a number of patients but confined within a trial and/or institution). If implementation succeeds and the innovation is widely employed, then the diffusion process can be regarded as having reached its final stage. For instance, this occurs with the FDA-approval and market release of a new medical device, or the international uptake of new surgical technique or the introduction of a new service delivery model in routine surgical practice. This progressive 'natural selection' process underlying the innovation chain has inspired our measure of innovation value, the innovation index, that is meant to express the influential role of a surgical innovation as a function of the degree to which it has reached its final stage of implementation.

S2.3.2 Innovation index

In order to assess the level of success reached by innovative ideas in the ten groups defined by type of surgical procedures, we partitioned the innovation cycle into eight stages corresponding to eight ordered categories, numbered in descending order from eight to one (Table S6). Each article can be classified according to the method and level of evidence presented in the corresponding medical work. In particular, the least mature stage of a generic innovation path is the theoretical description of an idea or its laboratory evaluation (stage eight). An idea can then be applied and assessed through an animal study (stage seven), and subsequently through a cadaveric study (stage six), and so on along all the stages described in Figure 4 in the main text. The initial idea, if successful, will end up being employed in a randomized control trial (stage one), which represents the final implementation stage at which an innovation can be regarded as having implications for medical practice. For each surgical specialty, we

quantified the average progress of innovative ideas towards their final implementation stage using what we have called the innovation index (main text).

Stage	Description
1	Randomised controlled trial
2	Non-randomised controlled trial
3	Observational study with controls
4	Observational study without controls
5	First-in-human study
6	Cadaveric study
7	Animal study
8	Description of idea / laboratory evaluation

Table S6. Description of stages into which articles can be classified according to level of evidence.

The value of the innovation index based on the eight stages of innovation varies from zero to 2·718. In principle, the maximum value is reached by a surgical specialty that manages to monopolize the whole production of articles across all categories (from eight to one). In this limiting case, the values p_c^g / P_c for any category *c* are equal to one, and the innovation index corresponds to a harmonic series with eight terms, converging to 2·718. We then normalized the value of the Innovation Index by dividing it by the maximum that was observed across all surgical specialties, i.e., 1·11. The normalized values will then range between zero and one. We then multiplied the normalized values by 100 to obtain percent values.

To understand the logic behind the innovation index, let us consider the simple case in which we have only two surgical specialties. The first specialty ($g=1$) accounts for *2/3* of the whole scientific production at each stage from eight to five, and for *1/3* of the whole production at each stage from four to one. The second group ($g=2$) accounts for *1/3* of the production at each stage from eight to five, and for 2/3 of the production at each stage from four to one (see Table S7 for details). The innovation

index i_1 for the group *1* is then equal to *70*, while the innovation index i_2 for group *2* is equal to *100*. This means that group *2* contributes more than group *1* to relatively mature stages of innovation in surgery. As a result, group *2* is more successful than group *1* in bringing innovations to the stage at which they have implications for medical practice and thus create value for patients. Table S8 illustrates how the contributions to each stage are combined into a final normalized value of the innovation index.

Category	***c***	*8*	*7*	*6*	*5*	*4*	*3*	*2*	*1*
Weight	***1/c***	1/8	1/7	1/6	1/5	1/4	1/3	1/2	1/1
Contribution Group 1	p_c^1/P_c	2/3	2/3	2/3	2/3	1/3	1/3	1/3	1/3
Contribution Group 2	p_c^2/P_c	1/3	1/3	1/3	1/3	2/3	2/3	2/3	2/3

Table S7. Calculation of contributions to implementation stages in two illustrative examples.

Group 1	$\sum_{c\in S}\frac{1}{1}\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}+\frac{1}{3}\times\frac{1}{3}+\frac{1}{4}\times\frac{1}{3}+\frac{1}{5}\times\frac{2}{3}+\frac{1}{6}\times\frac{2}{3}+\frac{1}{7}\times\frac{2}{3}+\frac{1}{8}\times\frac{2}{3}=1.12$	(1.12/1.6) ×100	70
Group 2	$\sum_{c\in S}\frac{1}{1}\times\frac{2}{3}+\frac{1}{2}\times\frac{2}{3}+\frac{1}{3}\times\frac{2}{3}+\frac{1}{4}\times\frac{2}{3}+\frac{1}{5}\times\frac{1}{3}+\frac{1}{6}\times\frac{1}{3}+\frac{1}{7}\times\frac{1}{3}+\frac{1}{8}\times\frac{1}{3}=1.6$	(1.6/1.6) ×100	100

Table S8. Calculation of innovation index in two illustrative examples.

S2.3.3 National Inpatient Sample (NIS) count of robotic procedures in 2012

The 7,296,968 hospital discharges included in the National Inpatient Sample (NIS) 2012 data set have been extracted according to the ICD-9-CM procedure codes within the principal procedure field (PR1), as shown in Table S9.

Surgical procedure	IC9 Codes for principal procedures
Hysterectomy	All 68, all 69, (exc, 695-699), all 653, all 654, all 655, all 656, all 664, all 665, all 667
Prostatectomy	603, 604, 605, 606,
Nephrectomy	All 553, all 554, all 555
Pulmonary lobectomy	All 32
Cardiac surgery	All 36, 35 and 37
Colectomy	All 48, all 457, 452 all 49, all 458
Cystectomy	576. all 577
TORS	8131, 8101, 066, all 27, all 28, all 29, all 30, 316
Hepatectomy	All 50 (excl., 505, 506, 509)
Thyroidectomy	060, 061, 062, 063, 064, 065, 066, 064

Tables S9. IC9 codes for the selection of principal surgical procedures in NIS 2012.

In addition to this filter, we also used the secondary procedure fields (from PR2 to PR15) to extract hospital discharges that relied upon one or more secondary robotic procedures identified through the codes reported in Table S10.

	IC9 Codes for secondary procedures
Robotic assisted procedures	1741, 1742, 1743, 1744, 1745, 1749

Table S10. IC9 codes for the selection of robotic secondary procedures in NIS 2012.

Table S11 shows the rank of the ten surgical robotic procedures performed in 2012 as recorded in the NIS data set.

Surgical procedure	Count of robotic procedures
Hysterectomy	9,949
Prostatectomy	8,027
Nephrectomy	2,673
Pulmonary Lobectomy	844
Cardiac surgery	634
Colectomy	615
Cystectomy	419
TORS	145
Hepatectomy	55
Thyroidectomy	25

Table S11. Number of surgical procedures performed in 2012 according to the NIS data set.

S3. Robustness checks

S3.1 Robustness of the ranking by size, structural depth, and structural width

In order to assess the robustness of our analysis based on of the ranking of surgical specialties according to the median of the distributions of cascade size, structural depth and structural width, we quantified the Kendall rank correlation coefficients to measure the similarity between the median-based rankings and alternative rankings obtained using different descriptive statistics of the same distributions, including the 75th percentile, the coefficient of variation (CV), the quartile coefficient of dispersion, the median absolute deviation (MAD) and its normalized version, the mean absolute difference (MD), and the relative mean absolute difference (RMD). Table S12 reports results based on such tests, and indicates that, across all alternative descriptive statistics, observations have a statistically significantly similar rank.

Descriptive statistic	Kendall's tau correlation coefficient (size)	Kendall's tau correlation coefficient (structural depth)	Kendall's tau correlation coefficient (structural width)
75th percentile	0·719** (p=0·005)	0·778** (p=0·0009)	0·750** (p=0·004)
CV	-0·045 (p=0·928)	-0·689** (p=0·005)	0·295 (p=0·279)
Quartile coefficient of dispersion	0·629* (p=0·015)	0·644** (p=0·005)	-0·367 (p=0·241)
MAD	1** (p=0·0001)	0·822** (p=0·004)	0·614* (p=0·019)
normalised MAD	0·744** (p=0·005)	0·778** (p=0·0009)	-0·735** (p=0·005)
MD	0·629* (p=0·015)	0·733** (p=0·002)	0·796** (p=0·002)
RMD	0·089 (p=0·787)	0·689** (p=0·005)	-0·341** (p=0·207)

Table S12. Kendall's tau correlation coefficients. Tests of independence between rankings of the ten surgical specialties obtained using the median of their distributions of cascade size, structural depth and width, on the one hand, and the rankings based on other selected descriptive statistics. *: rank correlation coefficient is statistically significant at the 0·05 level; **: rank correlation coefficient is statistically significant at the 0·01 level (2-tailed tests).

S3.2 Test of independence between distributions of level of evidence

Figure 10 (panel f) in the main text reports our findings on the ranking by innovation index. We noted that the ranking comes as no surprise. For instance, robotic cardiac surgery is increasingly employed as it offers a minimally invasive approach for coronary revascularization and other procedures such as valve repair and closure of atrial septal defect through a closed chest.[225,226] By contrast, TORS, though a potentially promising treatment modality for oropharyngeal cancer and obstructive sleep apnoea, for now lacks the high-level evidence needed to support its use against traditional treatments (chemoradiotherapy and continuous positive airways pressure, respectively). In addition, the majority of the robotic instruments (not originally designed for transoral surgery) limit the applicability of TORS to a select minority of patients that can tolerate transoral access.[227,228]

For each surgical specialty, Table S13 reports the number of articles associated with level of evidence, the number of seed articles generating cascades, and the fraction of seed article with an associated level of evidence. The table suggests that not all cited articles across all specialties can be placed within the surgical innovation funnel.

Specialty	Count of articles with level of evidence	Count of cascades (i.e., seed articles)	Coverage (%)
Prostatectomy	623	835	72%
Cardiac surgery	310	374	83%
Nephrectomy	239	279	86%
Hysterectomy	180	194	93%
Cystectomy	116	149	78%
Thyroidectomy	77	87	89%
Colectomy	61	79	77%
Pulmonary lobectomy	57	70	81%
TORS	44	48	92%
Hepatectomy	37	44	84%

Table S13. Count of cascades associated with level of evidence across surgical specialties.

Figure S1 shows the distributions of ordinal categories associated with implementation stages according to level of evidence for each surgical specialty. For each of the 45 possible pairs of surgical specialties, we performed a two-way Chi-squared test of independence between the corresponding distributions of levels of evidence (Figure S1,a). In particular, the distributions of prostatectomy and cardiac surgery show a statistically significant difference with respect to all other procedures except one (thyroidectomy and hepatectomy, respectively). For robustness check, we also performed a Mann-Whitey U test of independence for each of the 45 pairs of specialties (Figure S1,b). The null hypothesis of lack of independence is rejected when p-value $< 0{\cdot}05$. Both tests produce very similar results, with only few exceptions.

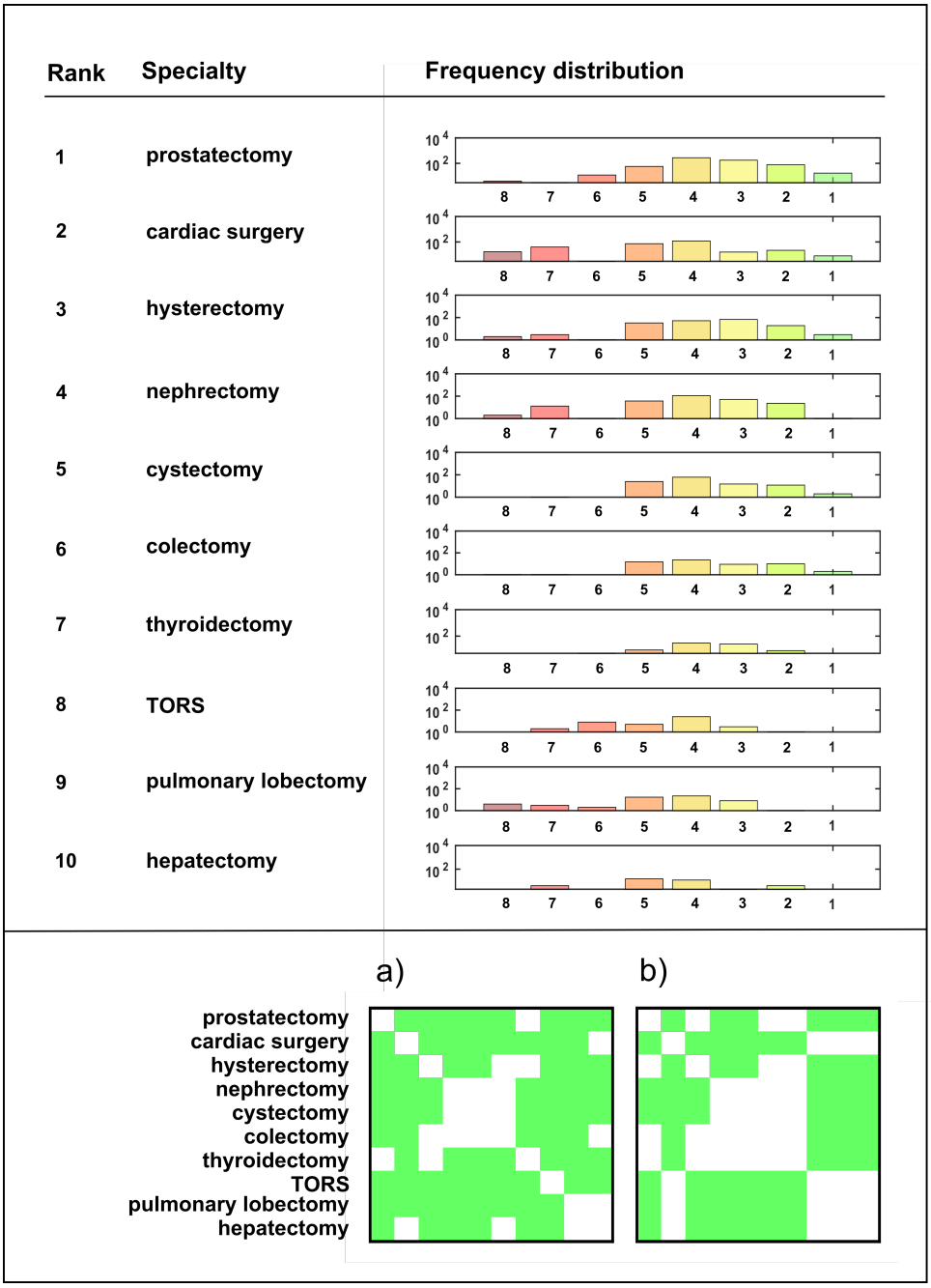


Figure S1. Frequency distributions of implementation stages and independence tests. Top panel: distributions of ordinal categories associated with implementation stages according to level of evidence for each surgical specialty. Bottom panel: a) two-way Chi-squared tests of independence and b) Mann-Whitney U tests of independence between distributions of implementation-based categories. Green squares correspond to values of the test statistic that reject the null hypothesis of non-independence with p-value $< 0{\cdot}05$.

S3.3 Spearman's and Kendall's tau tests for correlation between rankings

We further calculated Spearman's rank correlation coefficients (Table S14) and Kendall's tau coefficients (Table S15) to compare rankings of surgical specialties by number of citations to seed articles, fraction of citations within one year, cascade size, structural depth, structural width, innovation index, and count of robotic procedures in NIS. Results suggest that the rankings according to cascade size and structural depth are statistically significantly similar to the ranking by innovation index. By contrast, the ranking of surgical specialties by number of citations does not correlate significantly with either of the other rankings. Notice that results from Kendall's tau tests are consistent with the results reported in the main text.

	Innovation index	Count of procedures (NIS)
Citations	0·236 (p = 0·511)	
Fraction of citations	-0·188 (p = 0.603)	
Cascade size	0·758* (p = 0·011)	
Structural depth	0·782** (p = 0·008)	
Structural width	0·624 (p = 0·054)	
Innovation index		0·673* (p = 0·033)

Table S14. Spearman's rank correlation coefficients testing the independence between various pairs of rankings of surgical specialties. *: *p*-value < 0·05; **: *p*-value <0·01.

	Innovation index	Count of procedures (NIS)
Citations	0·2 (p = 0·421)	
Fraction of citations	-0·156 (p = 0·531)	
Cascade size	0·556* (p = 0·025)	
Structural depth	0·511* (p = 0·04)	
Structural width	0·378 (p = 0·128)	
Innovation index		0·511* (p = 0·04)

Table S15. Kendall's tau correlation coefficients testing the independence between various pairs of rankings of surgical specialties. *: p-value < 0·05; **: p-value <0·01.

Notice that the ranking of specialties by our measure of innovation index was found to closely match the one based on the NIS data set. This comes as no surprise. For example, robotic prostatectomy has long been the commonest approach for prostate cancer in the United States but increasingly also in Europe supported by level 1 evidence of superiority over traditional approaches.[229] By contrast, robotic thyroidectomy is a procedure originally developed in the Far East for cultural reasons with a remarkably poor uptake in the Western world.[82]

References

1. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 2014; **9**(6): e98679.
2. Newman MEJ. Networks An Introduction. New York: Oxford University Press; 2010.
3. Wiener H. Structural determination of paraffin boiling points. *J Am Chem Soc* 1947; **69**(1): 17–20.
4. Goel SA, Hofman J, Watts, DJ. The Structural Virality of Online Diffusion. *Management Science* 2016; **61**(1): 180–96.
5. Hutchison K, Rogers W, Eyers A, Lotz M. Getting clearer about surgical innovation: A new definition and a new tool to support responsible practice. *Ann Surg* 2015; **262**(6): 949–54.
6. Dunphy SM, Herbig PR, Howes ME. The innovation funnel. *Technological Forecasting and Social Change* 1996; **53**: 279–92.
7. Hansen MT, Birkinshaw J. The innovation value chain. *Harvard Business Review* 2007; **85**(6): 121–30.
8. Kim JE, Jung SH, Kim GS, et al. Surgical outcomes of congenital atrial septal defect using da VinciTM surgical robot system. *Korean J Thorac Cardiovasc Surg* 2013; **46**(2): 93–7.
9. Whellan DJ, McCarey MM, Taylor BS, et al. Trends in robotic-assisted coronary artery bypass grafts: A study of the Society of Thoracic Surgeons Adult Cardiac Surgery Database, 2006 to 2012. *Ann Thorac Surg* 2016; **102**(1): 140–6.

10. Arora A, Chaidas K, Garas G, et al. Outcome of TORS to tongue base and epiglottis in patients with OSA intolerant of conventional treatment. *Sleep Breath* 2016; **20**(2): 739–47.

11. Arora A, Kotecha J, Acharya A, et al. Determination of biometric measures to evaluate patient suitability for transoral robotic surgery. *Head Neck* 2015; **37**(9): 1254–60.

12. Robertson C, Close A, Fraser C, et al. Relative effectiveness of robot-assisted and standard laparoscopic prostatectomy as alternatives to open radical prostatectomy for treatment of localised prostate cancer: a systematic review and mixed treatment comparison meta-analysis. *BJU Int* 2013; **112**(6): 798–812.

13. Arora A, Garas G, Sharma S, et al. Comparing transaxillary robotic thyroidectomy with conventional surgery in a UK population: A case control study. *Int J Surg* 2016; **27**: 110–7.

Appendix 2

S.1 Data

Drawing on the Web of Science (WOS) database, we searched for all publications concerning robotic surgery according to the following search query:

- Topic*: "robotic OR robotic OR robot assisted OR robotic assisted OR robotically assisted OR robot-assisted OR robotic-assisted OR robotically-assisted"*;
- Research Area*: "Surgery"*;
- Document Types*: "Article"*.

The search was performed on the 17th of January 2017, and produced 3,889 publications (peer-reviewed articles) published between July 1988 and January 2017. Figure S29 shows the distribution of the retrieved publications across the years of retrieved.

Of the 3,889 articles collected, 23 did not specify the affiliations of the co-authors. Thus, we focused on the remaining 3,866 articles with specified affiliations of co-authors. Based on these 3,866 articles, a process of manual disambiguation of name variants of organisations was performed. Out of 2,590 name variants, we filtered out 1,700 unique names. To this end, each name variant was searched for on the WOS "Organisations - Enhanced List" tool with a view to identifying the preferred name associated with the corresponding organization. If such name was found, then it was used as a unique name. In particular, 816 preferred names were found corresponding

to 1,485 name variants. Not all organizations in the WOS database were included in this list by the WOS team. For 1,105 name variants we could not find a corresponding preferred name using the above mentioned tool. A manual search on Google search was then performed that took into account the associated geographical location (country and city) of the organization and, when necessary, also the affiliated authors' names. This resulted in the identification of additional 884 unique names of organizations. Thus, in total we obtained 1,700 unique organizations.

Moreover, these 1,700 organisations were classified into 642 academic institutions, 11 academic systems, 863 health providers, 99 enterprises, 53 research institutes, 22 governmental entities, 8 research foundations, and 2 research working groups. These institutional categories were produced partly based on the WOS *Organisation Type* classification produced through the InCites© (Clarivate Analytics, London, UK) data intelligence tool, and partly manually (mainly for enterprises and health providers that were not included in InCites).

The 1,700 organisations were nested within 62 countries, and within 6 geographical regions. These regions were based on the standard international allocation of countries into continents, with the exception of Russia and Middle Eastern countries that were classified as transcontinental countries. In particular, Russia was classified as belonging to Europe given the geographical proximity of its capital to the European borders.

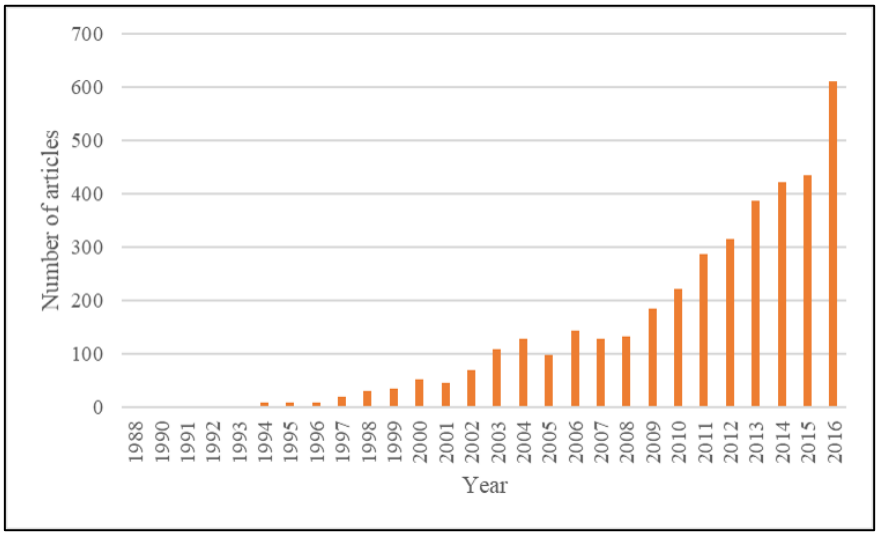


Figure S29. Time distribution of publications.

For each of the top 20 countries ranked by number of publications, Table S1 shows the number of organisations contributing to research.

Country	Number of publications	Number of organisations
USA	1846	520
Germany	359	137
South Korea	352	58
Italy	280	127
Japan	224	112
France	204	103
China	179	71
United Kingdom	167	68
Canada	159	33
Netherlands	88	32
Austria	86	20
Turkey	79	49
Belgium	61	17
Spain	46	41
Switzerland	40	14
Australia	35	31
Brazil	34	26
Taiwan	34	22
Romania	30	17
India	27	18

Table S5. Top 20 countries ranked by number of published articles.

