Heliyon 6 (2020) e04519

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Evidence-based policy making: determining what is evidence

Peter W.B. Phillips [a], David Castle [b], Stuart J. Smyth [c,*]

[a] Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan, 101 Diefenbaker Place, Saskatoon, Saskatchewan, S7N 5B5, Canada
[b] University of Victoria, PO Box 1700 STN CSC, Victoria, BC, V8W 2Y2, Canada
[c] Department Agricultural and Resource Economics, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan, S7N 5A8, Canada

ARTICLE INFO

Keywords:
Bibliometrics
Delphi
Evidence-based policy
Literature analysis
Social network analysis
Survey articles
Agricultural economics
Agricultural policy
Agricultural technology
Network analysis
Knowledge representation
Information science
Agriculture

ABSTRACT

In the era of evidence-based policy, framing and assessing the core evidence is fundamental to our ability to use research in support of public policy. In a world of almost exponentially expanding scholarly publication, it is becoming harder to define what is known. This article reviews the basic theories of knowledge, the context for sorting through and summarizing that knowledge and a number of options available, and used, to assemble the knowledge base for research and policy work. The authors undertook a summative process in the domain of biotechnology, agriculture and development and offer insights into the comparative methods and their impacts on the outcome. A population sample of 421 articles was gathered. Four methods—expert Delphi, citation analysis, social network analysis and peer evaluation—were used to select the 51 pieces for inclusion and analysis in the core literature. That analysis shows that each process delivered a different set of evidence. The potential for bias in knowledge assessment can challenge policy makers in their process of reviewing evidence that rationalizes policy.

1. Introduction

Isaac Newton wrote that 'If I have seen farther it is by standing on the shoulders of giants.' Today that is a real challenge, as millions of researchers are adding millions of new findings to our stock of knowledge annually. Knowing where to stand and how to find, assess and use the voluminous mass of knowledge is a key concern to scientists and policymakers alike.

Any research or policy process should begin with an assessment of what is known about a problem or topic, in order to provide a base for further investigation. Given the profusion of knowledge and venues for codifying and storing that knowledge, researchers and scholars are forced to be selective about how they survey what is known. The advent of the Thomson Reuter World of Science (WOS) and Google Scholar, to name the most prominent, offer tools to cast one's net well beyond the literature anyone might readily know or have read. But the ability to more readily identify the volume and diversity of candidate literature in-and-of-itself does not make research any easier—the rise in the absolute number of potentially-relevant articles and reports makes surveying the base for new research more problematic. The algorithms driving the search engines in the WOS and Google Scholar both offer indicators of relevance and impact and allow one to refine a search to distinguish literature based on a wide range of quantitative and qualitative factors (e.g. keywords, provenance, timeliness, impact factors, citations), but that often still leads to a large body of literature to digest. In response, there is a growing demand for, and supply of, pre-sorted and structured literature assessments, both in traditional peer-reviewed journal articles and more recently in structured collections.

All of this matters profoundly to our effort to build and deliver evidence-based policy. Governance systems around the world—including, among others, research management systems, human and environmental risk regulators, policy systems, competition bureaus and trade regimes—have all adopted norms that require evidence in the form of objective data and subjective analysis to support and audit decisions. As with any assessment effort, there is a great opportunity for bias to creep in and affect the results. The assessment of scholarly literature is not immune to this.

This article reviews relevant bits of the theory of knowledge generation, examines the context for researchers today and reports on our efforts to define a core set of literature and ideas about the use of biotechnology in agriculture and its impact on development—a relatively contemporary and still contested area of knowledge. The authors of this

* Corresponding author.
E-mail address: stuart.smyth@usask.ca (S.J. Smyth).

https://doi.org/10.1016/j.heliyon.2020.e04519
Received 24 February 2020; Received in revised form 5 June 2020; Accepted 17 July 2020

paper at that time had each spent more than 15 years contributing to that literature and engaging with the broader research community to define the impacts of this emerging technology on the research system, on users and on the broader economy and society. We here report on the multi-method approach to consolidating the core literature. We set out the challenges of dealing in a new and rapidly advancing knowledge area and show the results of our efforts to identify a number of themes and seminal articles that helped us to understand how the knowledge in this domain has emerged and consolidated.