S.2 The network

We built up the global collaboration network (see Figure 12 in the main manuscript) among the 1,700 organisations that carried out research on robotic surgery. The nodes of the network are the organisations, and links refer to collaborations between organisations. Specifically, a link exists between any two organisations when one or more scholars affiliated with one organisation co-authored at least one article with one or more scholars affiliated with the other organisation. Out of the 1,700 organisations, 1,543 were connected through at least one collaboration. The largest connected component includes 1,334 organizations, and in total the network comprises 6,000 collaboration links.

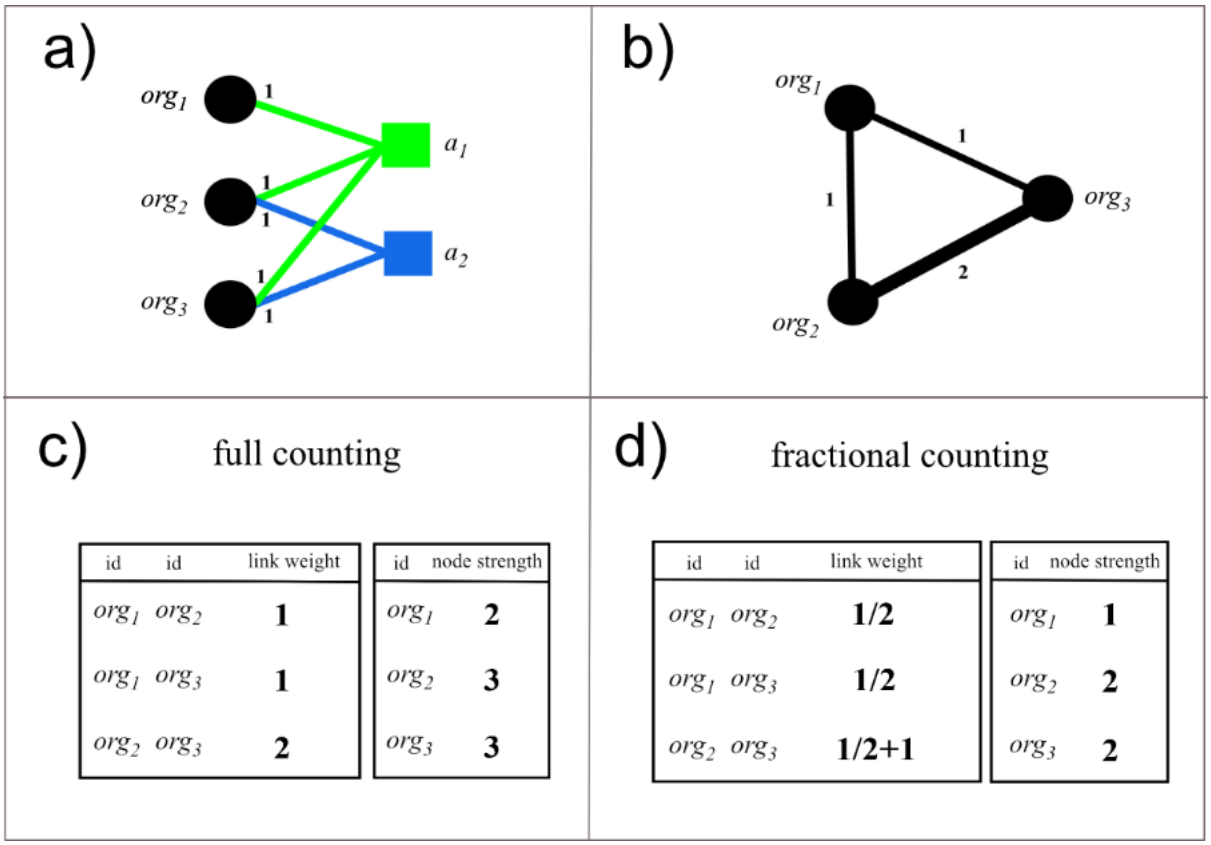
To construct the network, we first define $\boldsymbol{A}$ as the undirected and binary bipartite authorship matrix defining connections between each organisation org and each publication p such that element $a_{a,org} = 1$ if and only if organisation org contributed to article a, and $a_{a,org} = 0$ otherwise (see Figure S2.a).[1] We then define $\boldsymbol{C}$ as the undirected and weighted projected one-mode collaboration matrix, where each element $c_{ij} > 0$ if and only if organisation i collaborated with the organisation j. That is, $c_{ij} > 0$ when organisation *i* and organisation *j* co-authored at least one publication (see Figure S2.b).

The weight of each collaborative tie between any two organisations is measured according to the method proposed by Newman (2001).[2] In particular, the total weight c_{ij} of the collaborative tie between two organisations *i* and *j* is assumed to be equal to

the sum of the strengths w_{ij} contributed by each of the articles co-authored by the two organisations, where strength w_{ij} is defined as:

$$w_{ij} = \frac{1}{n_a - 1},$$

and n_a is the number of organisations appearing in the byline of article a. In other words, the total weight c_{ij} of the collaborative tie between organisations i and j is (i) proportional to the number of articles these organisations co-authored; and (ii) inversely proportional to the number of organisation involved in each article (see Figure S2,d). Note that all articles published by one single organization were excluded from the above measure as they did not contribute to the co-authorship network.



id	id	link weight
org_1	org_2	1
org_1	org_3	1
org_2	org_3	2

id	node strength
org_1	2
org_2	3
org_3	3

id	id	link weight
org_1	org_2	1/2
org_1	org_3	1/2
org_2	org_3	1/2+1

id	node strength
org_1	1
org_2	2
org_3	2

Figure S2. Authorship network, collaboration network, full and fractional counting method for computing weights of collaborative ties. a) The bipartite network connecting organisations to

publications; b) The weighted undirected one-mode collaboration network among organisations; c) Weights of links and node strength (i.e., the sum of the weights of the links incident upon the node) computed according the full counting method; d) Weights of links and node strength computed according to the fractional counting method (Newman, 2001).[2] Notice that in the case of fractional counting, node strength is equal to the number of articles the corresponding organisation published.

S.3 Measures

In this Section, we provide details on various dependent and independent measures used in the analysis. These measures include: innovativeness; research impact; the clustering coefficient; efficiency; closeness centrality; geographical dispersion; and strength of industrial collaboration.

S.3.1 The innovation index

We classified each article in terms of its level of evidence following the procedure proposed by Garas et al. (2017).[3] Each articles was associated with one out of 8 categories corresponding to distinct surgical innovation stages as shown in Table S2.

Surgical Innovation Stage	Description
1	Randomised controlled trial
2	Non-randomised controlled trial
3	Observational study with controls
4	Observational study without controls
5	First-in-human study
6	Cadaveric study
7	Animal study
8	Description of idea / laboratory evaluation

Table S2. Description of surgical innovation stages with which articles can be associated according to their level of evidence.

We could assign level of evidence to 3,389 articles. Figure S.3 shows the temporal distributions of these articles in each evidence-based category.

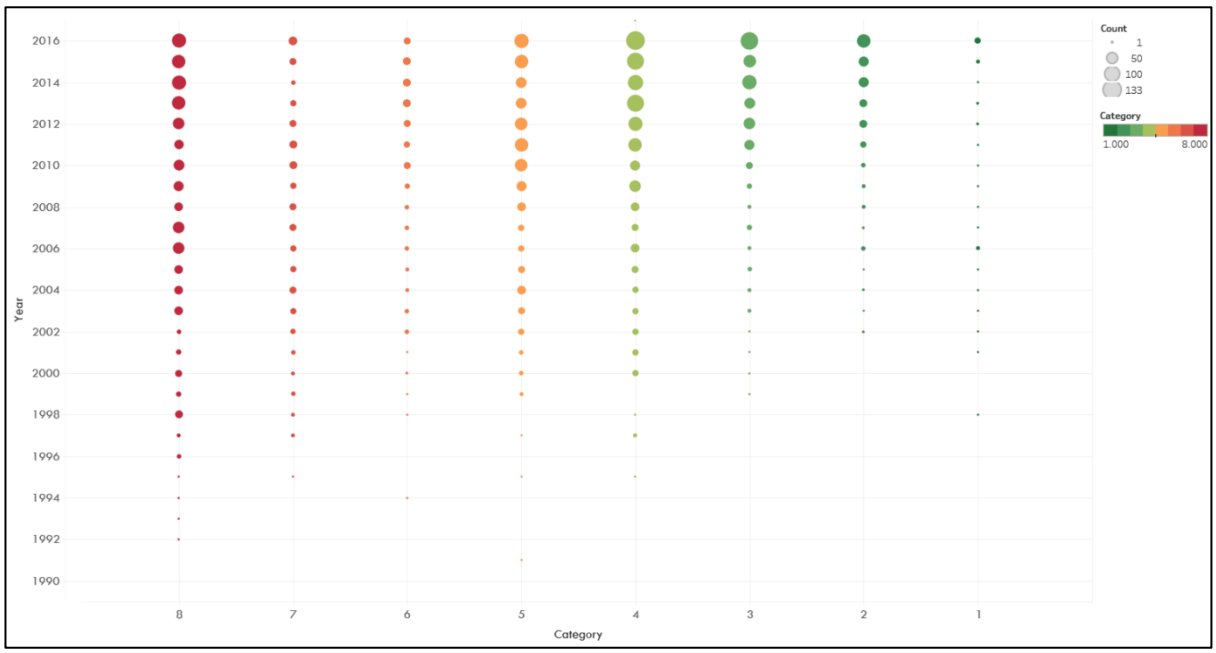


Figure S3. Number of articles per level of evidence over time.

We then computed the *innovation index* following the definition given in Garas *et al*. (2017).[3] For each organization *org*, the innovation index ii_{org} can be defined as

$$ii_{org} = \frac{\sum_{c \in S} \frac{1}{c} \times \frac{p_c^{org}}{P_c}}{\max_{org \in G} \left(\sum_{c \in S} \frac{1}{c} \times \frac{p_c^{org}}{P_c} \right)} \times 100,$$

where $c \in S = [1, \ldots, 8]$ is an integer value labelling the ordinal category associated with the implementation stage according to level of evidence, p_c^{org} is the count of publications of organisation *org* associated with category c, and $P_c = \sum_{org \in G} p_c^{org}$ is the total number of publications associated with category c across all organisations in the data set (i.e., in the set G). Figure S4 shows the ranking of countries according to the average of innovation indices of all the organisations within their borders. The figure also highlights the most innovative organisations within and between countries.

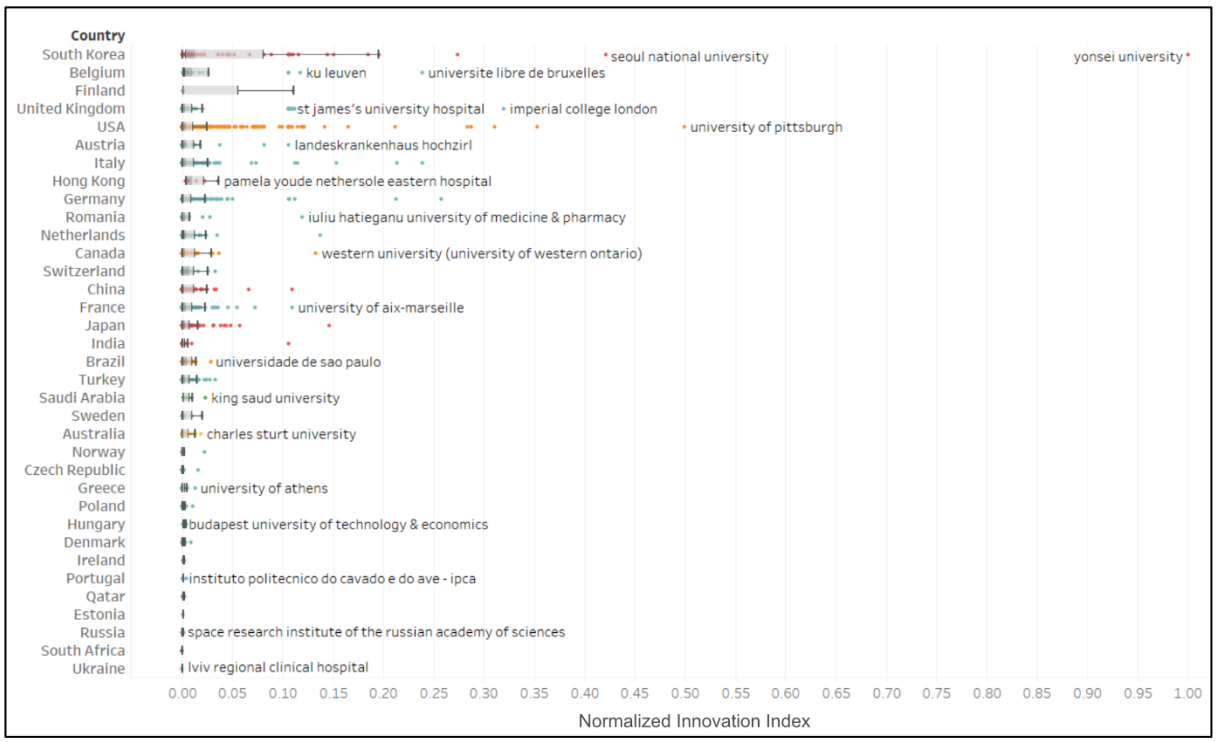


Figure S4. Ranking of countries and organisations according to normalized innovation index. To obtain the normalized innovation index, we divided each organisation's innovation index by the maximum value observed in the data set. Countries have been ranked according to the sum of organisations' normalised innovation index. Grey shaded areas correspond to distances between lower and upper quartiles (interquartile range). Darker shaded areas denote distances between lower quartile and median. Whiskers correspond to values within 1.5 times the interquartile range. To obtain the normalised innovation index, we divided each organisation's innovation index by the maximum value observed in the data set.

S.3.2 Research impact

For each organisation, we measured its academic performance using the sum snc_{org} of the normalised citations received by all articles published by the organisation across the years:

$$snc_{org} = \sum_{y}\sum_{a} nc_{a_{org,y}},$$

where $nc_{a_{org,y}}$ refers to the normalised citations received by article $a_{org,y}$ published by (authors affiliated with) organisation *org* in year *y*. To obtain normalised citations, the citation count for each publication in a given year was divided by the average number of citations obtained by all articles published in the same year:

$$nc_{a_{org,y}} = \frac{cit_{a_{org,y}}}{< cit_{a,y} >},$$

where the average $< cit_{a,y} >$ was computed across all articles published by all organisations in year *y*.

S.3.3 The local clustering coefficient

Network analysis provides a set of measures for quantifying a number of structural properties at different levels of a system.[1] For example, measures can be calculated at a node's level to detect its topological position (e.g., broker between otherwise disconnected partners) in the ego-centered network (i.e., the network including connections between the node and its partners as well as connections between these partners), or in the whole network (e.g., how close the node is to all other others in the network). Network analysis can therefore provide valuable tools for uncovering which nodes are central to the innovation process ('knowledge catalysts'), which have a more peripheral role, and which act as connectors between otherwise unconnected groups ('knowledge brokers'). By constructing and examining each node's ego-centered network, we aim to detect the extent to which a node's neighbours are also

connected with each other, and in so doing we engage with current debates on social capital and the relative benefits of closed versus open social structures.[1]

Using the weighted network as defined above, we computed the local clustering coefficient of each organisation, which measures the extent to which the organisation's ego-centered network is a closed collaborative structure.[1-5] Formally, the local clustering coefficient cc_{org} of organisation org in a binary network is defined as the ratio between the number of actual triangles t_{org} containing the organisation org and its k_{org} neighbours, and the maximum possible number of such triangles, $t_{org}^{max} = k_{org}(k_{org} - 1)/2$:

$$cc_{org} = \begin{cases} \frac{2 \times t_{org}}{k_{org}(k_{org}-1)} & \text{for } k_{org} \geq 2 \\ 0 & \text{for } k_{org} = 0,1 \end{cases}.$$

In our study, we computed the clustering coefficient only for organisations with $k_{org} \neq 0,1$ (i.e., 1,203 organizations). Moreover, we computed the generalised clustering coefficient to account for the (fractional) weight of links according to the definition proposed in Opsahl and Panzarasa (2009).[4] In particular, we used the arithmetic mean as a method to obtain the weights of triangles and triplets, respectively in the numerator and denominator of the generalized clustering coefficient.

S.3.5 Efficiency

To measure an organisation's opportunities of brokerage, we also computed efficiency, namely the degree to which the organisation's ego-centered network is

non-redundant.[6] Efficiency thus captures the extent to which an organisation's ego-centered network is an open structure. To measure efficiency, we first computed effective size, a measure originally introduced by Burt (1992) to capture the extent to which each of the first neighbours of a node is redundant with respect to the other neighbours. Following Latora *et al*. (2013),[5] an organisation *org*'s effective size, es_{org}, can be defined as

$$es_{org} = k_{org} - (k_{org} - 1)cc_{org},$$

where k_{org} is the degree of organisation *org* (i.e., the number of organisations with which *org* collaborated), and cc_{org} is the local binary local clustering coefficient (i.e., the coefficient presented in S.3.3 computed using the unweighted network).

The efficiency eff_{org} of organisation *org* (i.e., the normalized effective size) can now be computed as the ratio between the organisation's effective size and degree:

$$eff_{org} = \begin{cases} \frac{es_{org}}{k_{org}} = 1 - \frac{k_{org} - 1}{k_{org}} cc_{org} & k_{org} > 0 \\ 0 & k_{org} = 0 \end{cases}$$

Notice that, as with the local clustering coefficient (Section S.3.3), efficiency was computed only for organisations with $k_{org} \neq 0, 1$ (i.e., 1,203 organisations).

S.3.4 Closeness centrality

Drawing on the weighed network defined above, we computed the closeness centrality of each organisation. The closeness centrality of an organisation is a

function of the inverse of the shortest distances separating the organisation from each of all other organisations in the network (i.e., the smallest number of links separating the organisation from each of the others).[1] Formally, the closeness centrality l_{org} of a given organization *org* is defined as:

$$l_{org} = \frac{1}{n-1} \sum_{j(\neq org)} \frac{1}{d_{org,j}},$$

where n is the total number of organisations in the data set, and $d_{org,j}$ is the length of the shortest path between organisation org and organisation j. In other words, we measured the closeness centrality of organisation *org* as the mean of the inverse distances from *org* to all other nodes in the network.

Once again, we used the generalised version of closeness centrality for undirected networks to account for the (fractional) weights of links according to the definition proposed in Opsahl *at al.* (2010).[7] In particular, we set the alpha tuning parameter in the generalised measure equal to one (see Equation 8 in Opsahl et al. (2010)). In this case, the Dijkstra shortest paths were used, and the identification procedure of these paths relied simply on tie weights and disregarded the number of intermediary nodes along the paths.

S.3.5 Geographical entropy

In addition to the network-based measures described above, for each organisation we measured: (i) the diversity of the geographical locations of collaborators; and (ii) the strength of collaboration with industrial partners.

To measure the geographical diversity of collaborators, we computed geographical entropy. To this end, let us define Γ_{org} as the set of countries in which the collaborating organisations of a given organisation *org* are located, i.e., $\Gamma_{org} = \{\gamma_1, \gamma_2, \dots, \gamma_f\}$. Formally, the geographical entropy of organisation *org* can be defined as follows:

$$\varepsilon_{org} = -\sum_{\gamma_i \in \Gamma_{org}} p_{org,\gamma_i} \, log(p_{org,\gamma_i}),$$

where p_{org,γ_i} is the fraction of organisation *org*'s collaborative effort towards organisations located in country γ_i (i.e., the ratio between the sum c_{org,γ_i} of the weights of links between *org* and all organisations in country γ_i and the sum $\sum_{\gamma_i \in \Gamma_{org}} c_{org,\gamma_i}$ of the weights of links between *org* and all organisations in Γ_{org}). Notice that each weight c_{org,γ_i} is measured once again according to the method proposed by Newman (2001) (see Section S.2).

Given the set Γ of countries in the data set and the norm $\|\Gamma\|$ of such set, an organisation's geographical entropy can range from zero (when all international collaborators are located in the same country) to $log(\|\Gamma\| - 1)$ (when the international collaborators are uniformly distributed across $\|\Gamma\| - 1$ countries, i.e., all countries except the organisation's own). Similarly, given the set Γ_{org} of countries with which the organisation *org* collaborates, entropy is maximised when the organisation equally distributes its collaborative efforts across all countries in Γ_{org}. That is, for each country $\gamma_1, \gamma_2, \dots, \gamma_f \in \Gamma_{org}$, ε_{org} is maximised when each fraction

of collaborative effort is equal to $p_{org,\gamma_i} = \frac{1}{\|\Gamma_{org}\|}$, where $\|\Gamma_{org}\|$ is the norm of the set Γ_{org}.

Thus, geographical entropy enables us to assess organisations beyond the mere count of international collaborators, and to distinguish between organisations that, while having the same number of international collaborators, differ in terms of their geographical dispersion. This is important as some organisations may have a large number of international collaborators, yet all confined within the same geographical boundaries, whereas others may have few collaborators that nonetheless are widely geographically dispersed.

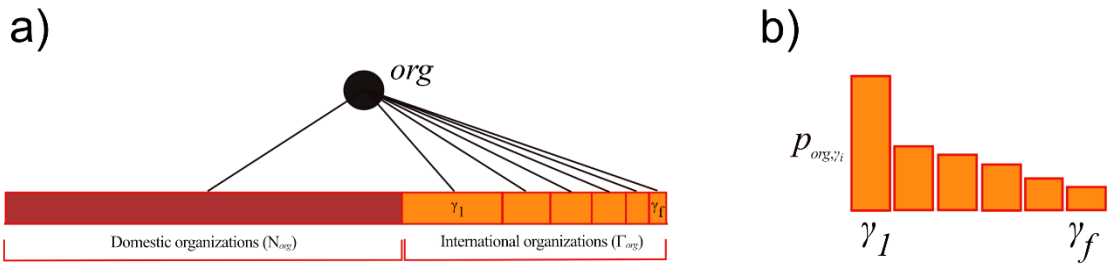


Figure S30. Representation of geographical dispersion. a) An organisation *org* can collaborate with domestic or international organisations, D_{org} and Γ_{org}, respectively. Each of the international organisations can, in turn, be associated to the foreign country $\gamma_1, \gamma_2, \ldots, \gamma_f \in \Gamma_{org}$ in which it is located. The size of each cell γ_i is proportional to the effort $c_{i\gamma_i}$devoted by organisation *org* towards country γ_i (i.e., proportional to the sum of weights of the links between *org* and all organisations in country γ_i). b) The distribution of organisation *org*'s efforts across foreign countries. The value of each p_{org,γ_i} is equal to the ratio between $c_{org,\gamma_i}$and the sum of org's efforts towards all countries in Γ_{org}, i.e., $\sum_{\gamma_i \in \Gamma_{org}} c_{org,\gamma_i}$. Based on the distribution of p_{org,γ_i}, it is possible to compute the geographical entropy ε_{org}.

S.3.6 Industrial collaboration

To assess the association between research impact and innovation on the one hand, and collaborative links with industry on the other, for each organisation we measured the strength of industrial collaborations. Indeed collaboration with industry represents a potentially relevant determinant of research performance that may not be captured by more traditional network metrics.[8]

The strength of an organisation's industrial collaborations was computed as the sum of the organisation's collaborative efforts towards industrial partners. Formally, the strength of industrial collaborations of a given organisation *org* is defined as:

$$ic_{org} = \sum_{j \neq org, j \in I} c_{org,j},$$

where $c_{org,j}$ denotes the weight of the collaborative link between organisation *org* and organisation *j*, and *I* represents the set of all organisations classified as "Industry". Notice that each weight $c_{org,j}$ is measured once again according to the method proposed by Newman (2001) (see Section S.2).