One can trace the scientific method to its roots in the 1660s (Shapin and Schaffer, 1985), when Robert Boyle, the first president of the Royal Society, asserted that matters of science should be the outcome repeated 'witnessed' empirical experimentation transmitted in a formal essay, written in a plain, functional style. The minimum criteria for the essay was that it would provide adequate information both to allow replication and to delimit the specific findings and implications of the experiment. Those essays are the focus of our investigation.

1.1. The scholarly literature

The actual number of journals and articles published can be measured with Ulrich's Periodicals Directory, named after Carolyn Farquhar Ulrich, who originated the list as Head of Periodicals at the New York Public Library. The 2019 Directory offers information on more than 246,000 serials publications online, in print and on microform. Björk et al. (2009) estimated that in 2006 a subset of 23,750 journals form Ulrich's list collectively published approximately 1,350,000 articles. Extrapolating using the observed 2.5 per cent per year growth between 2003 and 2013 (National Science Board, 2014), there may have been more 1.6 million new articles published last year and an accumulated stock of more than 50 million articles.

Given the expansive nature of knowledge generation, it only makes sense that most literatures are quite diverse and distributed. One way to manage this is though the synthesis and integration of work in survey journal articles, handbooks and research collections, where the personal knowledge, preferences, motives and style of the authors and editors are fundamentally important. It would be impossible in almost any field to truly review, assess and order the knowledge of every published work. Selection and culling is inevitable.

Editors of these types of synthesis pieces generally offer good rationales for why they pick the articles and chapters that comprise their collection, but the resulting literatures clearly reflect personal preferences and often seem quite idiosyncratic. Some collections reviewed in preparation for this article make their contribution to the literature by articulating the underlying logic of a field by following the developmental trajectory of ideas and methods through the years. Other collections take the 'all-star' approach by including predominantly high-impact and highly-cited articles from their field. This approach works reasonably well for those fields where the literature is mature and exhibits a relatively high degree of convergence on the core pieces and the key in-fill articles, but is more challenging when a field is new and still evolving. Our work sought to explore alternate approaches to consolidating a literature.

1.2. Constructing a literature base for agriculture, biotechnology and development

The authors of this article were commissioned by Edward Elgar to produce both a handbook and a companion research collection on agriculture, biotechnology and development (Smyth et al., 2014; Phillips et al., 2015). The handbook ultimately contained 51 commissioned chapters by 85 scholars from around the world. While the individual chapters represented the contemporary state of knowledge, they were unambiguously the creations of their authors, as some explored their topic through theory, some through primary research and some through survey and assessment of what others had previously reported. In contrast, the research collection is explicitly a collection of key articles and publications on this topic. The pool of literature for the research collection is necessarily new—transgenic technologies were first developed and reduced to practice in the early to mid-1980s, with the first commercial application of the technology occurring in 1994. Given the focus on understanding the impact of the technology on agriculture and development, we had less than 20 years of real-world experience to draw upon. The earliest appropriate candidate articles were written in the past 25 years, and most of the more important work emerged in the past decade. Even so, two and half decades of scholarship on agricultural biotechnology provides a deep repository of scholarship structured by themes and trends in patterns resembling fields of research with deeper histories.

Some argue that agricultural biotechnology, especially transgenic modification and the resulting genetically modified (GM) plant varieties may be the most important scientific and technological development in agriculture in the modern era. In the first 20 years of use, the technology was adapted, adopted and diffused into 13 commercial crops, introduced in 30 countries and cultivated by more than 18 million farmers on all continents and on more than 2.1 billion hectares of land (James, 2017). Many assert that this is the fastest and most comprehensive introduction of a new technology and its related products in the history of agricultural development. Since the rise of molecular genetics and recombinant DNA techniques in the 1970s, there has been a steady translation of scientific knowledge into practical applications. These incremental and enabling steps included techniques of gene identification and isolation, identification of markers and promoters, development of DNA constructs, amplification through the polymerase chain reaction, gene splicing and micro propagation and plant tissue culturing. These developments, while important for the scientific context, are only the background to the story of adaptation and adoption in global agriculture.