S.3.7 Other control variables

We also controlled for a measure of volume, i.e., the 'scholarly output' in WOS, here referred to as 'number of articles in WOS'. This was measured through the number of all articles published by each organization that the InCites© intelligence tool could

retrieve in the WOS database. Note that this control variable refers to each organisation's overall scholarly output in WOS beyond the specific field of robotic surgery. This, of course, should not be confused with the number of publications in our data set that are linked to any given organisation. Moreover, WOS does not track all the organisations we were able to detect in our set of publications (within robotic surgery). Out of the 1,700 organisations in our data set, only 814 could be associated with a total scholarly output in WOS.

S.4 Statistical analysis

Maximum-likelihood estimates of two hierarchical three-level random-intercept models were computed in which organisations were nested within countries, in turn nested within geographical regions. A random-intercept structure was combined with heteroskedastic level-1 residuals by letting the variances of these residuals be a function of the organisation's institutional type. For organisation *org*, country γ, and geographical region *r*, we estimated the following three-level linear random-intercept models:

$$snc_{org,\gamma,r} = \beta_1 + \beta_2 cc_{org,\gamma,r} + \beta_3 l_{org,\gamma,r} + \beta_4 \varepsilon_{org,\gamma,r} + \beta_5 ic_{org,\gamma,r} + \beta_6 v_{org,\gamma,r} + \sum_{k=7}^{13} d_{k,org,\gamma,r}\beta_k + \zeta_{\gamma,r}^{(2)} + \zeta_r^{(3)} + \varepsilon_{org,\gamma,r}$$

and

$$ii_{org,\gamma,r} = \beta_1 + \beta_2 cc_{org,\gamma,r} + \beta_3 l_{org,\gamma,r} + \beta_4 \varepsilon_{org,\gamma,r} + \beta_5 ic_{org,\gamma,r} + \beta_6 v_{org,\gamma,r} + \sum_{k=7}^{13} d_{k,org,\gamma,r}\beta_k + \zeta_{\gamma,r}^{(2)} + \zeta_r^{(3)} + \varepsilon_{org,\gamma,r},$$

where $cc_{org,\gamma,r}, l_{org,\gamma,r}, \varepsilon_{org,\gamma,r}, ic_{org,\gamma,r},$ and $v_{org,\gamma,r}$ are level-1 covariates (as defined in Section S.3) for organisation *org* in country γ and region *r*, $v_{org,\gamma,r}$ is a level-1 covariate measuring the number of all articles published by organisation *org* in country γ and region *r* and recorded in the WOS database, $d_{k,org,\gamma,r}$ is the dummy variable for the *k*-th institutional type of organisation *org* in country γ and region *r*, β_1 to β_{13} are fixed parameters, $\zeta_{\gamma,r}^{(2)}$ is the country-level (i.e., level-2) random intercept, $\zeta_r^{(3)}$ is the region-level (i.e., level-3) random intercept, and $\varepsilon_{org,\gamma,r}$ is the organisation-level (i.e., level-1) error term.[8]

We let all observed covariates for all organisations in region *r* be denoted by the matrix $\boldsymbol{X_r}$. It was assumed that the region-level random intercept $\zeta_r^{(3)}$ has zero expectation and variance $\psi^{(3)}$, given the covariates $\boldsymbol{X_r}$. Similarly, we assumed that the country-level random intercept $\zeta_{\gamma,r}^{(2)}$ has zero expectation and variance $\psi^{(2)}$, given the random intercept $\zeta_r^{(3)}$ and the covariates $\boldsymbol{X_r}$. That is, we assumed that: (i) there is zero correlation between covariates and the random intercept at the region level (i.e., level-3 exogeneity); there is zero correlation between covariates and the random intercept at the country level (i.e., level-2 exogeneity); there is zero correlation between random intercepts $\zeta_r^{(3)}$ and $\zeta_{\gamma,r}^{(2)}$ across regions and countries; the variance of the random intercept at the region level is homoskedastic given the covariates; and (iv) the variance of the random intercept at the country level is homoskedastic given the covariates and the region-level random intercept.

Moreover, it was assumed that the level-1 error term $\varepsilon_{org,\gamma,r}$ has zero expectation, given the covariates $\boldsymbol{X_r}$ and the random intercepts $\zeta_r^{(3)}$ and $\zeta_{\gamma,r}^{(2)}$. That is, we assumed

that there is zero correlation between covariates and level-1 residuals (i.e., level-1 exogeneity), and zero correlation between level-1 residuals and both random intercepts $\zeta_r^{(3)}$ and $\zeta_{\gamma,r}^{(2)}$. However, unlike the two random intercepts, the level-1 residuals were not assumed to be homoskedastic. To this end, we combined a three-level random-intercept structure with heteroskedastic level-1 residuals by letting the variances $\vartheta^{(t)}$of these residuals be a function of the organisation's institutional type $t_{org,\gamma,r}$, given the covariates and level-3 and level-2 random intercepts. That is, the conditional variance of the responses given the covariates and random intercepts, or the conditional variance of the total residual, depends on the value of the covariate $t_{org,\gamma,r}$, and the total residual is therefore heteroskedastic. Finally, the random effects at each level were assumed to be mutually independent and independent of the level-1 residual error term.

S.5 Results

For the sake of completeness, Tables S3-S8 report estimates for all parameters, including the remaining fixed effects (i.e., intercept and institutional types) as well as the estimated random-effect parameters (i.e., the standard deviations of the level-3 and level-2 random intercepts and of the level-1error term, $\sqrt{\psi^{(3)}}$), $\sqrt{\psi^{(2)}}$, $\sqrt{\vartheta^{(t)}}$ respectively). To check for robustness, we also estimated a number of additional three-level random-intercept models. First, we replaced the local clustering coefficient with efficiency, a measure that directly captures the extent to which an organisation collaborates with non-redundant partners that are otherwise disconnected with each other. Second, we introduced a level-1 interaction term between (standardised) geographical entropy and the local clustering coefficient.

S.5.1 Estimates from mixed-effects regression models of research impact

Table S3 reports the full range of estimates from the hierarchical three-level random-intercept model of research impact including the local clustering coefficient as a covariate (as in the main manuscript).

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	1.754686	0.986096	0.583233	5.27906
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	1.645659	1.159483	0.413626	6.54744
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	20.37447	0.706987	19.03487	21.80834
Academic system	5.749711	1.93462	2.973309	11.11865
Industry	3.104733	0.894139	1.765569	5.459635
Government	5.184769	1.134586	3.376451	7.961564
Health	8.482241	0.567122	7.440451	9.669898
Research Institute	13.80572	1.863823	10.59604	17.98764
Number of observations = 639 Log pseudolikelihood = 2678.401 Wald $\chi^2(10)$ = 458.82 Log likelihood = -2678.401				

Table S3. Maximum-likelihood estimates from three-level random-intercept linear models of normalized citations with the local clustering coefficient as covariate.

An increase by one standard deviation in (normalised) closeness centrality was associated with an expected increase of 6.0782 units in normalized. After controlling for the other covariates, a one-unit increase in geographical entropy was associated

with an expected increase in normalised citations by 10.5599 units. Estimates suggested that there was no significant difference in (normalized) citation count between two organisations that differed in number of publications, after controlling for the other covariates.

Table S4 reports the full range of estimates from the hierarchical three-level random-intercept model of research impact in which the local clustering coefficient has been replaced by efficiency. Results are in qualitative agreement with the ones reported in the main manuscript. A surge in opportunities of knowledge brokerage are associated with an expected increase in research impact, after controlling for the other covariates; yet the association is not statistically significant.

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	1.809977	1.022269	0.598297	5.475565
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	1.774071	1.100912	0.525717	5.986737
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	20.33417	0.702829	19.00227	21.75942
Academic system	5.733436	1.915343	2.978914	11.03499
Industry	3.270703	0.934917	1.867792	5.727348
Government	5.072872	1.117939	3.29359	7.813366
Health	8.503419	0.571205	7.454445	9.700002
Research Institute	13.7129	1.856314	10.51725	17.87955
Number of observations = 639 Log pseudolikelihood = -2678.2335 Wald $\chi^2(10)$ = 458.98 Log likelihood = -2678.2335				

Table S4. Maximum-likelihood estimates from three-level random-intercept linear models of normalized citations with efficiency as covariate.

S.5.2 Estimates from mixed-effects regression models of innovation

Table S5 reports the full range of estimates from the hierarchical three-level random-intercept model of innovation including the local clustering coefficient as a covariate.

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	5.95E-09	5.35E-08	1.30E-16	0.272217
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	2.97E-09	1.08E-08	2.39E-12	3.69E-06
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	1.357163	0.046495	1.269027	1.45142
Academic system	2.729741	0.865797	1.466034	5.082749
Industry	0.105642	0.028544	0.062208	0.179403
Government	0.142781	0.035741	0.087417	0.233209
Health	0.519942	0.033448	0.45835	0.58981
Research Institute	0.665177	0.090613	0.50931	0.868743
Number of observations = 639 Log pseudolikelihood = -914.73492 Wald $\chi^2(10)$ = 181.68 Log likelihood = -914.73492				

Table S5. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with the local clustering coefficient as covariate.

As clustering increases by one standard deviation, an organisation's innovation value decreases by 0.0820 units. An increase by one standard deviation in (normalised) closeness centrality was associated with an expected increase of 0.2166 units in innovation value. After controlling for the other covariates, a one-unit increase in geographical entropy was associated with an expected increase in innovation value by 0.1629 units. Estimates suggested that there was no significant difference in

innovation value between two organisations that differed in number of publications, after controlling for the other covariates.

Table S6 reports the full range of estimates from the hierarchical three-level random-intercept model of innovation in which the local clustering coefficient has been replaced by efficiency. Results are in qualitative agreement with the ones reported in the main text. A surge in opportunities of knowledge brokerage are associated with an expected increase in innovation value, after controlling for the other covariates. Moreover, unlike what was found in the case of research impact, the association between brokerage and innovation is now statistically significant.

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	3.97E-11	4.13E-10	5.66E-20	0.027876
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	0.150211	0.063125	0.065916	0.342303
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	1.330701	0.046708	1.242232	1.425469
Academic system	2.712254	0.858253	1.458747	5.042903
Industry	0.098093	0.034607	0.049129	0.195857
Government	0.146177	0.040164	0.08531	0.250469
Health	0.515374	0.034133	0.452634	0.58681
Research Institute	0.643473	0.087905	0.492319	0.841033
Number of observations = 639 Log pseudolikelihood = -912.65477 Wald $\chi^2(10)$ = 174.24 Log likelihood = -912.65477				

Table S6. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with the efficiency as covariate.

S.5.3 Estimates from mixed-effects regression models with interaction effects

To check for robustness, we further estimated a number of additional three-level random-intercept models. In these models we introduced a level-1 interaction term between (standardised) geographical entropy and the clustering coefficient. Tables S7 and S8 report estimates from such models of research impact and innovation, respectively. Once again, results (and especially the main effects of the interacted covariates) remained qualitatively similar to the ones reported in the main text. In addition, the parameter estimate for the interaction term was found to be negative and statistically significant in both models. That is, the expected negative effects on an organisation's research impact and innovation of an increase by one standard deviation in clustering are amplified as the geographical diversity of an organisation's collaboration network also increases by one standard deviation, after controlling for the other covariates. In other words, these results suggest that the benefits of open structures can be further amplified when an organisation's collaborators are spatially distributed across many and different countries.

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	1.69721	1.125712	0.462549	6.227501
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	2.409696	0.939308	1.122434	5.173253
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	19.66172	0.676594	18.37936	21.03357
Academic system	3.627777	1.421989	1.682647	7.821469
Industry	1.879379	0.539752	1.070414	3.299718
Government	4.666399	1.045263	3.008255	7.238511
Health	8.768805	0.617094	7.639028	10.06567
Research Institute	12.06206	1.66891	9.197057	15.81954
Number of observations = 639 Log pseudolikelihood = -2658.2566 Wald $\chi^2(10)$ = 549.88 Log likelihood = -2658.2566				

Table S7. Maximum-likelihood estimates from three-level random-intercept linear models of research impact with interaction effects.

Random-effects parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Geographical region ($\sqrt{\boldsymbol{\psi}^{(3)}}$)	2.58E-09	2.60E-08	6.54E-18	1.014793
Country ($\sqrt{\boldsymbol{\psi}^{(2)}}$)	0.166526	0.056729	0.085411	0.324676
Level-1 residual by type of organisation ($\sqrt{\boldsymbol{\vartheta}^{(t)}}$)				
Academic	1.308692	0.04497	1.223455	1.399866
Academic system	2.645774	0.840066	1.419992	4.929691
Industry	0.076094	0.025267	0.039693	0.145878
Government	0.118277	0.034155	0.067159	0.208306
Health	0.534018	0.036643	0.46682	0.610891
Research Institute	0.58007	0.080519	0.441902	0.761438
Number of observations = 639 Log pseudolikelihood = -903.79064 Wald $\chi^2(10)$ = 203.47 Log likelihood = -903.79064				

Table S8. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with interaction effects.

References

1. Newman MEJ. *Networks. An Introduction.* Oxford, UK: Oxford University Press; 2010.

2. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*. 2001;64(1):016132.

3. Garas G, Cingolani I, Panzarasa P, et al. Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery. *PLoS One.* 2017;12:e0183332.

4. Opsahl T, Panzarasa P. Clustering in weighted networks. *Social Networks*. 2009;31:155-63.

5. Latora V, Nicosia V, Panzarasa P. Social cohesion, structural holes, and a tale of two measures. *Journal of Statistical Physics*. 2013;151:745-64.

6. Burt RS. *Structural Holes. The Social Structure of Competition*. Cambridge, MA: Harvard University Press; 1992.

7. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. 2010;32:245-51.

8. Savage N. Industry links boost research output. *Nature.* 2017;552:S11-S13.

9. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling. Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC; 2004.

Appendix 3

S.1 Data

Drawing on the Web of Science® (WOS) database, we searched for all publications concerning robotic surgery according to the following search query:

- Topic*: "robotic OR robotic OR robot assisted OR robotic assisted OR robotically assisted OR robot-assisted OR robotic-assisted OR robotically-assisted"*;
- Research Area*: "Surgery"*;
- Document Types*: "Article"*.

The search was performed on the 17th of January 2017, and produced 3,889 publications (peer-reviewed articles) published between July 1988 and January 2017. Figure S31 shows the distribution of the retrieved publications across the years.

Based on these 3,889 articles, a process of manual disambiguation of name variants of organisations was performed. Out of 2,590 name variants, we filtered out 1,700 unique names. To this end, each name variant was searched for using the WOS® "Organizations - Enhanced List" tool with a view to identifying the preferred name associated with the corresponding organisation. If such name was found, then it was used as a unique name. In particular, 816 preferred names were found corresponding to 1,485 name variants. Not all organisations in the WOS® database were included in this list by the WOS® team. For 1,105 name variants we could not find a corresponding preferred name using the above-mentioned tool. A manual search on

Google search was then performed that took into account the associated geographical location (country and city) of the organisation and, when necessary, also the affiliated authors' names. This resulted in the identification of an additional 884 unique names of organisations. Thus, in total we obtained 1,700 unique organisations.

These 1,700 organisations were nested within 61 countries (China and Hong Kong were aggregated into one country), and within 7 geographical regions. These regions were based on the standard international allocation of countries into continents, with the exception of Russia and Middle Eastern countries that were classified as transcontinental countries. In particular, Russia was classified as belonging to Europe given the geographical proximity of its capital to the European borders.

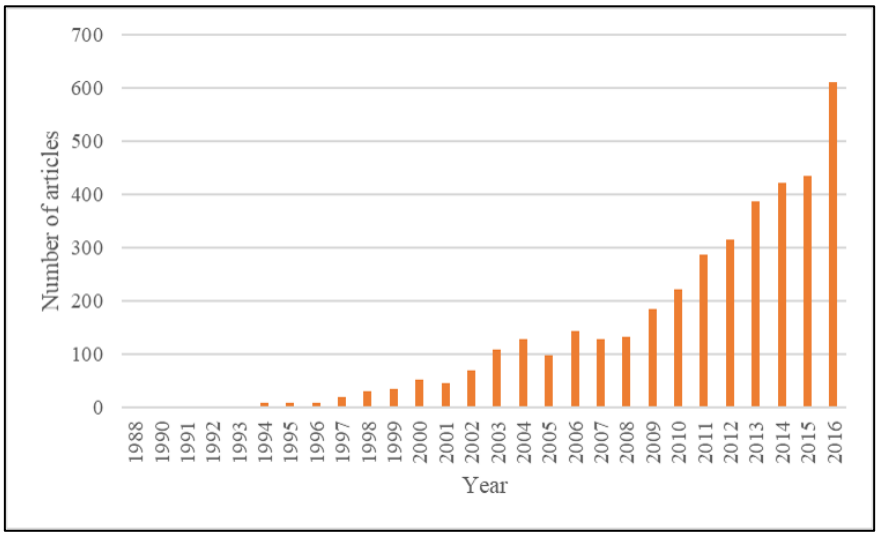


Figure S31. Time distribution of publications

Of the 3,889 articles collected, 23 did not specify the affiliations of the co-authors. Thus, we focused on the remaining 3,866 articles with specified affiliations of co-authors. Moreover, among the 3,866 articles, 3,306 are articles published solely by domestic organisations, of which 2,088 are articles involving only non-collaborating domestic organisations and 1,778 collaborating domestic organisations. The

remaining 560 articles resulted from international collaborations. Table S1 shows the total number of articles, the number of organisations collaborating at least once with international organisations, and the number of domestic organisations of the top 20 countries ranked by number of articles in robotic surgery resulted from international collaborations.

Country	Number of articles with international collaborations	Number of articles	Number of organisations collaborating at least once with international organisations	Number of domestic organisations
USA	350	1846	188	520
Italy	95	280	60	127
Germany	94	359	59	137
France	86	204	60	103
South Korea	75	352	33	58
Canada	73	159	19	33
United Kingdom	67	167	39	68
Japan	59	224	36	112
Austria	39	86	10	20
China-Hong Kong	34	179	31	71
Brazil	27	34	21	26
Belgium	26	61	11	17
Netherlands	22	88	21	32
Switzerland	22	40	14	14
Australia	16	35	12	31
Spain	16	46	16	41
Turkey	14	79	13	49
Singapore	13	27	5	13
Israel	12	16	10	11
Greece	11	21	9	12

Table S1. Total number of articles, number of organisations collaborating at least once with international organisations, and number of domestic organisations of the top 20 countries ranked by number of articles resulting from international collaborations.

The 560 articles that resulted from international collaborations involved from a minimum of 2 countries to a maximum of 19 countries, with a median of 2 countries. The organisations involved in these 560 articles range from a minimum of 2 organisations to a maximum of 43 organisations per article, with a median value of 3 organisations per article. Given an article resulting from international collaborations, a country contributes in proportion to the share of domestic organisations appearing as the affiliations of the co-authors in the byline. Across the 560 articles here considered, a country can contribute to a given article through a minimum of one organisation to a maximum of 21 organisations, and a median of one organisation per article.

S.2 Building the international collaboration network

To construct the collaboration network among countries, first we built the collaboration network among the 1,700 organisations that carried out research on robotic surgery. The nodes of the network are the organisations, and links refer to collaborations between organisations. Specifically, a link exists between any two organisations when one or more scholars affiliated with one organisation co-authored at least one article with one or more scholars affiliated with the other organisation. Out of the 1,700 organisations, 1,543 were connected through at least one collaboration. The largest connected component includes 1,334 organisations, and in total the network comprises 6,000 collaboration links.

To construct the inter-organisational network, we first define $\boldsymbol{A^{org}}$ as the undirected and binary bipartite authorship matrix defining connections between each organisation org and each article a such that element $a_{a,org} = 1$ if and only if

organisation org contributed to article a, and $a_{a,org} = 0$ otherwise.[1] We then define $\boldsymbol{C^{org}}$ as the undirected and weighted projected one-mode collaboration matrix, where each element $c_{h,k}^{org} > 0$ if and only if organisation h collaborated with organisation k. That is, $c_{h,k}^{org} > 0$ when organisation h and organisation k co-authored at least one publication.

The weight of each collaborative tie between any two organisations is measured according to the method proposed by Newman (2001).[2] In particular, the total weight $c_{h,k}^{org}$ of the collaborative tie between two organisations h and k is assumed to be equal to the sum of the strengths $w_{h,k}$ contributed by each of the articles co-authored by the two organisations, where strength $w_{h,k}$ is defined as:

$$w_{h,k} = \frac{1}{n_a - 1},$$

and n_a is the number of organisations appearing in the byline of article a. In other words, the total weight $c_{h,k}^{org}$ of the collaborative tie between organisations h and k is: (i) proportional to the number of articles these organisations co-authored; and (ii) inversely proportional to the number of organisations involved in each article. Note that all articles published by one single organisation were excluded from the above measure as they did not contribute to the inter-organisational network.

Next, to build the collaboration network among countries, we define $\boldsymbol{A^{\gamma}}$ as the undirected and unweighted matrix of connections among all countries, where element $a_{i,j} = 1$ if at least one organisation in country i collaborated with the at least one organisation in the country j, and $a_{i,j} = 1$ otherwise. That is, letting O_i and O_j be, respectively, the sets of organisations in country i and country j, the entry $a_{i,j}$ of

matrix $\boldsymbol{A^{\gamma}}$ is equal to one when there are at least two institutions, $h \in O_i$ and $k \in O_j$, such that at least one scholar affiliated with O_i and one scholar affiliated with O_j co-authored at least one publication.

Finally, to obtain the weights $c_{i,j}^{\gamma}$ of the links connecting countries, we then define $\boldsymbol{C^{\gamma}}$ as the undirected and weighted collaboration matrix, where each element $c_{i,j}^{\boldsymbol{\gamma}} > 0$ if and only if country i collaborated with country j through at least one co-authored publication. More specifically, for any pair of countries i and j, the weight $c_{i,j}^{\boldsymbol{\gamma}}$ of the link connecting i and j is equal to the sum of the weights of all links connecting any organisation $h \in O_i$ in country i to any organisation $k \in O_j$ in country j, such that $c_{h,k}^{org} > 0$. Formally, we have:

$$c_{i,j}^{\gamma} = \sum_{h \in O_i, k \in O_j} c_{h,k}^{org}$$

where, as usual, $O_i$and $O_j$are, respectively, the sets of organisations in country i and country j.

The overall inter-organisational network comprises 6,000 collaborative links connecting 1,543 organisations worldwide. According to the type of collaboration, we can distinguish between the network connecting organisations involved only in domestic collaborations and the network connecting organisations involved in international collaborations. The former comprises 2,674 links and 1,162 organisations, while the latter 3,513 links (of which 1,116 are within national borders and 2,397 across countries) and 772 organisations (of which 513 also collaborated within national borders). The network among countries as well as our network study

draw on the 2,397 international links connecting 772 organisations across different countries.

S.3 Outcome measures

In this Section, we provide details on the dependent measures used in the analysis. These are the innovation index and research impact.

S.3.1 The innovation index

We classified each article in terms of its level of evidence following the procedure proposed by Garas et al. (2017).[3] Each articles was associated with one out of 8 categories corresponding to distinct surgical innovation stages as shown in Table S2.

Surgical innovation stage	Description
1	Randomised controlled trial
2	Non-randomised controlled trial
3	Observational study with controls
4	Observational study without controls
5	First-in-human study
6	Cadaveric study
7	Animal study
8	Description of idea / laboratory evaluation

Table S2. Description of surgical innovation stages with which articles can be associated according to their level of evidence.

We could assign level of evidence to 3,389 articles. Figure S.3 shows the temporal distributions of these 3,389 articles in each category. Among these articles, only 3,376 showed the affiliations of co-authors. Moreover, among the articles with associated level of evidence, 1,808 involved only non-collaborating domestic organisations, 1,077 involved only collaborating domestic organisations, and 491 resulted from international collaborations.

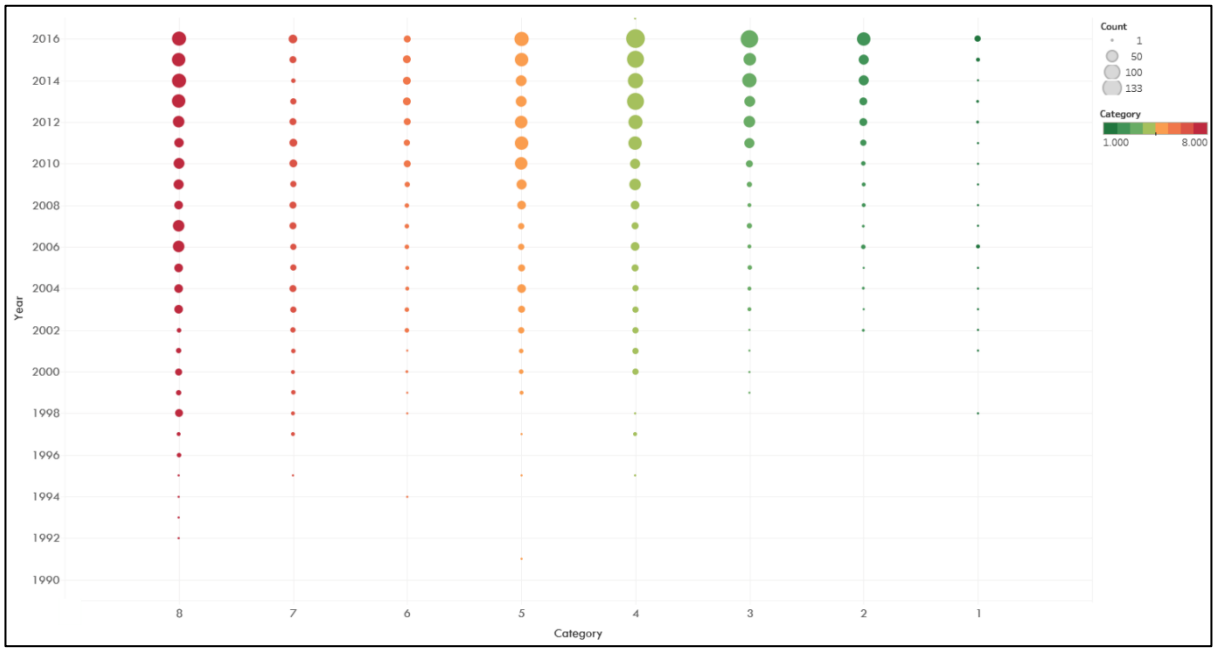


Figure S2. Number of articles per level of evidence over time.

We then computed the *innovation index* following the definition given in Garas *et al.* (2017).[3] For each organisation $org_i \in O_i$ in country i, we computed the innovation index ii_{org_i} as a function of the innovative value of all articles published by authors affiliated with org_i. We then measured the innovation index ii_i of country i as the sum of the innovation indexes of all organisations residing in i. Formally, we have

$$ii_i = \sum_{org_i \in O_i} ii_{org_i} = \sum_{org_i \in O_i} \left[\frac{\sum_{c \in S} \frac{1}{c} \times \frac{p_c^{org_i}}{P_c}}{\max_{org \in O} \left(\sum_{c \in S} \frac{1}{c} \times \frac{p_c^{org}}{P_c} \right)} \times 100 \right],$$

where $c \in S = [1, \ldots, 8]$ is an integer value labeling the ordinal category associated with the implementation stage according to level of evidence, $p_c^{org_i}$ is the count of publications of organisation org_i in country i associated with category c, and $P_c = \sum_{org \in O} p_c^{org}$ is the total number of publications associated with category c across all organisations in the data set (i.e., in the set O). Figure S4 shows the ranking of countries according to the average of innovation indices of all the organisations

within their borders. The figure also highlights the most innovative organisations within and between countries.

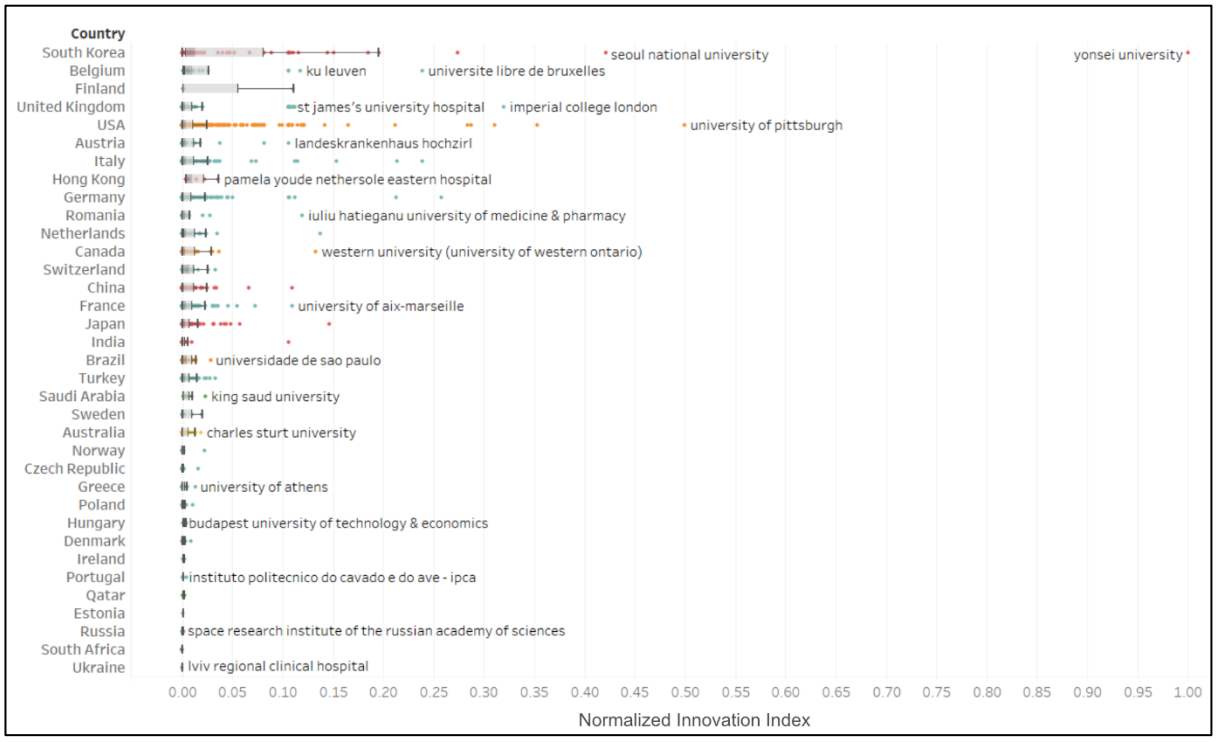


Figure S3. Ranking of countries and organisations according to normalised innovation index. Countries have been ranked according to the sum of organisations' innovation index. For each country, the boxplot of organisational innovation indices are depicted. Grey shaded areas correspond to distance between lower and upper quartiles (interquartile range), darker shade area denotes distance between lower quartile and median. Whiskers are showed to display all points within 1.5 times the interquartile range. To obtain the normalised innovation index, we divided each organisation's innovation index by the maximum value observed in the dataset.

S.3.2 Research impact

For each organisation, we measured its academic performance using the sum snc_{org} of the normalised citations received by all the articles published by the organisation across the years:

$$snc_{org} = \sum_{y} \sum_{a_{org,y}} nc_{a_{org,y}},$$

where $nc_{a_{org,y}}$ refers to the normalised citations received by article *a* published by (authors affiliated with) organisation *org* in year *y*. To obtain normalised citations, the citation count for each publication in a given year was divided by the average number of citations obtained by all articles published in the same year:

$$nc_{a_{org,y}} = \frac{cit_{a_{org,y}}}{< cit_{a,y} >},$$

where the average $< cit_{a,y} >$ was computed across all articles published by all organisations in year *y*. Finally, we measured the research impact snc_i of country *i* as the sum of the normalised citations received by each organisation $org_i \in O_i$ in country *i*. Formally, we have

$$snc_i = \sum_{org_i \in O_i} \sum_{y} \sum_{a_{org_i}} nc_{a_{org_i,y}},$$

where $nc_{a_{org_i,y}}$ refers to the normalised citations received by article *a* published by (authors affiliated with) organisation org_i in country *i* and in year *y*.

S.4 Breaking down research impact and innovation into domestic and foreign components

The overall academic performance of a country can be broken down into sub-components according to the nature of the institutional authorship. To this end, we distinguished among: (i) the fraction of the overall academic performance attributable to articles authored by a singular domestic institution (no collaboration); (ii) the fraction of the overall academic performance attributable to articles resulting from collaborations among uniquely domestic organisations (i.e., domestic collaboration); and (iii) the fraction of the overall academic performance attributable to articles

resulting from collaborations among domestic and foreign organisations (i.e., international collaboration). Moreover, we further broke down the fraction of performance resulting from international collaborations into a component attributed to domestic organisations and a component attributed to foreign organisations (see Figure S4).

S.4.1 Breaking down a country's sum of normalised citations (*snc*) into its domestic and foreign components

We denote by snc_i the sum of normalised citations obtained by country i, and by a_i any article in which the sole author or at least one of the co-authors is affiliated with an organisation in country i. We want to break down snc_i into its domestic and foreign components in order to understand how much of the overall academic success of a country can be attributed to international collaborations as opposed to solely domestic endeavours. For each country, we can then distinguish between the portion of the country's overall academic success that originates from domestic organisations collaborating with foreign ones ($intDSNC_i$), the portion of success attributable to the foreign organisations collaborating with domestic ones ($FSNC_i$), and finally the portion of success related to domestic organisations not collaborating internationally ($exDSNC_i$). In turn, the foreign component can be broken down into regional subcomponents ($FSNC_{i,r}$) to characterise the geographical dispersion of international collaboration and the specific regional contribution to a country's success. Formally, we have:

$$\mathrm{snc}_i = \sum_{a_i=1}^{A_i} nc^{a_i}\left(\frac{N_i^{a_i}}{N^{a_i}} + \sum_{j\neq i}\frac{N_j^{a_i}}{N^{a_i}}\right) = \sum_{a_i=1}^{A_i} nc^{a_i}\left(\frac{N_i^{a_i}}{N^{a_i}} + \sum_{j\neq i} f_{i,j}^{a_i}\right) = \sum_{a_i=1}^{A_i} nc^{a_i}\left(d_i^{a_i} + f_i^{a_i}\right)$$
$$= \sum_{a_i=1}^{A_i} nc^{a_i} d_i^{a_i} + \sum_{a_i=1}^{A_i} nc^{a_i} f_i^{a_i} = DSNC_i + FSNC_i,$$

where nc^{a_i} refers to normalised citations of article a_i, A_i is the total number of articles published by country i, $N_i^{a_i}$ is the number of unique domestic organisations with which co-authors of article a_i, are affiliated, $N_j^{a_i}$ is the number of unique organisations from country j in article a_i, and N^{a_i} is the total number of unique organisations in article a_i.

If article a_i to which country i contributed shows exclusively organisations from country i, then the domestic component $d_i^{a_i} = 1$ and the foreign component $f_i^{a_i} = 0$. If, instead, article a_i to which country i contributed shows organisations both from country i and from other countries, then the domestic component would be $0 < d_i^{a_i} < 1$, and the foreign component would be $0 < f_i^{a_i} < 1$. In both cases, for each article a_i the condition would hold that $d_i^{a_i} + f_i^{a_i} = 1$.

Regarding the values taken by $d_i^{a_i}$, let us distinguish between the case of articles in which there are only domestic organisations and the case of articles in which both domestic and foreign organisations appear. Formally,

$$DSNC_i = exDSNC_i + intDSNC_i,$$

where

$$exDSNC_i = \sum_{a_i} nc^{a_i} d_i^{a_i} \text{ for each } a_i \text{ such that } d_i^{a_i} = 1;$$
$$intDSNC_i = \sum_{a_i} nc^{a_i} d_i^{a_i} \text{ for each } a_i \text{ such that } d_i^{a_i} < 1.$$

Thus, $exDSNC_i$ refers to the contribution of the normalised citations obtained by country i attributable to the articles in which only domestic organisations appear, and

$intDSNC_i$ is the contribution of citations attributable to the domestic organisations across all articles in which domestic organisations co-appear with foreign ones. Moreover, the foreign component $f_i^{a_i}$ can be further broken down into specific contributions attributable to the various foreign regions (e.g., continents) with which the focal country collaborates. Each foreign country j can be attributed to a region $= 1, 2, \dots, R$. We can then sum up the foreign contributions according to the region each country belongs. In this way, we can break down the overall foreign component $FSNC_i$ into regional sub-components. Formally,

$$FSNC_i = \sum_{a_i=1}^{A_i} nc^{a_i} f_i^{a_i} = \sum_{a_i=1}^{A_i} \sum_{j \neq i} nc^{a_i} f_{i,j}^{a_i} = \sum_{a_i=1}^{A_i} \sum_{j \neq i} nc^{a_i} \frac{N_j^{a_i}}{N^{a_i}} = \sum_r \sum_{a_i=1}^{A_i} \sum_{j \neq i} nc^{a_i} \frac{N_{j,r}^{a_i}}{N^{a_i}} =$$
$$= \sum_r \sum_{a_i=1}^{A_i} \sum_{j \neq i} nc^{a_i} f_{i,j,r}^{a_i} = \sum_r FSNC_{i,r}.$$

Thus, country i's sum of normalised citations snc_i can be broken down into its domestic and foreign components as follows

$$\text{snc}_i = DSNC_i + FSNC_i = exDSNC_i + intDSNC_i + \sum_r FSNC_{i,r}.$$

S.4.2 Breaking down a country's innovation index (*ii*) into its domestic and foreign components

We denote by ii_i the innovation index of country i, and by a_i^c any article in category c in which the sole author or at least one co-author is affiliated with an organisation in country i. Once again, we want to break down this summary statistics into domestic and foreign components in order to understand the extent to which the overall innovation index of a country can be attributed to international collaborations as opposed to domestic ones. For each country, we can then distinguish between the portion of a country's overall innovation index that originates from domestic

organisations collaborating with foreign ones ($intDII_i$), the portion that relates to foreign organisations collaborating with domestic ones (FII_i), and finally the portion attributable to articles in which only domestic organisations appear ($exDII_i$). In turn, the foreign component of innovation can be further broken down into regional subcomponents ($FII_{i,r}$) to characterise the geographical dispersion of international collaboration and the specific regional contribution to a country's innovativeness. Formally, we have

$$ii_i = \sum_{c=1}^{C} \frac{1}{c} \frac{A_i^c}{A^c},$$

where A^c is the overall number of articles that belong to category c across all countries, A_i^c is the number of articles to which country i contributed and that belong to category c, and C is the total number of categories used to measure innovation. We have

$$A_i^c = \sum_{a_i^c=1}^{A_i^c} \left(\frac{N_i^{a_i^c}}{N^{a_i^c}} + \sum_{j \neq i} \frac{N_j^{a_i^c}}{N^{a_i^c}} \right) = \sum_{a_i^c=1}^{A_i^c} \left(\frac{N_i^{a_i^c}}{N^{a_i^c}} + \sum_{j \neq i} f_{i,j}^{a_i^c} \right) = \sum_{a_i^c=1}^{A_i^c} \left(d_i^{a_i^c} + f_i^{a_i^c} \right)$$
$$= \sum_{a_i^c=1}^{A_i^c} d_i^{a_i^c} + \sum_{a_i^c=1}^{A_i^c} f_i^{a_i^c} = D_i^c + F_i^c,$$

where a_i^c is an article in category c and to which country i contributed, $N^{a_i^c}$ is the total number of unique organisations appearing in article a_i^c, $N_i^{a_i^c}$ is the number of unique domestic organisations appearing in article a_i^c, and $N_j^{a_i^c}$ is the number of unique foreign organisations appearing in article a_i^c. Thus, D_i^c is the total fraction of articles in category c contributed by country i that is attributable to domestic organisations, and F_i^c is the fraction attributable to foreign organisations. Notice that, for each article a_i^c, the condition holds that

$$\left(\frac{N_i^{a_i^c}}{N^{a_i^c}} + \sum_{j \neq i} \frac{N_j^{a_i^c}}{N^{a_i^c}}\right) = \left(d_i^{a_i^c} + f_i^{a_i^c}\right) = 1.$$

Thus, if article a_i^c belonging to category c shows exclusively organisations from country i, then the domestic component $d_i^{a_i^c} = 1$ and the foreign component $f_i^{a_i^c} = 0$. If, by contrast, a_i^c belonging to category c shows organisations both from country i and from foreign countries, then the domestic component would be $0 < d_i^{a_i^c} < 1$ and the foreign component $0 < f_i^{a_i^c} < 1$.

Regarding the domestic component D_i^c, we can now distinguish between the portion that derives from articles that are exclusively domestic and the one deriving from the domestic component of the articles in which both domestic and foreign organisations appear. We have

$$D_i^c = D_{i,ex}^c + D_{i,int}^c,$$

where

$$D_{i,ex}^c = \sum_{a_i^c} d_i^{a_i^c}, \text{ for each } a_i \text{ such that } d_i^{a_i^c} = 1;$$
$$D_{i,int}^c = \sum_{a_i^c} d_i^{a_i^c} \text{ for each } a_i \text{ such that } d_i^{a_i^c} < 1.$$

Thus, $D_{i,ex}^c$ is the sum of the contributions of domestic organisations across the subset of articles in which they are the sole contributors (i.e., the total number of articles in category c in which only domestic organisations appear), and $D_{i,int}^c$ is the sum of the contributions of domestic organisations across all articles in which they appear jointly with foreign ones.

In turn, the foreign component $f_i^{a_i^c}$ can be further broken down into specific contributions related to the regions in which the foreign countries are located. To this

end, each foreign country j is attributed to a region $= 1, 2, \dots, R$. We can then sum up the foreign contributions according to the region each country belongs to. We have

$$f_i^{a_i^c} = \sum_{j \neq i} f_{i,j}^{a_i^c} = \sum_{j \neq i} \frac{N_j^{a_i^c}}{N^{a_i^c}} = \sum_r \sum_{j \neq i} \frac{N_{j,r}^{a_i^c}}{N^{a_i^c}} = \sum_r \sum_{j \neq i} f_{i,j,r}^{a_i^c}.$$

In this way, we can break down the overall foreign component F_i^c into regional sub-components:

$$F_i^c = \sum_{a_i^c=1}^{A_i^c} f_i^{a_i^c} = \sum_{a_i^c=1}^{A_i^c} \sum_r \sum_{j \neq i} f_{j,r}^{a_i^c} = \sum_r \sum_{a_i^c=1}^{A_i^c} \sum_{j \neq i} f_{j,r}^{a_i^c} = \sum_r F_{i,r}^c.$$

Thus, a country i's innovation index ii_i can be broken down into its domestic and foreign components as follows

$$ii_i = \sum_{c=1}^{C} \frac{1}{c} \frac{A_i^c}{A^c} = \sum_{c=1}^{C} \frac{1}{c} \frac{D_{i,ex}^c + D_{i,int}^c + F_i^c}{A^c} = \sum_{c=1}^{C} \frac{1}{c} \frac{D_{i,ex}^c + D_{i,int}^c + \sum_r F_{i,r}^c}{A^c}$$
$$= \sum_{c=1}^{C} \frac{1}{c} \frac{D_{i,ex}^c}{A^c} + \sum_{c=1}^{8} \frac{1}{c} \frac{\sum_r D_{i,int}^c}{A^c} + \sum_{c=1}^{C} \frac{1}{c} \frac{\sum_r F_{i,r}^c}{A^c} = exDII_i + intDII_i + FII_i,$$

where $exDII_i + intDII_i$ is the total portion of country i's innovation that is attributable to domestic organisations, and FII_i is the foreign component. In turn, FII_i can be further broken down into its regional components:

$$FII_i = \sum_{c=1}^{C} \frac{1}{c} \frac{\sum_r F_{i,r}^c}{A^c} = \sum_{c=1}^{C} \frac{1}{cA^c} \sum_r F_{i,r}^c = \sum_r \sum_{c=1}^{C} \frac{1}{cA^c} F_{i,r}^c = \sum_r FII_{i,r}.$$

Finally, we have:

$$ii_i = \sum_{c=1}^{C} \frac{1}{c} \frac{A_i^c}{A^c} = exDII_i + intDII_i + \sum_r FII_{i,r}.$$

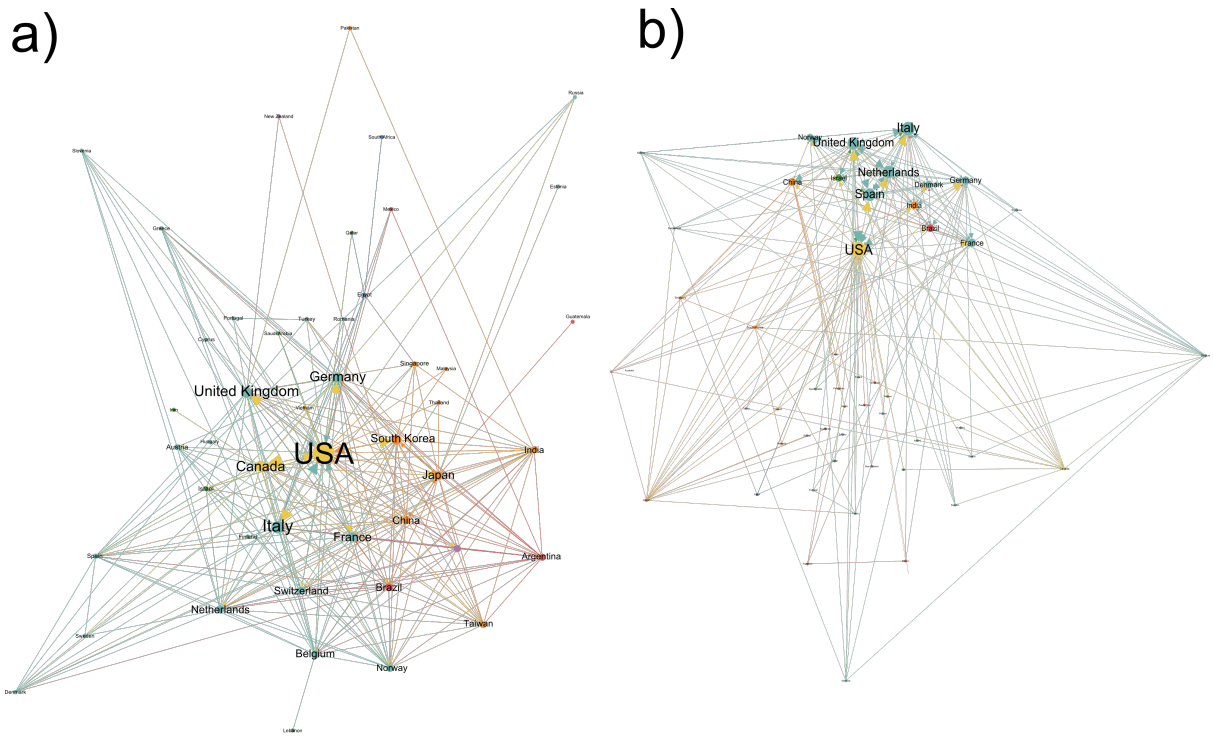


Figure S4. Network of foreign contributions to countries' research impact (a) and innovation (b). The width of each link is proportional to the pointing country's contribution to the success of the pointed one. The size of each node is proportional to the sum of the node's weights of incoming links (i.e., the in-strength).