Uptake of these technologies in agriculture started in the early 1980s, when a small number of public sector scientists, entrepreneurial start-ups and multinational agro-chemical companies began to assess the potential for this technology in farming. It was generally agreed that biotechnology could help scientists and breeders overcome some of the previously insurmountable hurdles in plant breeding. Throughout the 1980s, scientists tailored biotechnology techniques to fit agricultural practices and sought applications that could be scaled to industrial agricultural systems. At the 1983 Miami Winter Symposia on the Molecular Genetics of Plants, three of the four world-leading researchers unveiled their accomplishments related to transgenic tobacco (Fraley et al., 1983; Framond et al., 1983; Schell et al., 1983), thereby paving the way for gene technology to be translated to a wider array of agricultural applications. GM tobacco was first produced on limited acres in China in 1993, quickly followed by a commercial GM tomato cultivar in the US in 1994 and then widespread introduction of GM varieties of canola, corn, soybeans and cotton beginning in 1995. Those four large-area crop types alone accounted for more than 99% of the area planted to GM crops in the first 20 years of commercial use of the technology.

In spite of the promise and enthusiasm (some say hype), the reality has been disappointing to many. While GM technologies were touted as a key tool for breeding a wide range of crops, only four main modified commodity crops have reached global proportions. Moreover, GM crops have been unevenly adopted throughout the world. The net effect is that, two decades on, the technology is more or less in a steady state, rather than generating major technological transformation. The focus of the research collection was to explain why there was a patchwork of adoption and a technological plateau.

2. Method

When we started the process, we already had a good sense of the scale, scope and key content of the literature. Even so, there were thousands of articles to sort and select from. Early on in the process, the three of us agreed that we needed some method that would reduce our

personal biases and would enable us to more fully capture the nuances of the literature. With the insights of more than 85 authors and collaborators from 18 countries from our Handbook, we had substantive summaries of the state of the models, methods and metrics related to the actors, issues and impacts of biotechnology on the global crops-based agricultural system. This provided a useful compendium of references that could fit in this research collection. Given our sense of the general absence of method in most of the research collections we had surveyed, we adopted a methodical and structured approach to assemble our collection in order to avoid an idiosyncratic and potentially unfocused sampling.

We decided to use an expert opinion process coupled with citation and social network analyses to structure our choices. Three rounds of a Delphi exercise were administrated in June–December 2012, using SurveyMonkey® as the survey tool. This process was evaluated by the University of Saskatchewan Behavioural Research Ethics Board and approved on 29 May 2012 (Beh-REB 09-256).

In the first round of the Delphi participants were offered a set of themes and asked to both identify experts that could be invited to participate and additional themes for consideration (Appendix 1). The respondents identified 75 new experts that were added to the exercise. In reviewing the list, the editors identified an additional 23 experts that had been missed; they were also added. Of the 68 active respondents, 48 gave advice on the identification of themes to be covered in the volume. A qualitative assessment of the 100 ideas proffered identified five main frequently repeated themes. The largest group (~40%) broadly related to impact evaluations, ranging from product-specific, farm-level assessments to ecosystem and global effects, variously focusing on producers, consumers and the broader socio-economic system. The second largest group (~30%) related to the political economy of regulation and governance, including processes related to product assessments, the politics of biotechnology and the regulation of international trade. The third group (~20%) captured the economic, commercial and social impacts of intellectual property rights and the public and private strategies governments and firms use, including public-private partnerships. Two other minor themes emerged: a desire to see the philosophical and ethical dimension of biotechnology examined; and the potential to use the volume to advance the use of foresight tools and to identify its lessons.

We used the advice on themes to construct the pool of candidate articles to be considered for inclusion in the volume. After the first survey, the editors produced a list of 19 potential topics and themes for the volume. The second survey process asked participants to nominate key articles and to identify the thematic focus of the articles. Round two was explicitly about identifying the source population for subsequent voting, yielding 190 unique articles. An additional 234 articles were added to the list by the editors using a snow-ball technique for identifying literature cited in the nominations.

In November 2012 we began the third and final round of the Delphi exercise to solicit votes on the articles most important to biotechnology, agriculture and development. We circulated a list of 421 articles to 283 experts and received 1,783 votes from 52 respondents. Forty-five articles received no votes. The remaining 366 articles exhibited a power log effect, with an average of 4.8 votes (max = 23; min = 1; mode = 1); the top 35% of the articles gained 71% of the votes. After the voting was complete, the top 51 articles, each having received at least nine votes (i.e. about 17% support from our respondents), were identified and selected for the volume.