S.5 Formal definitions of covariates

S.5.1 Network degree

Given the set of countries Γ, we define the degree k_i of country i as the number of links connecting the country to other countries in Γ. Formally, we have

$$k_i = \sum_{j \in \Gamma} a_{i,j},$$

where $a_{ij} = 1$ if there is at least one organisation in country i that collaborated with the at least one organisation in country j ($i \neq j$), and $a_{ij} = 0$ otherwise. Thus, $a_{ij} = 1$ implies that there are at least two scholars, one affiliated with an organisation in country i and the other in country j, that co-authored at least one publication, thus

generating a collaboration link between the corresponding countries.[90]

S.5.2 The local clustering coefficient

Network analysis provides a set of measures for quantifying a number of structural properties at different levels of a system.[1] For example, measures can be calculated at a node's level to detect its topological position (e.g., broker between otherwise disconnected partners) in the ego-centred network (i.e., the network including connections between the node and its partners as well as connections between these partners), or in the whole network (e.g., how close the node is to all other others in the network). By constructing and examining each node's ego-centred network, we aim to detect the extent to which a node's neighbours are also connected with each other, and in so doing we engage with current debates on social capital and the relative benefits of closed versus open social structures.[4]

Using the weighted network as defined above, we computed the local clustering coefficient of each country, which measures the extent to which the country's ego-centred network is a closed collaborative structure.[1,5] Formally, the (unweighted) local clustering coefficient cc_i of country i in a binary network is defined as the ratio between the number of actual triangles t_i containing country i and its k_i neighbours, and the maximum possible number of such triangles, $t_i^{max} = k_i(k_i - 1)/2$:

$$cc_i = \begin{cases} \frac{2 \times t_i}{k_i(k_i - 1)} & \text{for } k_i \geq 2 \\ 0 & \text{for } k_i = 0,1 \end{cases} .$$

In our study, we computed the clustering coefficient only for countries with $k_i \neq 0, 1$.

S.5.3 Efficiency

To measures a country's opportunities of brokerage, we also measured the country's efficiency, namely the degree to which the country's ego-centred network is non-redundant.[4] Efficiency thus captures the extent to which a country's ego-centred network is an open structure. To measure efficiency, we first computed effective size, a measure originally introduced by Burt (1992)[5] to capture the extent to which each of the first neighbours of a node is redundant with respect to the other neighbours. Following Latora *et al*. (2013),[4] a country *i*'s effective size, es_i, can be defined as

$$es_i = k_i - (k_i - 1)cc_i,$$

where k_i is the degree of country i (i.e., the number of countries with which i collaborated), and cc_i is the local binary local clustering coefficient (i.e., the coefficient presented in S.5.2 computed using the unweighted network).

The efficiency eff_i of country i (i.e., the normalised effective size) can now be computed as the ratio between the country's effective size and degree:

$$eff_i = \begin{cases} \frac{es_i}{k_i} = 1 - \frac{k_i - 1}{k_i} cc_i & k_i > 0 \\ 0 & k_i = 0 \end{cases}$$

Notice that, as with the local clustering coefficient (Section S.3.3), efficiency was computed only for countries with $k_i \neq 0, 1$.

S.5.4 Geographical entropy

In addition to the network-based measures described above, for each country we measured the diversity of the geographical locations of collaborators. To this end, let us define Γ_i as the set of countries in which the collaborating organisations of a given

country i are located, i.e., $\Gamma_i = \{\gamma_1, \gamma_2, \ldots, \gamma_f\}$. Formally, the geographical entropy of country i can be defined as follows:

$$\varepsilon_i = -\sum_{\gamma_i \epsilon \Gamma_i} p_{i,\gamma_i} log(p_{i,\gamma_i}),$$

where p_{i,γ_i} is the fraction of country i's collaborative effort towards organisations located in country γ_i (i.e., the ratio between the sum c_{i,γ_i} of the weights of links between i and all organisations in country γ_i and the sum $\sum_{\gamma_i \in \Gamma_i} c_{i,\gamma_i}$ of the weights of links between i and all countries in Γ_i). Notice that each weight c_{i,γ_i} is measured once again according to the method proposed by Newman (2001) (see Section S.2).

Given the set Γ of countries in the data set and the norm $\|\Gamma\|$ of such set, a country's geographical entropy can range from zero (when all international collaborators are located in the same country) to $log(\|\Gamma\| - 1)$ (when the international collaborators are uniformly distributed across $\|\Gamma\| - 1$ countries, i.e., all countries except the focal one). Similarly, given the set Γ_i of countries with which country i collaborates, entropy is maximised when the country equally distributes its collaborative efforts across all countries in Γ_i. That is, for each country $\gamma_1, \gamma_2, \ldots, \gamma_f \in \Gamma_i$, ε_i is maximised when each fraction of collaborative effort is equal to $p_{i,\gamma_i} = \frac{1}{\|\Gamma_i\|}$, where $\|\Gamma_i\|$ is the norm of the set Γ_i.

Thus, geographical entropy enables us to assess countries beyond the mere count of international collaborators, and to distinguish between countries that, while having the same number of international collaborators, differ in terms of their geographical dispersion. This is important, as some countries may have a large number of

international collaborators, yet all confined within the same country, whereas others may have few collaborators that nonetheless are widely geographically dispersed.

S.5.5 Control variables

Other country-level characteristics may influence performance. In this study we controlled for the following two additional variables. First, for each country we measured the average contribution of international collaborators to a scientific publication. Given country *i* and article a_i, let $N_i^{a_i}$ be the number of unique domestic organisations in country i with which co-authors of a_i are affiliated, and N^{a_i} be the total number of unique organisations contributing to article a_i. We define $p_{a_i}^{org_i} = \frac{N_i^{a_i}}{N^{a_i}}$ as the fraction of organisations contributing to article a_i that reside in country *i*. The average contribution of international organisations to country *i*'s publication can thus be defined as

$$\langle p_{a_i}^{int} \rangle = 1 - \frac{\sum_{a_i} \frac{N_i^{a_i}}{N^{a_i}}}{A_i},$$

where A_i is the total number of articles published by country *i*. Thus, $\langle a_i^{int} \rangle$ captures the degree to which international collaboration contributed on average to an article published by country *i*.

Second, for each country we constructed a measure of volume by computing the average number of publications per domestic contributing organisation. Formally, we have

$$\langle A_{org_i} \rangle = \frac{A_i}{N_i},$$

where N_i is the total number of unique organizations in country i. Finally, we controlled for each country's geographical regions, by estimating fixed effects for the corresponding categories (see section S.1).

S.6 Regression analysis

In the main manuscript, we reported estimates from two cross-section models of research impact and innovation, measured as usual by normalized citation count and innovation index, respectively. In both models, the covariates of brokerage and geographical dispersion of collaborators were measured in the usual way as efficiency and geographical entropy, computed on the network between countries and standardised for the sake of interpretation. We also estimated parameters for two control variables (i.e., relevance of international contribution and average number of articles per domestic institution) and fixed effects for geographical regions.

In both models, we treated both efficiency and geographical entropy as endogenous covariates. The intuitive justification is that some unobserved covariates (e.g., each country's accumulated knowledge or ability to collaborate) might account for variations both in the country's ability to broker between collaborators, to amplify geographical diversity of collaborators, and secure success (i.e., efficiency and entropy might be correlated with the level-one country-specific residuals). To correct for such level-one endogeneity of efficiency and entropy, in both models we used the following (excluded) instrumental country-level variables: (i) number of unique domestic organisations collaborating at least once with international organisations; (ii) number of unique domestic organisations collaborating at least once with one another. We also squared each of these variables and created an interaction between them. In

total, we thus used five excluded instrumental variables. To avoid problems of multi-collinearity arising from high correlation between variables, before squaring and multiplying variables (i) and (ii) to estimate interaction effects, we centred them around their means (i.e., we subtracted the mean from the variable).

Table S3 reports the means, standard deviations, minimum and maximum values of, and zero-order correlations between, all (non-centred) variables used in the regression models.

Variable	Mean	St. Dev.	Min	Max	1	2	3	4	5
1. Research impact	86·09 9·52	273·46 24·72	0 0	2034·80 173·53	0·94				
2. Innovation index	0·31 1·64	0·21 1·07	0 0·33	0·8 6·07	-0·22 0·47	-0·23 0·41	-0·26		
3. Rel. int. coll.	0·53	0·302	0·09	1·00	-0·01	-0·06	0·18	-0·30	
4. Articles per inst.	1·14	0·89	0	2·79	0·42	0·50	-0·16	0·55	-0·72
5. Efficiency									
6. Geog. entropy									

Table S3. Means, standard deviations, minimum and maximum values, and zero-order correlations between all (non-centred) variables in the regression models.

S.6.1 Results

Tables S4 and S5 report the estimated coefficients and robust standard errors for the two models. Both models are efficient generalised methods of moments (GMM) instrumental-variables two-step estimators that are robust to heteroskedasticity.

Variables	**Coeff.**	**Robust st. error**	**z**	**p>|z|**	**95% Conf. interval**	
Rel. int. coll.	68·272	158·432	0·43	0·667	-242·248	378·793
Articles per inst.	33·668	48·883	0·69	0·491	-62·141	129·478
Efficiency	211·320	63·625	3·32	0·001	86·617	336·022
Geog. entropy	4.39	55.760	244·527	0·000	135·238	353.815
Africa	-75·966	237·91	-0·32	0·749	-542·261	390·329
Asia	-10·371	88·571	-0·12	0·907	-183·967	163·225
Central and South America	-126·965	104·473	-1·22	0·224	-331·729	77·798
Europe	-139·831	82·295	-1·70	0·089	-301·127	21·465
Middle East	-160·874	122·330	-1·32	0·188	-400·636	78·888
North America	524·986	491·652	1·07	0·286	-438·635	1488·606
Constant	74·600	106·865	0·70	0·485	-134·852	284·051
No. Observations: 56 $F(10, 45) = 3·26$ Centred $R^2 = 0·572$ Uncentred $R^2 = 0·614$ Root MSE = 184·4						

Table S4. Instrumental-variables models of countries' research impact. The models are efficient GMM instrumental-variables estimators with standard errors robust to heteroskedasticity. Reported values of the square root of the mean squared error (MSE) refer to the estimated standard deviation $\hat{\sigma}_e$ of the idiosyncratic disturbance.

Variables	Coeff.	Robust st. error	z	p>|z|	95% Conf. interval	
Rel. int. coll.	8·109	11·384	0·71	0·476	-14·203	30·421
Articles per inst.	-2·363	4·174	-0·57	0·571	-10·545	5·819
Efficiency	18·819	6·108	3·08	0·002	6·848	30·791
Geog. entropy	30·850	4·641	6·65	0·000	21·753	39·948
Africa	-2·048	20·654	-0·10	0·921	-42·530	38·433
Asia	9·151	9·557	0·96	0·338	-9·581	27·883
Central and South America	-2·890	11·528	-0·25	0·802	-25·483	19·704
Europe	-1·968	9·771	-0·20	0·840	-21·119	17·183
Middle East	-3·022	13·736	-0·22	0·826	-29·944	23·899
North America	47·184	43·343	1·09	0·276	-37·767	132·135
Constant	7·116	10·899	0·65	0·514	-14·246	28·479
No. Observations: 56 $F(10, 45) = 7·17$ Centred $R^2 = 0·519$ Uncentred $R^2 = 0·587$ Root MSE = 17·63						

Table S5. Instrumental-variables models of countries' innovation. The models are efficient GMM instrumental-variables estimators with standard errors robust to heteroskedasticity. Reported values of the square root of the mean squared error (MSE) refer to the estimated standard deviation $\hat{\sigma}_e$ of the idiosyncratic disturbance.

S.6.2 Specification tests

Table S6 reports findings from a number of specification tests concerning regressor endogeneity, underidendification, overidentifying restrictions, weak identification, and instrument relevance. We also estimated alternative models with better finite-sample properties and more robust to the presence of weak instruments than the instrumental-variables and two-step efficient GMM estimators.

S.6.2.1 Tests of regressor endogeneity

C (GMM distance) tests for regressor endogeneity in both models lead to strong rejection of the null hypothesis that regressors are exogenous (p-value < 0·01 in both models). We also performed the robust Durbin-Wu-Hausman tests of joint endogeneity of regressors, and produced results that reject exogeneity of the covariates (p-value < 0·05). These tests are robust to

S.6.2.2 Tests of underidentification, overidentifying restrictions, weak identification, and instrument relevance

We tested whether our endogenous regressors are not identified by the excluded instruments after partialling-out the included instruments. This test is equivalent to testing whether the matrix of reduced form coefficients for the endogenous regressors is full rank vs. less than full rank. We calculated the LM version of the Kleibergen-Paap rk statistic, which has been proposed as a generalisation of the Anderson canonical correlation rank statistic to the non-i.i.d. case.[6] Across both models the test strongly rejects the null hypothesis of underidentification (p-value < 0·01).

Instruments are valid if they are (i) exogenous and (ii) relevant. To test whether the excluded instruments are exogenous, we computed the Hansen's *J* test of overidentifying restrictions. The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. Findings from the Hansen's *J* test do not reject the null hypothesis, and thus do not cast doubt on the validity of the instruments.

Relevance means that, after controlling for the remaining exogenous regressors in the structural equation, the excluded instruments can account for significant variation in the endogenous regressors. Instruments that are only marginally relevant are typically referred to as weak instruments.[7] Findings (not reported here) from the *F*-tests for the joint significance of the excluded instruments in the first-stage regressions of each of the endogenous regressors reject the joint null hypothesis of zero coefficients, yielding values above 10 (i.e., the "rule of thumb" suggested by Staiger and Stock for rejecting the hypothesis of weak instruments).[8] Similar findings (not reported here) were obtained based on the Angrist-Pischke multivariate *F*-test of excluded instruments.

Moreover, values (not reported here) of the Shea's partial R^2 (and of its adjusted version) for all models also do not seem to flag a problem of weak excluded instruments. We also obtained the Kleibergen-Paap rk Wald *F* statistics (robust to heteroskedasticity of standard errors), and across both models we calculated the Cragg-Donald Wald *F* statistic. Finally, across both models, the Anderson-Rubin Wald test[9] and the Stock-Wright Lagrange multiplier *S* test[7] (both robust to presence of weak identification, as well as to heteroskedasticity of standard errors) also reject their null hypothesis that the coefficients of the endogenous regressors in the main equation are jointly equal to zero (see Table S6).

Moreover, to circumvent potential biases originating from weak instruments and the small sample, we fitted the GMM "continuously updated" estimator (CUE) (which in turn extends the limited-information maximum-likelihood estimator to account for heteroskedasticity of disturbances).[10] Recent research has indeed suggested that both

estimators have a better finite-sample performance than instrumental-variables and GMM estimators when instruments are weak.[11] Estimates (not reported here) from the CUE models are nearly identical to those reported in Table S5.

Tests	**Model (1)** **Research impact**	**Model (2)** **Innovation**
C (GMM distance) test	$\chi^2_2 = 15 \cdot 419$***	$\chi^2_2 = 11 \cdot 921$***
Robust Durbin-Wu-Hausman test	$F(2, 43) = 4 \cdot 214$*	$F(2, 43) = 4 \cdot 879$*
Kleibergen-Paap rk *LM* statistic	$\chi^2_4 = 16 \cdot 887$**	$\chi^2_4 = 16 \cdot 887$**
Hansen *J* statistic	$\chi^2_3 = 3 \cdot 366$ (*p*-value=0·220)	$\chi^2_3 = 3 \cdot 805$ (*p*-value= 0·283)
Kleibergen-Paap rk Wald *F* statistic	7·840	7·840
Cragg-Donald Wald *F* statistic (n. r.)	2·169	2·169
Anderson-Rubin Wald test	$F(5, 42) = 3352 \cdot 58$***	$F(5, 42) = 1089 \cdot 73$ ***
Stock-Wright *LM S* test	$\chi^2_5 = 12 \cdot 80$*	$\chi^2_5 = 13 \cdot 87$*

Table S6. Specification tests. The first two statistics are tests of endogeneity of regressors. Both tests strongly reject the null hypothesis that efficiency and geographical entropy are exogenous covariates. The remaining six statistics are tests of underidentification, overidentifying restrictions, weak identification, and instrument relevance. The tests strongly reject the null hypothesis that the models are underidentified and that the excluded instruments are weak. The tests also do not reject the joint null hypothesis that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. Note: n. r. = test is non-robust to heteroskedasticity; * p < 0·05, ** p < 0·01, *** p < 0·001.

S.7 Simulation modelling

We carried out a number of simulations to shed light on how the UK's research impact and innovation index might be affected as a result of a progressive substitution of the actual output produced in collaboration with other EU countries.

The data set comprises 346 articles to which at least one of the EU27 countries contributed. The EU27 members appearing in our data set are: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

In total, the United Kingdom has contributed to 67 articles (52 associated with level of evidence), of which: (i) 32 articles (25 with level of evidence) in collaboration with 13 of the EU27 countries (i.e., Belgium, Cyprus, Denmark, France, Germany, Greece, Ireland, Italy, Netherlands, Poland, Portugal, Slovenia and Spain); (ii) 29 articles (24 with level of evidence) in collaboration with USA; 14 articles (10 with level of evidence) in collaboration with Asian countries (i.e., China (incl. Hong Kong), India, Japan, Pakistan, Singapore, South Korea and Taiwan); and 20 articles (12 with level of evidence) in collaboration with "other" 11 Rest-of-the World (RoW) countries (i.e., Argentina, Australia, Bolivia, Brazil, Canada, Israel, New Zealand, Norway, Switzerland, Turkey, and Ukraine).

Of the 32 articles (25 with level of evidence) in collaboration with 13 (12 if considering only articles with level of evidence) of the EU27 members, 20 articles (17 with level of evidence) resulted from the participation of only EU27 countries. Collaborations involving only the UK and other non-EU27 countries (i.e., USA, Asian countries, and "RoW" countries) resulted, respectively, in 21, 9, and 10 articles (and 19, 7 and 6 articles with level of evidence). Finally, collaborations involving the UK, non-EU27 as well as EU27 countries resulted, respectively, in 8, 5, and 10 articles (and 6, 3 and 5 articles with level of evidence).

In order to assess the relevance of the EU27 countries to the UK's scientific success, we ranked all articles published by the UK in collaboration solely with any of the EU27 countries by normalised citations and innovation index. Starting with the worst performing article, we progressively replaced various percentages of such articles with articles corresponding to the 25th percentile, the median, or the maximum value of the distribution of normalised citations or level of evidence of all articles resulted from collaborations with non-EU27 countries. Moreover, to select the articles replacing those resulting from collaborations with EU27 members, we further distinguished among articles where at least one of the co-authors was affiliated with an organisation residing in: (i) the USA; (ii) one of the seven Asian countries; and (iii) one of the remaining "RoW" countries.

References

1. Newman MEJ. *Networks. An Introduction.* Oxford, UK: Oxford University Press; 2010.

2. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*. 2001;**64**(1):016132.

3. Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery. *PLoS One* 2017;**12**(8):e0183332.

4. Latora V, Nicosia V, Panzarasa P. Social cohesion, structural holes, and a tale of two measures. *Journal of Statistical Physics* 2013;**151**(3–4):745–64.

5. Burt RS. *Structural Holes. The Social Structure of Competition*. Cambridge, MA: Harvard University Press; 1992.

6. Kleibergen F, Paap R. Generalized reduced rank tests using the singular-value decomposition. *J Econom* 2006;**127**:97–126.

7. Stock JH, Wright JH. GMM with weak identification. *Econometrica* 2000;**68**:1055–1096.

8. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997;**65**:557–586.

9. Anderson T-W, Rubin H. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann Math Statist* 1949;**20**:46–63.

10. Hansen L, Heaton J, Yaron A. Finite sample properties of some alternative GMM estimators. *J Bus Econ Stat* 1996;**14**:262–280.

11. Hahn J, Hausman J, Kuersteiner G. Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econom J* 2004;**7**:272–306.

Appendix 4

1. Search strategy

Our study draws on the Web of Science® (WOS) All Databases Search Engine (Clarivate Analytics, Philadelphia, PA). A search has been performed the 7th December 2017 and a collection of 2,595 articles have been retrieved filtering out publications that have as Topic "aortic valve stenosis" and "aortic valve regurgitation". We then refined the search by a subset of categories Web of Science® categories to which an article can be assigned:

- "cardiac cardiovascular systems"
- "surgery"
- "radiology nuclear medicine medical imaging"
- "medicine general internal"
- "medicine research experimental"
- "transplantation"
- "pathology",
- "veterinary sciences" .

We focused the search to only peer-reviewed articles. In web of Science searching for topic means that the keywords or combination of keywords are searched within a record Title, Abstract, Author Keyword or Web of Science

Keywords Plus®. This last set of keywords are assigned by Web of Science team of experts rather than the articles authors to improve search engine results matching to searched topics.

2. Classification of articles

Two reviewers (G.G. and T.A.) independently critically appraised and classified the data for each study: level of evidence, country of study origin (based on corresponding author's affiliation), type of intervention (e.g. TAVI, aVR), manufacturer (e.g. Medtronic, Edwards Lifesciences), type of approach (e.g. transfemoral for TAVI or minimally-invasive for aVR), and for each intervention arm: number of patients, mean age (and standard deviation), number of patients dead and follow-up time at which mortality was reported, and number of patients developing a disabling cerebrovascular accident (CVA), major vascular complications, and moderate-severe paravalvular leak (for the last three outcomes 30-day rates were extracted). Any discrepancies were resolved by consensus.

When considering the type of intervention we found that aVR intervention appears in 376 different articles while TAVI intervention appears in 728 different articles.

3. Network analysis of citation network

a) Construction of the citation network

Given the bibliographic information provided by Web of Science Search®, namely the reference list for each article included in the set of 2,595 collected articles, and using VOS Viewer®, a software specifically designed for constructing citation and collaboration networks based on reference list, we extrapolated the citation network.

The citation network analysed in this study comprises 2,216 articles and 10,595 links (See Figure S1). In order for an article to be included in the citation network, one of the following two criteria needs to be met: an article makes a citation to at least one other article in the set of articles collected; an article is cited by at least one other article in the set of articles collected. Out of our set of collected articles, 1,879 articles made at least one citation to other articles while instead 1,306 received at least one citation from one of the other articles. Each link of the network represents a citation.

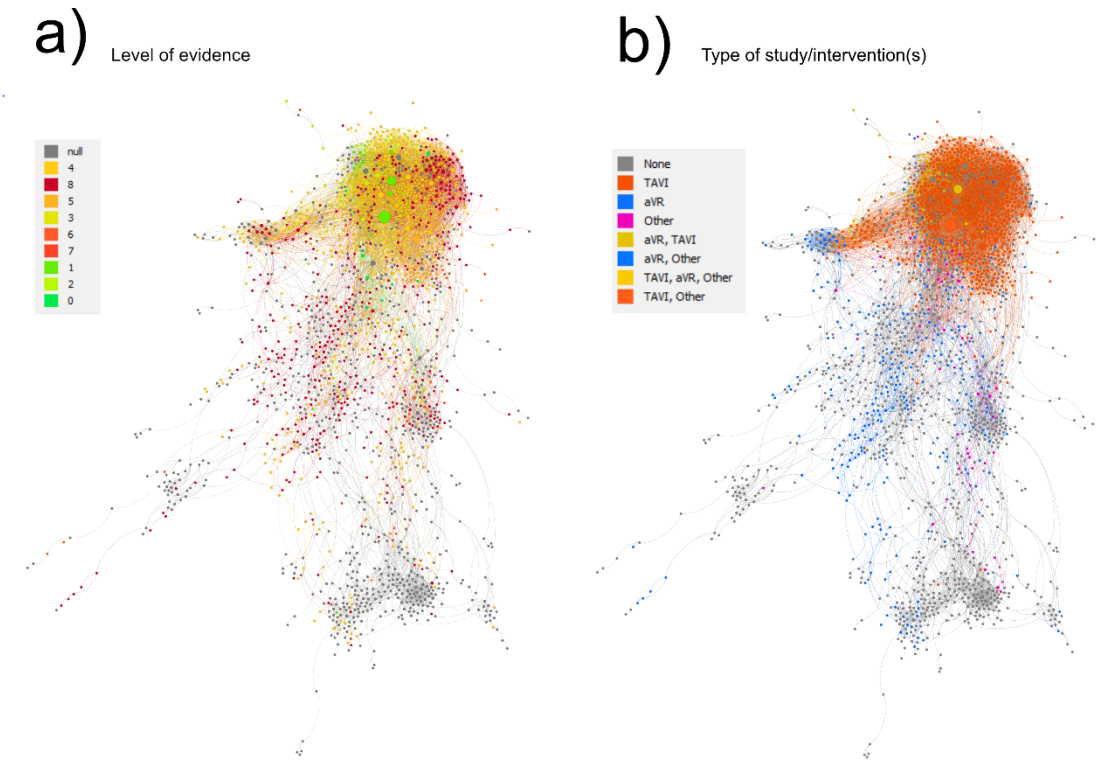


Figure S1. The giant component (2,006 articles, 10,491 citation links) of the citation network comprising articles concerning "aortic valve stenosis" and "aortic valve regurgitation". a) each node is colored according to the assigned level of evidence and sized according to the number of citations it received. b) each node is colored according to the type of intervention(s) outlined in the article and sized according to the number of citations it received.

b) Measuring nodes' network-based characteristics

Given a certain network, we can measure the local properties of its nodes. For the purposes of this study we measured the so-called *in-degree* centrality and a revised version of the so-called *virality* [1]. To better understand the definitions of the two measures, first we have to define the binary adjacency matrix A, that is used to formally model the connection structure of a network. Each entry a_{ij} of the matrix is a binary value. When a_{ij} is equal to one, this means that nodes i and j are connected by a link. When it equals to zero then no

connection exists between the two nodes i and j. The direction of the link in the context of citation networks is relevant and needs to be taken into account. Given the matricial representation of the network, the row order i of the entry a_{ij} represents the identificator of the node from which a link departs while the column order j represents the identificator of the node that is pointed by the link departing from i. Moreover in a citation network, mutual citations between the same couple of nodes cannot exist. So an article that cites another article, cannot be logically cited by the latter, being this precedent in terms of date of publication. Formally, if $a_{ij} = 1$ then $a_{ij} = 0$. For this reason, citation networks fall into the broader class of Directed Acyclic Graphs (DAG).

We them can define the in-degree centrality of a node as the sum of all links (citations) that point to that node. In other words, the in-degree centrality in the context of citations network, is the count of citations received by an article. Formally, the in-degree of a node can be defined as it follows

$$d_i^{in} = \sum_j a_{ji}$$

When speaking about virality instead it is important to consider another fundamental concept, the path going from one node to another node. A path can be defined as the number of links that are necessary to consider in order for the two nodes to be directly or indirectly connected strictly following the direction of the links and where the nodes at the extremes of each link are visited only once. There can be more than one possible path connecting two nodes. We then call the *shortest path* d_{ij} that path that is characterised by the

minimum number of links connecting the two nodes i and j. In other words, the shortest path is the minimum distance, or geodesic, among to nodes.

Given the definition of shortest path we can defined the virality as showed in the equation in the main manuscript, where after selecting the seed node s for which we want to evaluate the virality, we compute all the minimum distances connecting the node s to those nodes that belong to the citation cascade generated by it, generation after generation, and we evaluate the average value of all geodesics. As describe in Garas et al. (2017),[2] the virality defined in this way can be used as a proxy of the length of a cascade generated by a node. Longer a cascade is, more its seed article virality is characterised by a longer and persistent chain of indirect citations.

4. Uptake of TAVI interventions

We investigated the uptake of TAVI intervention interpolating an exponential function into data showing evolution over time of sum of number of patients included in clinical studies (See Figure 21 in the main text).

A linear trend model has been computed for natural log of *sum of number of patients* given the year of publication of studies including those patients as cohorts as the following formulas show:

$$P = \alpha e^{\beta Y}$$
$$\ln(P) = \ln(\alpha) + \beta Y$$

Term	Value	StdErr	t-value	p-value
Year	0.888201	0.0669125	13.2741	< 0.0001
intercept	-1778.63	134.484	-13.2256	< 0.0001

Observations: 13
Residual degrees of freedom: 11
SSE (sum squared error): 10.2227
MSE (mean squared error): 0.929341
R-Squared: 0.941939
Standard error: 0.964023
p-value (significance): < 0.0001

Table S6. Exponential trend best fitting model.

5. Comparison of aVR and TAVI mortality rates distributions and time trend models

We assessed differences of mortality rates per person-year distributions by type of intervention, controlling for age groups (three groups: <60, 61-80, >80-years). We performed two non-parametric statistical tests (Kruskal-Wallis and Wilcoxon signed-rank tests) in view of the heterogeneous distributions (See Figures S32 and S33) of mortality rates per person-year for each group.

Intervention	Median Mortality per person-year	Number of observations
TAVI	25.1	792
aVR	4.6	382

Table S2. Median values and number of observations of distribution of mortality rates per person-year by intervention.

Age range	Intervention	Median Mortality per person-year	Number of observations
<60	aVR	1.265	49
61-80	TAVI	22.2	123
61-80	aVR	6.62	114
81+	TAVI	26.15	340
81+	aVR	10.1	13

Table S3. Median values and number of observations of distribution of mortality rates per person-year by intervention and by age range.

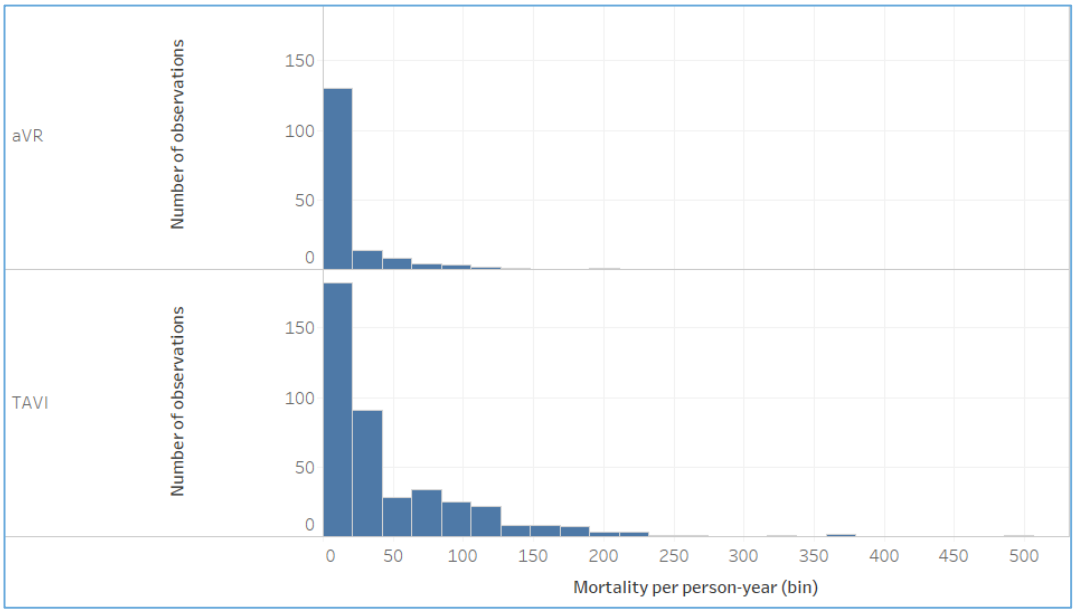


Figure S32. Histograms of mortality rates per person-year distributions by intervention.

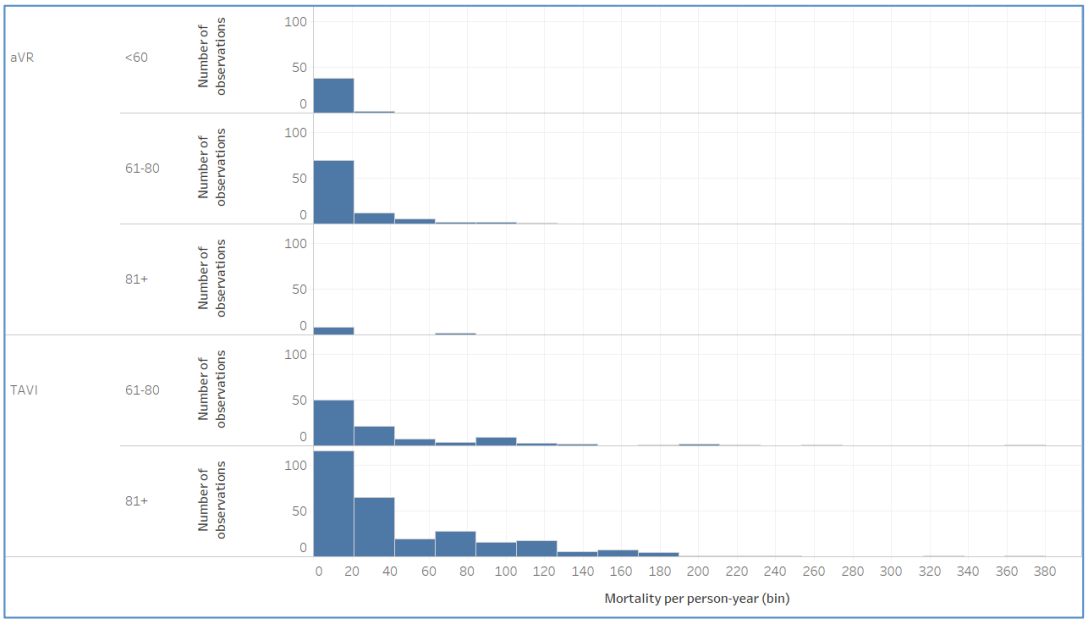


Figure S33. Histograms of mortality rates per person-year distributions by intervention and age group.

c) Difference between aVR and TAVI mortality rates per person-year (not age stratified)

Test type	Test result
Kruskal-Wallis rank sum test	KW chi-squared = 102.54, df = 1, p-value <2.2e-16
Wilcoxon rank sum test with continuity correction	W = 11913, p-value <2.2e-16

Table S4. Non-parametric statistical test for group differences in terms of mortality rates per person year.

d) Difference between aVR and TAVI mortality rates per person-year (age stratified)

Test type	Test result by age group		
	<60	61-80	81+
Kruskal-Wallis rank sum test	No observations for TAVI	Kruskal-Wallis chi-squared = 15.978, df = 1, p-value = 6.407e-05	Kruskal-Wallis chi-squared = 7.2514, df = 1, p-value = 0.00708
Wilcoxon rank sum test with continuity correction	No observations for TAVI	W = 3209, p-value = 6.442e-05	W = 708.5, p-value = 0.007125

Table S5. Non-parametric statistical test for group differences in terms of mortality rates per person year.

e) Time trend lines of aVR and TAVI mortality rates per person-year

Term	Value	StdErr	t-value	p-value
Year	-17.8609	4.56222	-3.91496	0.003538
intercept	35990.9	9174.64	3.92286	0.0034957
Observations:11 Residual degrees of freedom:9 SSE (sum squared error):20605.7 MSE (mean squared error):2289.53 R-Squared:0.630039 Standard error:47.849 p-value (significance):0.003538				

Table S6. Linear trend line model coefficients for TAVI intervention mortality rates per person-year.

Term	Value	StdErr	t-value	p-value
Year	-0.223474	0.425259	-0.525499	0.604058
intercept	457.128	852.013	0.536527	0.596533
Observations: 26 Residual degrees of freedom : 24 SSE (sum squared error): 6347.67 MSE (mean squared error): 264.486 R-Squared: 0.0113753 Standard error: 16.263 p-value (significance): 0.604058				

Table S7. Linear trend line model coefficients for aVR intervention mortality rates per person-year.

6. Comparison of aVR and TAVI innovation indices and relationship between mortality per person-year and implementation stage

We computed for each intervention group its innovation index (see Figure S34). We also analysed for each group the relationship between the implementation stage and the mortality per person-year using linear model (see Table S9 and S10).

Surgical Innovation Stage	Description
1	Randomised controlled trial
2	Non-randomised controlled trial
3	Observational study with controls
4	Observational study without controls
5	First-in-human study
6	Cadaveric study
7	Animal study
8	Description of idea / laboratory evaluation

Table S8. Implementation stages description.

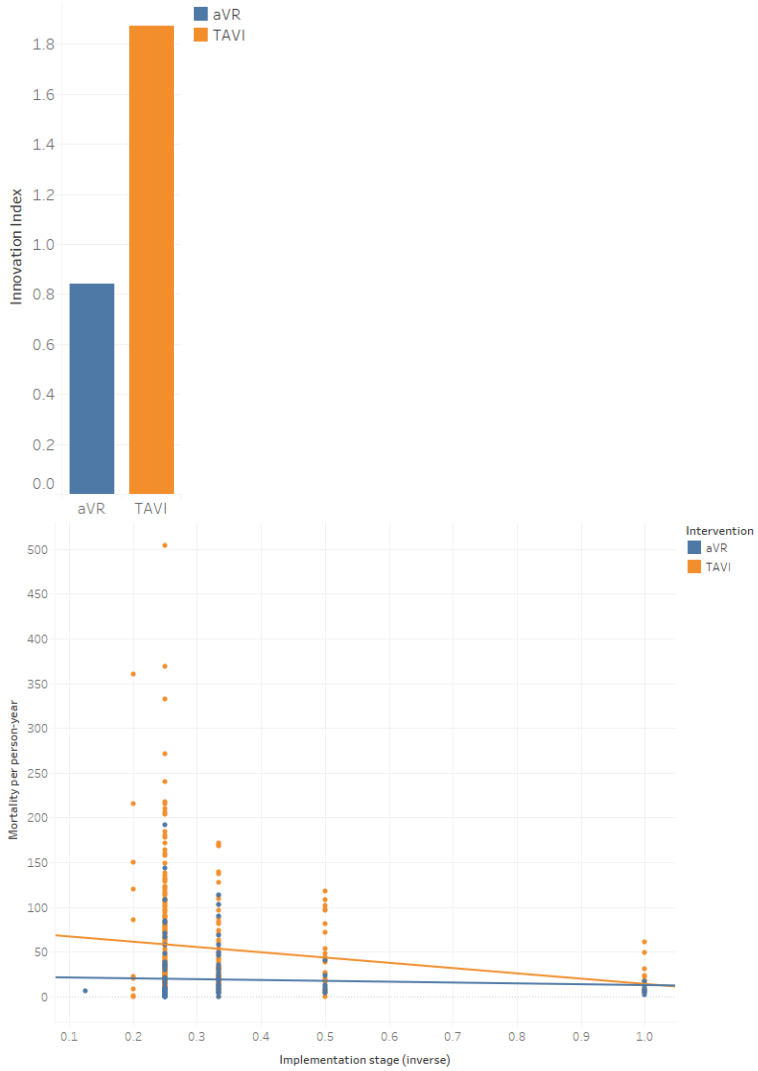


Figure S34. a) Innovation index for aVR and TAVI articles; b) relationship between intervention stage (inverse value) and mortality per person-year.

Term	Value	StdErr	t-value	p-value
Implementation stage (inverse)	-58.8983	21.5337	-2.73517	0.006601
intercept	73.0262	7.85121	9.30126	< 0.0001
Observations=306 Residual degree of freedom=304 SSE (sum squared error): 1.23431e+06 MSE (mean squared error): 4060.22 R-Squared: 0.024018 Standard error: 63.7198 p-value (significance): 0.006601				

Table S9. Linear model of the relationship between implementation stage (inverse) and mortality per person-year for TAVI intervention.

Term	Value	StdErr	t-value	p-value
Implementation stage (inverse)	-9.39137	14.3816	-0.653013	0.515018
intercept	22.2138	5.59062	3.97341	0.0001224
Observations=120 Residual degree of freedom=118 SSE (sum squared error): 116479 MSE (mean squared error): 987.111 R-Squared: 0.0036008 Standard error: 31.4183 p-value (significance): 0.515018				

Table S10. Linear model of the relationship between implementation stage (inverse) and mortality per person-year for aVR intervention.

7. Mortality rates per person-year and virality relationship by intervention

We first assessed the two intervention groups differences in terms of virality distributions (see Table S11) using the above cited non- parametric test we found the two groups to be not statistically different.

Test type	Test result
Kruskal-Wallis rank sum test	Kruskal-Wallis chi-squared = 0.019927, df = 1, p-value = 0.8877
Wilcoxon rank sum test with continuity correction	W = 51757, p-value = 0.8879

Table S11. Non-parametric statistical test for group differences in terms of virality distributions.

Non parametric test have been used in view of the non-normal distributions of virality (see Figure S35).

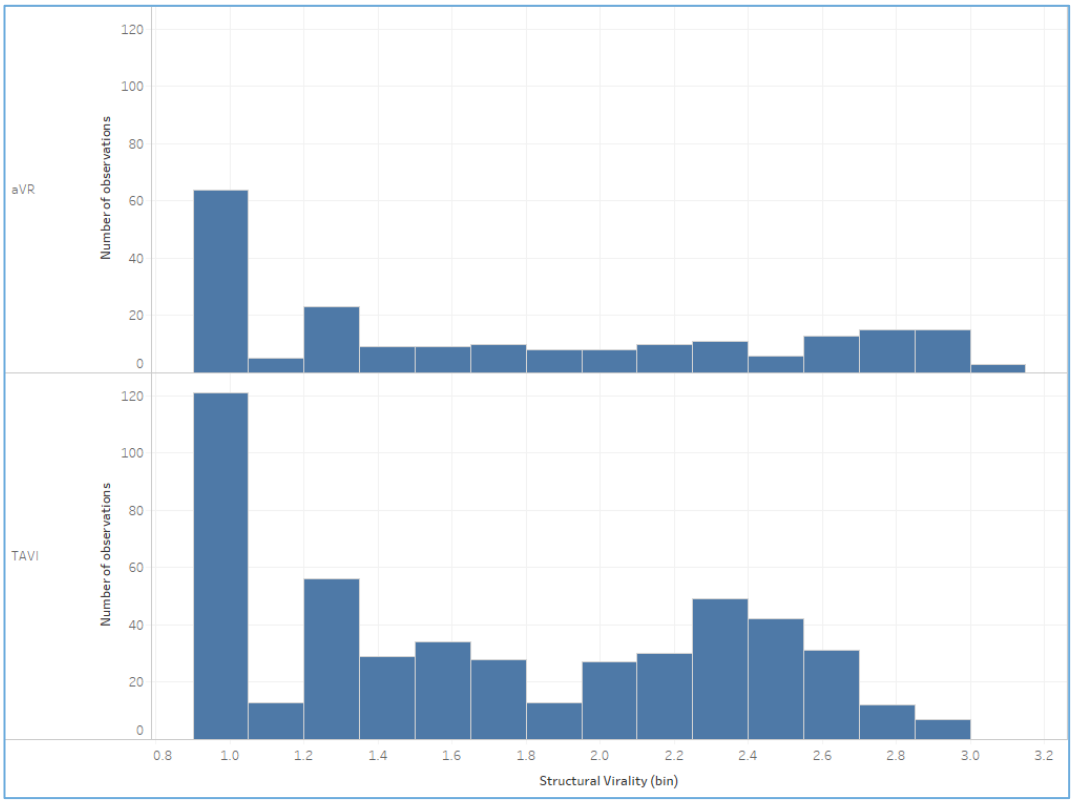


Figure S35. Histograms showing distribution of structural virality scores obtained by articles in aVR and TAVI intervention groups.

We then assessed for each intervention group the existence of differences in terms the nature of relationship of mortality rates per person-year with virality of the associated study (see Figure 23 in the main text).

We selected as best fitting model to describe the relationship a polynomial model of degree 2 for both groups. Table S12 and S13 reports the coefficients and p-values for each term of the polynomial model. In brackets comparison of the selected models with linear models are reported in correspondence of standard indicators of model squared errors and significance.

Term	Value	StdErr	t-value	p-value
Structural Virality^2	73.0574	16.4539	4.44012	< 0.0001
Structural Virality	-253.875	59.0014	-4.30286	< 0.0001
intercept	252.963	47.9324	5.2775	< 0.0001

Observations=256
Residual degree of freedom=253
SSE (sum squared error): 1.11574e+06 (Linear: 1.20268e+06)
MSE (mean squared error): 4410.02 (Linear: 4734.95)
R-Squared: 0.0749971 (Linear: 0.0029177)
Standard error: 66.408 (Linear: 68.811)
p-value (significance): < 0.0001 (Linear: 0.389428)

Table S12. Polynomial model of degree 2 of the relationship between structural virality and mortality per person-year for TAVI intervention.

Term	Value	StdErr	t-value	p-value
Structural Virality^2	-8.64376	8.43605	-1.02462	0.308171
Structural Virality	33.3916	32.2555	1.03522	0.30322
intercept	-11.5516	27.1011	-0.426242	0.670906

Observations=97
Residual degree of freedom=94
SSE (sum squared error): 80226.5 (Linear: 81122.5)
MSE (mean squared error): 853.474 (Linear: 853.922)
R-Squared: 0.0112724 (Linear: 0.0002297)
Standard error: 29.2143 (Linear: 29.2219)
p-value (significance):0.586951 (Linear: 0.882857)

Table S13. Polynomial model of degree 2 of the relationship between Structural Virality and mortality per person-year for aVR intervention.

8. Mortality rates per person-year and altmetric score relationship by intervention

We assessed the two intervention groups differences in terms of altmetric scores distributions using the above cited non-parametric test we found the two groups to be statistically different.

Test type	Test result
Kruskal-Wallis rank sum test	Kruskal-Wallis chi-squared = 50.96, df = 1, p-value = 9.427e-13
Wilcoxon rank sum test with continuity correction	W = 118570, p-value = 9.434e-13

Table S14. Non-parametric statistical test for group differences in terms of virality altmetric distribution.

Non-parametric tests were used in view of the heterogeneous distributions of altmetric scores (see Figure S6).

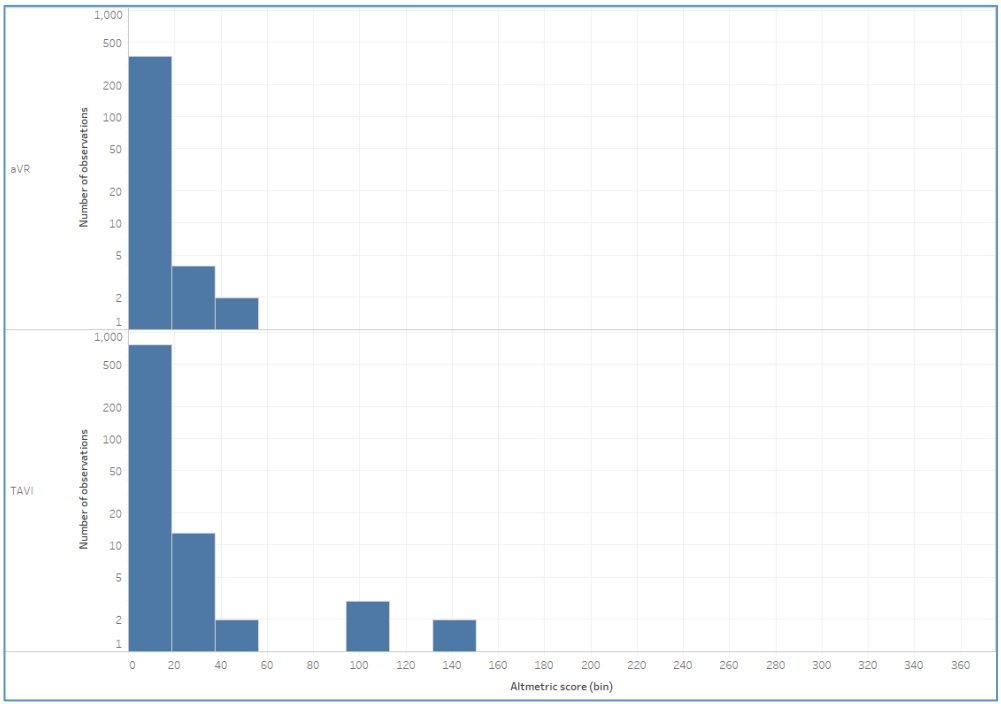


Figure S6. Histograms showing distribution of altmetric scores obtained by articles in aVR and TAVI intervention groups.

We also considered the nature of relationship between the indicator of virality and altmetric score using a Kendall's rank correlation test (after assessing the not normality of virality and altmetric score distributions) and found that two measures do not correlate with a correlation coefficient tau= -0.09798 (p-value = 0.0008234).

We then assessed for each intervention group the existence of differences in terms the nature of relationship of mortality rates per person-year with the altmetric score of the associated study (see Figure 24 in the main text).

We selected as best fitting model to describe the relationship a logarithmic model for the TAVI intervention group and a linear model for the aVR intervention group. Table S15 and S16 report the coefficients and p-values for each term of the models. In brackets comparison of the selected models with linear models are reported in correspondence of standard indicators of model squared errors and significance.

Term	Value	StdErr	t-value	p-value
ln(altmetric_score)	-4.01903	1.83543	-2.18969	0.0298328
intercept	45.0483	3.43585	13.1113	< 0.0001
Observations=188 Residual degree of freedom=186 SSE (sum squared error): 323536 (Linear: 325904) MSE (mean squared error): 1797.42 (Linear: 1810.58) R-Squared: 0.0259463 (Linear: 0.0188166) Standard error: 42.396 (Linear: 42.5509) p-value (significance): 0.0298328 (Linear: 0.0648109)				

Table S15. Logarithmic regression model for TAVI.

Term	Value	StdErr	t-value	p-value
altmetric_score	-0.029988	0.0557139	-0.53825	0.592892
intercept	16.2684	3.83268	4.24466	< 0.0001
Observations=50 Residual degree of freedom=50 SSE (sum squared error): 31437.6 MSE (mean squared error): 654.949 R-Squared: 0.0059995 Standard error: 25.592 p-value (significance): 0.592892				

Table S16. Linear regression model for aVR.

9. Mortality rates per person-year and citations relationship by intervention

We assessed for each intervention group the existence of differences in terms the nature of relationship of mortality rates per person-year with the citations received by the associated study.

We selected as best fitting model to describe the relationship a logarithmic model both for TAVI and aVR intervention. Table S17 and S18 report the coefficients and p-values for each term of the models. In brackets comparison of the selected models with linear models are reported in correspondence of standard indicators of model squared errors and significance.

Term	Value	StdErr	t-value	p-value
ln(Citations)	-5.46236	3.06828	-1.78027	0.0762461
intercept	68.35	7.01283	9.74642	< 0.0001
Observations=252 Residual degree of freedom=250 SSE (sum squared error): 1.17995e+06 (Linear: 1.18885e+06) MSE (mean squared error): 4719.81 (Linear: 4755.41) R-Squared: 0.0125187 (Linear: 0.00507) Standard error: 68.7009 (Linear: 68.9595) p-value (significance): 0.0762461 (Linear: 0.260107)				

Table S17. Logarithmic regression model for TAVI.

Term	Value	StdErr	t-value	p-value
ln(Citations)	-2.00048	2.95692	-0.676542	0.500432
intercept	18.8855	4.56977	4.1327	< 0.0001
Observations=92 Residual degree of freedom=90 SSE (sum squared error): 79840.9 (Linear: 80220.8) MSE (mean squared error): 887.121 (Linear: 891.342) R-Squared: 0.0003254 (Linear: 0.0050599) Standard error: 29.8554 (Linear: 29.7846) p-value (significance): 0.864491 (Linear: 0.500432)				

Table S18. Logarithmic regression model for TAVI.

References

1. Goel S, Anderson A, Hofman J, Watts DJ. The structural virality of online diffusion. Management Science. 2015 Jul 22;62(1):180-96.
2. Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery. PloS one. 2017 Aug 25;12(8):e0183332.

Appendix 5

10. Search strategy

Our study draws on the Web of Science® (WOS) All Databases Search Engine (Clarivate Analytics, Philadelphia, PA). A search has been performed the 2nd of June 2017 and a collection of 2,884 articles have been retrieved filtering out publications that have as Topic "aortic valve stenosis" and "aortic valve regurgitation". We then refined the search by a subset of categories Web of Science® categories to which an article can be assigned:

- "cardiac cardiovascular systems"
- "surgery"
- "radiology nuclear medicine medical imaging"
- "medicine general internal"
- "medicine research experimental"
- "transplantation"
- "pathology",
- "veterinary sciences".

We focused the search to only peer-reviewed articles. In web of Science searching for topic means that the keywords or combination of keywords are searched within a record Title, Abstract, Author Keyword or Web of Science Keywords Plus®. This last set of keywords are assigned by Web of Science

team of experts rather than the articles authors to improve search engine results matching to searched topics.

11. Classification of articles

Two reviewers (G.G. and T.A.) independently critically appraised and classified the data for each study: level of evidence, country of study origin (based on corresponding author's affiliation), type of intervention (e.g. TAVI, aVR), manufacturer (e.g. Medtronic, Edwards Lifesciences), type of approach (e.g. transfemoral for TAVI or minimally-invasive for aVR), and for each intervention arm: number of patients, mean age (and standard deviation), number of patients dead and follow-up time at which mortality was reported, and number of patients developing a disabling cerebrovascular accident (CVA), major vascular complications, and moderate-severe paravalvular leak (for the last three outcomes 30-day rates were extracted). Any discrepancies were resolved by consensus.

Given the bibliographic information we obtained from Web of Science Search®, whenever the affiliation of the corresponding author was available, we used it to attribute to an article a unique geographical region. The list countries appearing in the affiliations address for each of the 7 regions here considered (Africa, Asia, Central and South America, Europe, Middle East, North America and Oceania) are showed in Table S1.

Region	Countries
Africa	**Cameroon, Egypt, Ethiopia, Kenia, Morocco, South Africa, Uganda, Zimbabwe**
Asia	**China, India, Indonesia, Japan, Nepal, Pakistan, Singapore, South Korea, Sri Lanka, Taiwan**
Central and South America	**Argentina, Brazil, Chile, Guyana, Mexico, Uruguay**
Europe	**Austria, Belgium, Bosnia and Herzegovina, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, Yuguslavia**
Middle East	**Iran, Israel, Saudi Arabia, Syria, United Arab Emirates, Yemen**
North America	**Canada, USA**
Oceania	**Australia, New Zealand**

Table S1. List of regions and countries.

12. Network analysis of citation network

3.1 Construction of the citation network

Given the bibliographic information provided by Web of Science Search®, namely the reference list for each article included in the set of 2,595 collected articles, and using VOS Viewer®, a software specifically designed for constructing citation and collaboration networks based on reference list, we extrapolated the citation network.

The citation network analysed in this study comprises 2,216 articles and 10,595 links. In order for an article to be included in the citation network, one of the following two criteria needs to be met: an article makes a citation to at least one other article in the set of articles collected; an article is cited by at

least one other article in the set of articles collected. Out of our set of collected articles, 1,879 articles made at least one citation to other articles while instead 1,306 received at least one citation from one of the other articles. Each link of the network represents a citation.

3.2 Measuring nodes' network-based characteristics

Given a certain network, we can measure the local properties of its nodes. For the purposes of this study we measured the so-called *in-degree* centrality and a revised version of the so-called *virality* [1]. To better understand the definitions of the two measures, first we have to define the binary adjacency matrix A, that is used to formally model the connection structure of a network. Each entry a_{ij} of the matrix is a binary value. When a_{ij} is equal to one, this means that nodes i and j are connected by a link. When it equals to zero then no connection exists between the two nodes i and j. The direction of the link in the context of citation networks is relevant and needs to be taken into account. Given the matricial representation of the network, the row order i of the entry a_{ij} represents the identificator of the node from which a link departs while the column order j represents the identificator of the node that is pointed by the link departing from i. Moreover in a citation network, mutual citations between the same couple of nodes cannot exist. So an article that cites another article, cannot be logically cited by the latter, being this precedent in terms of date of publication. Formally, if $a_{ij} = 1$ then $a_{ij} = 0$. For this reason, citation networks fall into the broader class of Directed Acyclic Graphs (DAG).

We them can define the *in-degree centrality* of a node as the sum of all links (citations) that point to that node. In other words, the in-degree centrality in the context of citations network, is the count of citations received by an article. Formally, the in-degree of a node can be defined as it follows

$$d_i^{in} = \sum_j a_{ji}$$

When speaking about virality instead it is important to consider another fundamental concept, the path going from one node to another node. A path can be defined as the number of links that are necessary to consider in order for the two nodes to be directly or indirectly connected strictly following the direction of the links and where the nodes at the extremes of each link are visited only once. There can be more than one possible path connecting two nodes. We then call the *shortest path* d_{ij} that path that is characterised by the minimum number of links connecting the two nodes i and j. In other words, the shortest path is the minimum distance, or geodesic, among to nodes.

Given the definition of shortest path we can defined the virality as showed in the equation in the main manuscript, where after selecting the seed node s for which we want to evaluate the virality, we compute all the minimum distances connecting the node s to those nodes that belong to the citation cascade generated by it, generation after generation, and we evaluate the average value of all geodesics. As described by Garas et al. (2017)[2], the virality defined in this way can be used as a proxy of the length of a cascade generated by a

node. Longer a cascade is, more its seed article virality is characterised by a longer and persistent chain of indirect citations.

4 Statistical analysis of regional group differences

4.1 Number of observations

For some articles show more than one arm (intervention), from a minimum of 2 to a maximum of 3, we disentangled the clinical outcomes (mortality, percentage of patients with disabling CVA, percentage of patients with major vascular complications and percentage of patients with moderate or severe paravalvular leak) for each single intervention (i.e. an article comparing two TAVI interventions, will result in two observations of outcomes). For this reason, we evaluated the overall clinical performance of each regional group over the 792 observations originating from 728 articles concerning TAVI intervention(s). We instead evaluated all the other outcomes based on the 728 set of articles as unit of observation. Indeed the level of evidence as well as the count of citations, the structural virality and the altmetric score are associated to the single article, independently from the number of interventions that are exhibited.

Measure	# Total available observations
Mortality per person year	N=417 (N missing=375)
Africa	No observations
Asia	N=15
Central and South America	N=13
Europe	N=272
Middle East	N=8
North America	N=101
Oceania	N=8

% Patients with Disabling CVA	N=351 (N missing=441)
Africa	No observations
Asia	N=13
Central and South America	N=11
Europe	N=220
Middle East	N=4
North America	N=95
Oceania	N=8
% Patients with major vascular complications	N=296 (N missing=496)
Africa	No observations
Asia	N=13
Central and South America	N=9
Europe	N=190
Middle East	N=3
North America	N=76
Oceania	N=5
% Patients with moderate or severe paravalvular leak	N=437 (N missing=355)
Africa	No observations
Asia	N=14
Central and South America	N=11
Europe	N=281
Middle East	N=8
North America	N=114
Oceania	N=9
Citations	N=451(N missing=277)
Africa	No observations
Asia	N=13
Central and South America	N=4
Europe	N=293
Middle East	N=9
North America	N=122
Oceania	N=10
Structural Virality	N=451 (N missing=277)
Africa	No observations
Asia	N=13
Central and South America	N=4
Europe	N=293
Middle East	N=9
North America	N=122
Oceania	N=10
Altmetric	N=284 (N missing=444)
Africa	No observations
Asia	N=6
Central and South America	N=4
Europe	N=174
Middle East	N=5
North America	N=88
Oceania	N=7

Table S1. Number of observations by measure and regional group.

4.2 Distribution of measures

In this paragraph we provide evidence on the characteristic of Europe's and North America's distributions of the following variables: mortality per person year, percentage of patients with disabling CVA, percentage of patients with major vascular complications, percentage of patients with moderate or severe paravalvular leak, citations, structural virality and altmetric score. The aim here is to show how these two groups of observations, across all the measures here considered, are far from being characterised by normal distributions. Visual comparisons of the two univariate Kernel distribution estimates corresponding to each of the two groups are provided to support the statistical result reported in section 4.3.

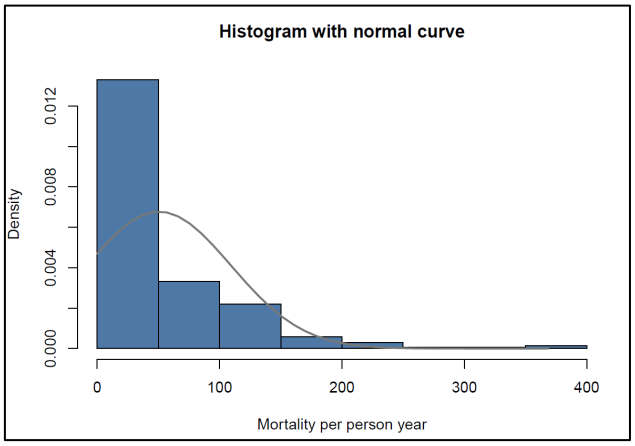


Figure S1. Histogram with corresponding normal curve for Europe's distribution of mortality per person year.

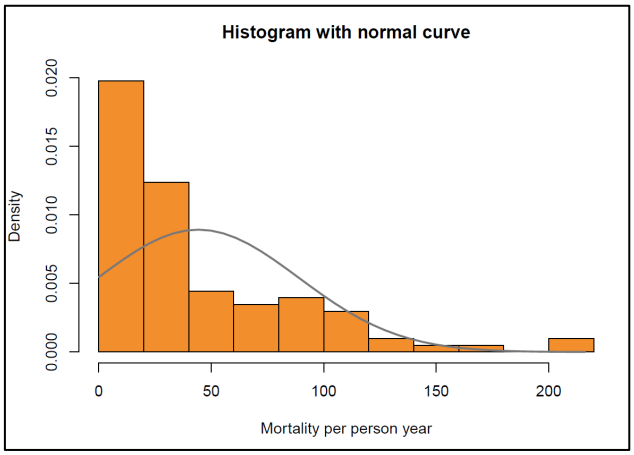


Figure S2. Histogram with corresponding normal curve for North America's distribution of mortality per person year.

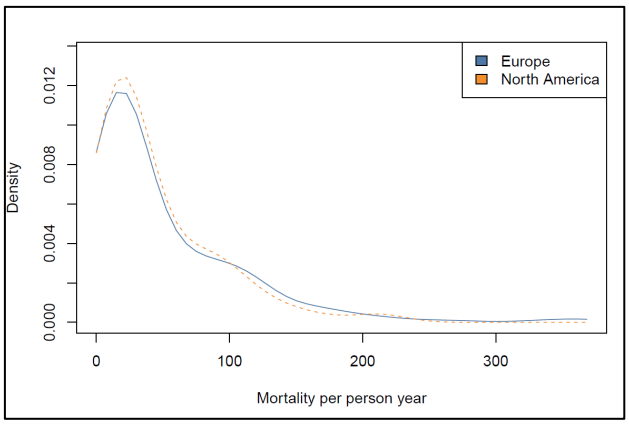


Figure S3. Comparison of univariate Kernel density estimates for mortality per person year distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

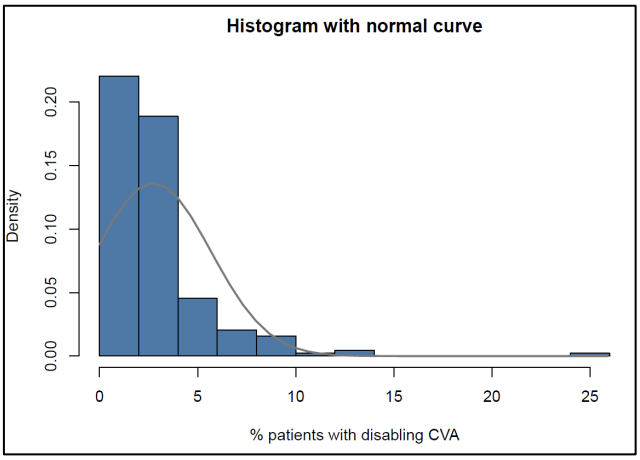


Figure S4. Histogram with corresponding normal curve for Europe's distribution of percentage of patients with disabling CVA.

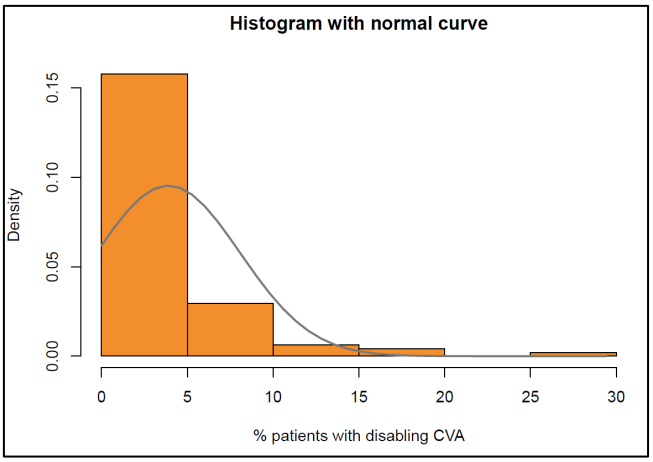


Figure S5. Histogram with corresponding normal curve for North America's distribution of percentage of patients with disabling CVA.

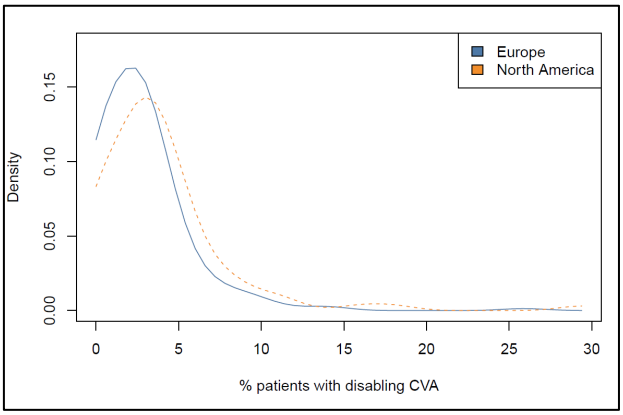


Figure S6. Comparison of univariate Kernel density estimates for percentage of patients with disabling CVA distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

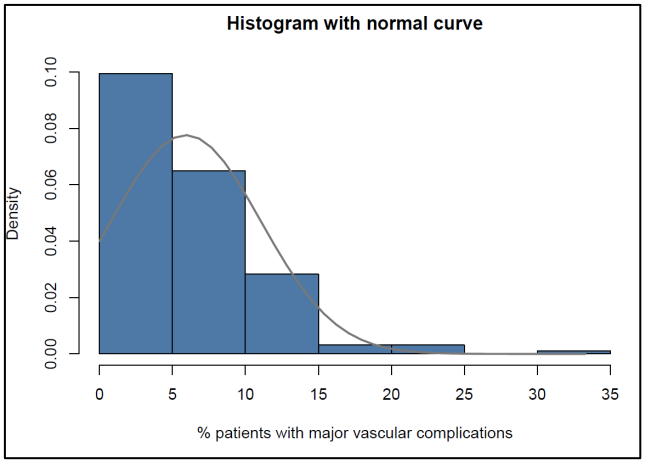


Figure S7. Histogram with corresponding normal curve for Europe's distribution of percentage of patients with major vascular complications.

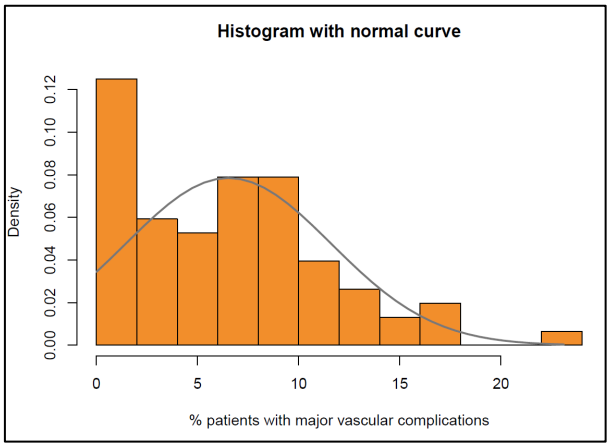


Figure S8. Histogram with corresponding normal curve for North America's distribution of percentage of patients with major vascular complications.

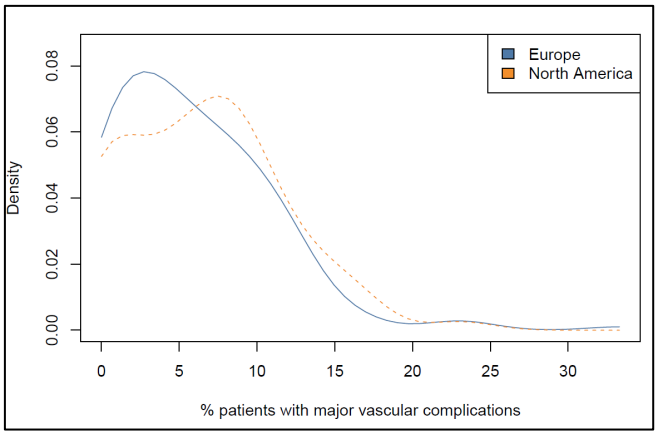


Figure S9. Comparison of univariate Kernel density estimates for percentage of patients with major vascular complications distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

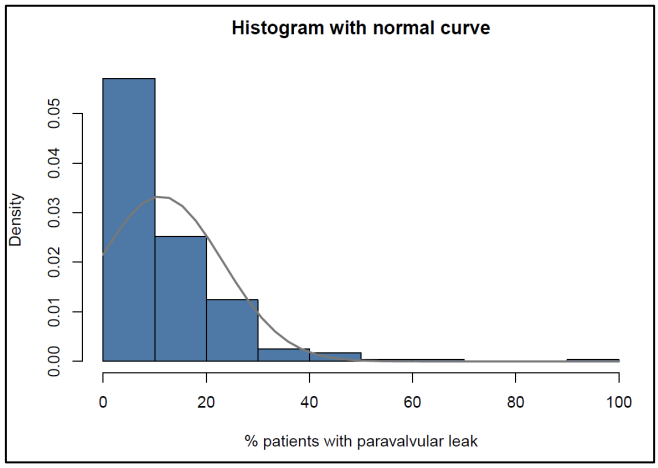


Figure S10. Histogram with corresponding normal curve for Europe's distribution of percentage of patients with moderate or severe paravalvular leak.

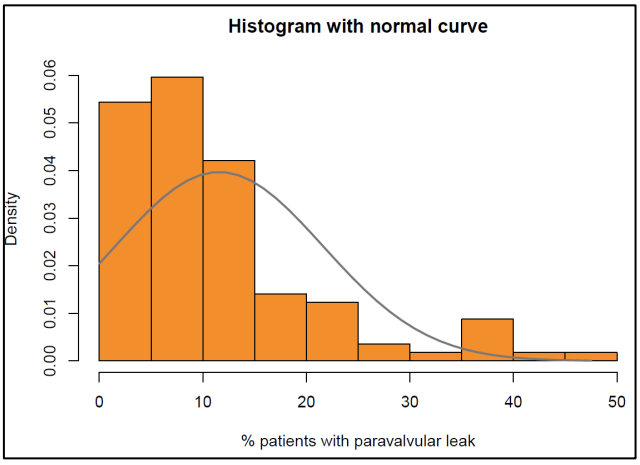


Figure S11. Histogram with corresponding normal curve for North America's distribution of percentage of patients with moderate or severe paravalvular leak.

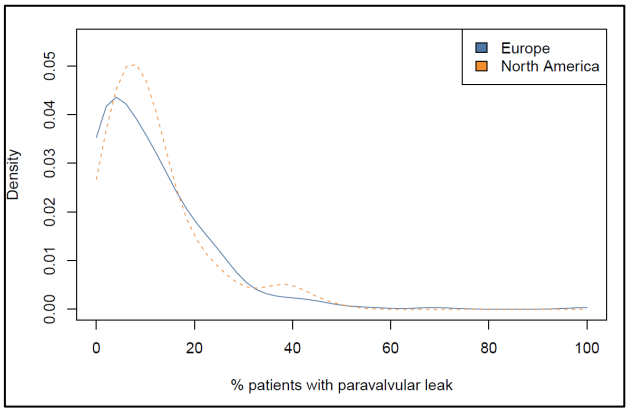


Figure S12. Comparison of univariate Kernel density estimates for percentage of patients with moderate or severe paravalvular leak distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

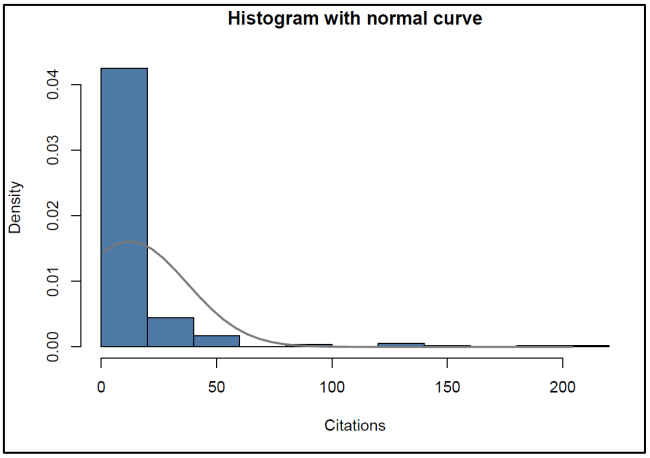


Figure S13. Histogram with corresponding normal curve for Europe's distribution of citations.

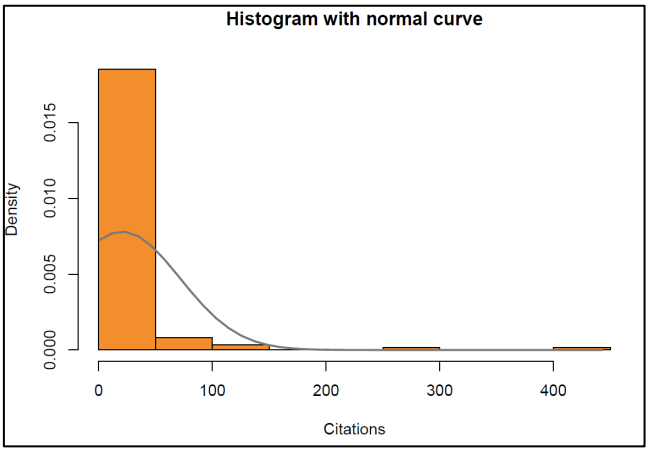


Figure S14. Histogram with corresponding normal curve for North America's distribution of citations.

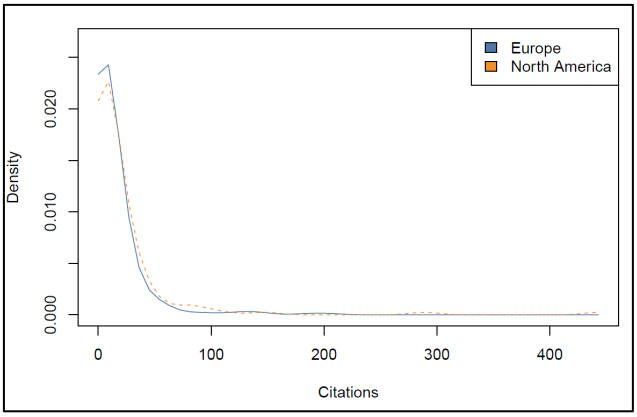


Figure S15. Comparison of univariate Kernel density estimates for citations distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

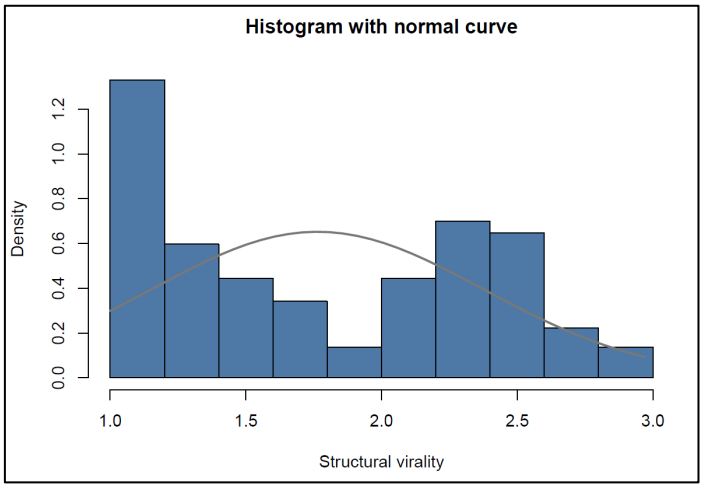


Figure S16. Histogram with corresponding normal curve for Europe's distribution of structural virality.

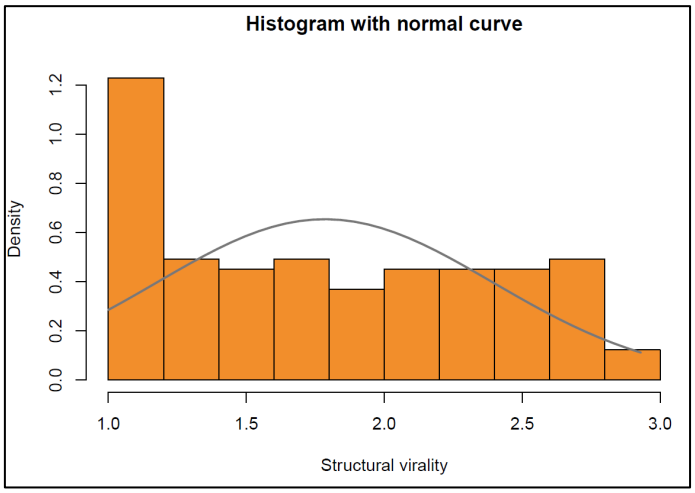


Figure S17. Histogram with corresponding normal curve for Europe's distribution of structural virality.

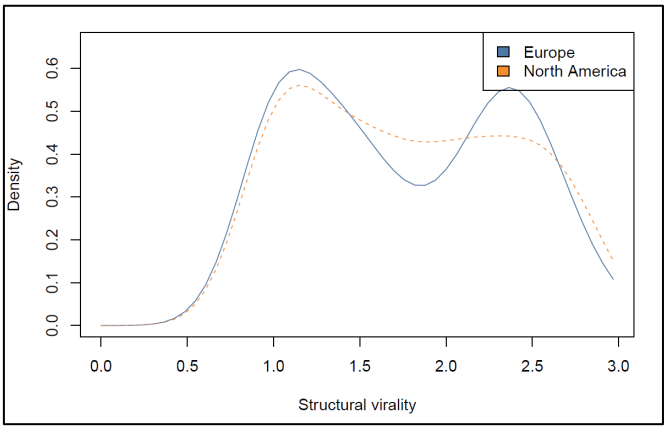


Figure S18. Comparison of univariate Kernel density estimates for structural virality distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

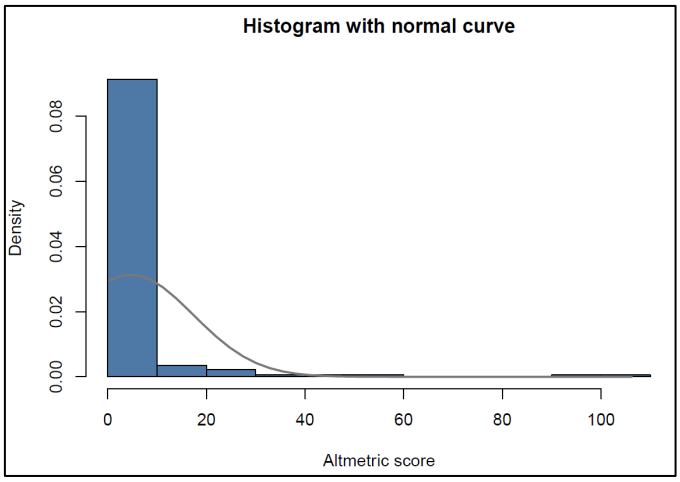


Figure S19. Histogram with corresponding normal curve for Europe's distribution of altmetric score.

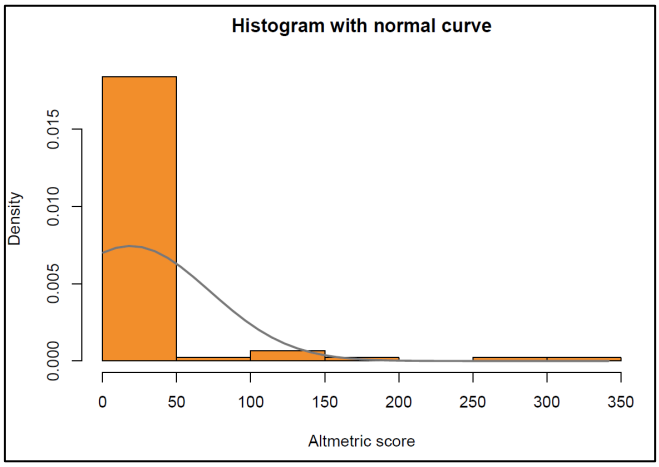


Figure S20. Histogram with corresponding normal curve for North America's distribution of altmetric score.

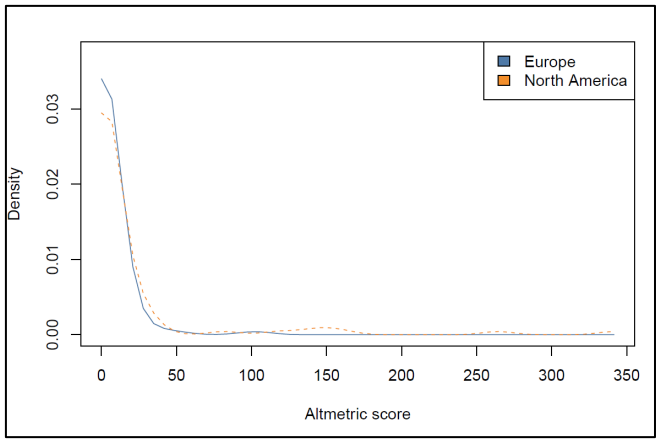


Figure S21. Comparison of univariate Kernel density estimates for altmetric score distributions for Europe and North America regions; smoothing parameter is set to be as the mean of the normal optimal values for the different groups; estimated curves are intentionally showed within the maximum range of observation across the different groups (right and left tails have been truncated).

4.3 Pairwise comparison of distributions: Wilcoxon one-sided test

We showed in section 4.2 of this supplementary material (Figures S1-S21), the variables used in this study are never characterised by a normal distribution for any of the two regional groups. Therefore, to make a statistical comparison of

the two regional groups we opted for a non-parametric test, allowing also for one-sided comparison.

We performed a two-sample one-sided Wilcoxon rank sum test tests with continuity correction to assess statistically the significance of the difference between Europe's and North America distributions of mortality per person year, percentage of patients with disabling CVA, percentage of patients with major vascular complications, percentage of patients with moderate or severe paravalvular leak, citations, structural virality and altmetric score.

The two groups have been first ranked in a decreasing order according to the median values of distributions for each measure, so that the test was aimed to assess if the first group in rank was significantly higher than the second. The significance values one-tailed right test for each pair of distributions (alternative hypothesis: distribution of group 1 (G1) is shifted to the right of distribution of group 2 (G2) are reported in the following table:

Distribution G1	Distribution G1	Measure	p-value
North America	Europe	Mortality per person year	0.470
North America	Europe	% patients with Disabling CVA	0.002*
North America	Europe	% patients with major vascular complications	0.149
North America	Europe	% patients with moderate or severe paravalvular leak	0.134
North America	Europe	Citations	0.021*
North America	Europe	Structural Virality	0.342
North America	Europe	Altmetric score	0.004*

4.4 Fisher's test

We also performed a Fisher test for dichotomous variables to assess group differences for three of the four measures indicating a clinical outcome: the percentage of patients with disabling CVA, the percentage of patients with major vascular complications and the percentage of patients with moderate or severe paravalvular leak.

The contingency tables that were used to computed odds ratio for each group for any of the three measures are presented below:

Region	Rank	Count of patients with disabling CVA	Count of patients without disabling CVA	Odds ratio
North America	1	1,475	50,485	0.029
Europe	2	2,023	100,226	0.020

Table S1. Contingency table showing the count of patients with and without disabling CVA clinical outcome across all studies accounted to Europe and North America.

Region	Rank	Count of patients with major vascular complications	Count of patients without major vascular complications	Odds ratio
North America	1	2,204	49,756	0.044
Europe	2	3,810	98,439	0.039

Table S2. Contingency table showing the count of patients with and without major vascular complications clinical outcome across all studies accounted to Europe and North America.

Region	Rank	Count of patients with moderate or severe paravalvular leak	Count of patients without moderate or severe paravalvular leak	Odds ratio
North America	1	4,215	47,745	0.088
Europe	2	7,782	94,466	0.082

Table S3. Contingency table showing the count of patients with and without moderate or severe paravalvular leak clinical outcome across all studies accounted to Europe and North America.

G1	G2	Measure	p-value
North America	Europe	% patients with Disabling CVA	3.23e-26*
North America	Europe	% patients with major vascular complications	4.83e-07*
North America	Europe	% patients with moderate or severe paravalvular leak	2.77e-04*

Table S4. Fisher test significance values by measure. Significance threshold is 0.05 (*). The alternative hypothesis is that first ranked group (G1) odds ratio is higher than second ranked group (G2).

Appendix 6

Peer-reviewed publications, copyright permissions forms, and research reports

Garas G, Cingolani I, Patel V, Panzarasa P, Darzi A, Athanasiou T. Evaluating the implications of Brexit for research collaboration and policy: A network analysis and simulation study. *BMJ Open*. 2019 Sep 10;9(9):e0250525.

Published open access (copyright: ©2019 Garas et al.)

Open access Research

BMJ Open

Evaluating the implications of Brexit for research collaboration and policy: a network analysis and simulation study

George Garas,[1] Isabella Cingolani,[2] Vanash M Patel,[1] Pietro Panzarasa,[3] Ara Darzi,[1] Thanos Athanasiou[1]

Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Beyond IDEAL: the importance of surgical innovation metrics. *Lancet*. 2019 Jan 26;393(10169):315.

Garas G, Cingolani I, Patel V, Panzarasa P, Alderson D, Darzi A, Athanasiou T. Surgical innovation in the Era of Global Surgery: A Network Analysis. *Ann Surg.* 2018 Dec 20. [Epub ahead of print]

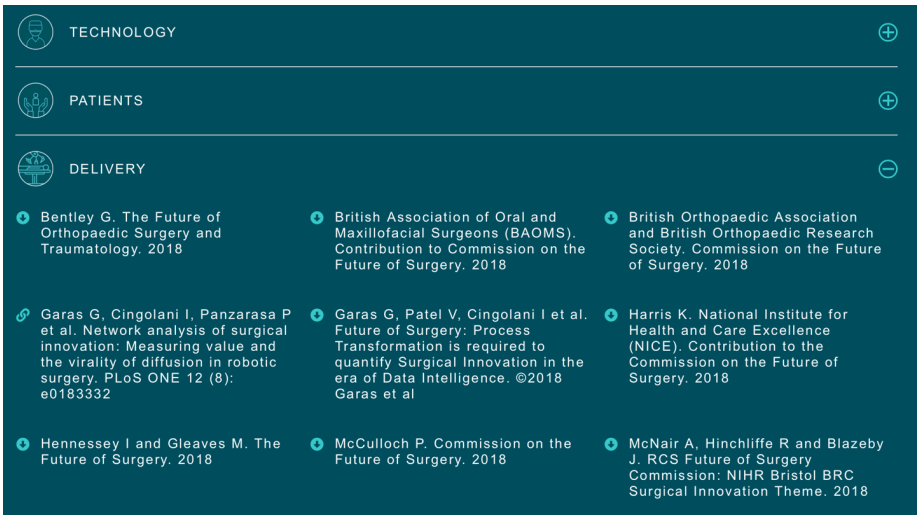
ORIGINAL ARTICLE

Surgical Innovation in the Era of Global Surgery

A Network Analysis

George Garas, MD, FRCS,† Isabella Cingolani, PhD,‡ Vanash Patel, MD, PhD, FRCS,**
Pietro Panzarasa, PhD,§ Derek Alderson, MD, FRCS,†¶
Ara Darzi, MD, KBE, FRCS,† and Thanos Athanasiou, MD, PhD, MBA, FRCS*†*

Garas G, Patel V, Cingolani I, Panzarasa P, Ashrafian H, Darzi A, Athanasiou T. Future of Surgery: Process Transformation is required to quantify Surgical Innovation in the era of Data Intelligence. The Royal College of Surgeons of England. *Future of Surgery*. London, United Kingdom; December 2018.



The Royal College of Surgeons of England. *Surgical Research Report 2019-2020.* London, United Kingdom; May 2019.

Enhancing healthcare value though surgical innovation metrics

Georgios Garas

FELLOWSHIP/SPONSOR:
The Dr Shapurji H Modi Memorial Research Fellowship

SUPERVISORS:
Professor Thanos Athanasiou and Professor Lord Ara Darzi

SITE OF WORK:
St Mary's Hospital, Imperial College London

George performing transoral robotic surgery (TORS) as part of his surgical innovation studies.

The NHS is experiencing a radical transformation driven by pressures to reduce costs. An ageing population has ever increasing healthcare needs with patient expectations growing. Measuring healthcare innovation value is more important than ever.

Currently, the rate of innovation occurring in surgery is beyond our

and using network analysis developed the first surgical innovation metrics: the innovation index and structural virality. These were subsequently validated using big data from the real world that exceeded seven million hospital stays per year (NIS®). These novel ideas and findings were published in some of the world's most prestigious peer-reviewed scientific journals including *The Lancet*.

Garas G, Markar SR, Malietzis G, Ashrafian H, Hanna GB, Zacharakis E, Jiao LR, Argiris A, Darzi A, Athanasiou T. Induced Bias Due to Crossover Within Randomized Controlled Trials in Surgical Oncology: A Meta-regression Analysis of Minimally Invasive versus Open Surgery for the Treatment of Gastrointestinal Cancer. *Ann Surg Oncol.* 2018;25(1):221-230.

Ann Surg Oncol (2018) 25:221–230
https://doi.org/10.1245/s10434-017-6210-y

Annals of SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

REVIEW ARTICLE – GASTROINTESTINAL ONCOLOGY

Induced Bias Due to Crossover Within Randomized Controlled Trials in Surgical Oncology: A Meta-regression Analysis of Minimally Invasive versus Open Surgery for the Treatment of Gastrointestinal Cancer

George Garas, MD, FRCS[1,2], Sheraz R. Markar, MD, PhD, MRCS[1], George Malietzis, MD, PhD, MRCS[1], Hutan Ashrafian, MD, PhD, MBA, MRCS[1], George B. Hanna, MD, PhD, FRCS[1], Emmanouil Zacharakis, MD, PhD, FRCS[3], Long R. Jiao, MD, FRCS[3], Athanassios Argiris, MD, PhD, FACP[4], Ara Darzi, MD, KBE, FRCS[1,2], and Thanos Athanasiou, MD, PhD, MBA, FRCS[1,2,5]

[1]Surgical Epidemiology Unit, Department of Surgery and Cancer, Imperial College London, St. Mary's Hospital, London, UK; [2]Department of Surgical Research and Innovation, The Royal College of Surgeons of England, London, UK; [3]Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital, London, UK; [4]Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA; [5]Health Technology Assessment Committee, National Institute of Health and Care Excellence, London, UK

Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Network analysis of surgical innovation: Measuring value and virality of diffusion in robotic surgery. *PLoS One*. 2017;12(8):e0183332.

PLOS ONE

RESEARCH ARTICLE

Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery

George Garas[1,2]*, Isabella Cingolani[3], Pietro Panzarasa[4], Ara Darzi[1,2], Thanos Athanasiou[1,2]

1 Surgical Innovation Center, Department of Surgery and Cancer, Imperial College London, St. Mary's Hospital, London, United Kingdom, **2** Department of Surgical Research and Innovation, The Royal College of Surgeons of England, London, United Kingdom, **3** Big Data and Analytical Unit, Imperial College London, St. Mary's Hospital, London, United Kingdom, **4** School of Business and Management, Queen Mary University of London, London, United Kingdom

* g.garas@imperial.ac.uk