As discussed below in our assessment of the methodology, after the fact we received some passionate pleas for other literature to be in the voting, signalling that regardless of the method, no volume of this sort is going to be the absolute last word on any subject. Time always elapses between expert consultation – in this case our Delphi – and the analysis and publication of the findings of the consultation. In this respect these methods offer systematic approaches to periodic, retrospective insights about the state of a field of research. What they do not purport to do is give contemporaneous accounts of trends that are, or ought to be, perpetuated. Following the identification of the articles, we undertook structural and content analysis of the collection to isolate the key themes and insights.

2.1. A comparative analysis of the methods

Before discussing the results of our Delphi and the implications for the content of the literature selected, it is important to compare and contrast our chosen method with three other potential ways of assembling literature for our research collection: editor/peer judgment; citation analysis; and data mining using social network analysis. This section assesses and compares the output of each of those processes against the Delphi results (Table 1).

Our Delphi approach identified 51 articles equal to approximately the top 12 per cent of the candidate articles. The average article was 7.9 years old and had 94 citations (or about 10.2 per year) and the articles ranged from 1 page to over 80 pages. The articles were presented in the Delphi exercise ordered alphabetically based on the last name of the first author. We found that the median article in our selection was listed at about number 200 in the voting list, virtually at mid-population, which suggests that there is no structural bias to the selection (in some instances, such as voting, there is evidence that respondents may be lazy and pick from the top half of the sample). The articles in this collection have been cited between zero and 525 times. A total of 96 authors were involved in these works. It is this body of work that we use to assess the effect of method on the outcome.

The most common approach for surveys or research collections is for a single editor or small team of editors to identify a specific topic they want to help define or shape through the compilation of a selection of literature. These editorial efforts are often quite focused on a narrow or limited range of approaches. In this sense, the article selection process often ends up being a self-limiting exercise, as the absolute pool of possible articles is delimited by the experience of the editors. In our case, we undertook a voting process to see how we as subject experts might choose if unconstrained by other rules. The challenge we faced was that we bring widely different perspectives and backgrounds to the task: one of us is a life-science innovation expert with an applied philosophy background; a second is a classically-trained international political economist with background in government; and the third is a former business consultant with an applied interdisciplinary research program. When we looked at the 421 articles and did our own separate voting (we did not vote in the Delphi process), we individually indicated interest in 93 articles for the collection. But we had common interest on only 19 of the articles (i.e. two or more editors supporting inclusion). We could have done a few extra rounds of voting to refine the sample into 'must haves' and the rest, but decided this first vote was appropriately indicative of the diversity of our views. Figure 1 shows the array of consensus that emerged in a first round of voting. In the interests of full disclosure, six of articles identified with a single vote had at least one of the co-editors as an author; two of these ended up in the final collection of 51 articles based on the Delphi vote and one would have been in a collection of highly cited articles. In the larger set of 93, the average age of the articles was about 9.5 years, the average number of citations was 127 (or about 10 per year) while the subset of 19 was more recent (average 7.1 years old) and had a lower average citation rate (53 citations or about 5.1 per year). Ultimately, the larger sample included about 57% of the articles that were identified by the Delphi and included in the final collection while the majority picks only included about 20% of the final collection.

A second common approach to assembling a research collection is to define a subject area and use the combined judgment of all of the other authors in the field, ultimately selecting articles based on the number of citations. We did a citation analysis of all of the articles and discovered that 55 articles had been cited 100 or more times, which is one measure of enduring impact. In that context, we found those articles were on average 13.6 years old and had an average of 554 citations (or 37.2 per year). Interestingly, only about 37 per cent of the most highly cited

Table 1. Comparative methods and the results.

	Source pop	Top voted articles in Delphi	Editors majority picks	Editors sample	>100 cites	SNA methodology
Articles N =	421	51	19	93	55	23
Average years since publication	8.6	7.9	7.1	9.5	13.6	9.2
Average cites per article	94.7	94	53	127	554	97
min cites	0	0	0	0	100	2
max cites	9,868	525	156	2,808	9,868	525
Average cites per year	7.6	10.2	5.1	10	37.2	10.1
Average number of articles in sample	211	200	134	194	223	175
Total votes cast and share	–	652 (36.6%)	182 (10.2%)	642 (36%)	359 (20.1%)	152 (8.5%)
% of collection of 51	–	100%	20%	57%	37%	12%

articles were included in our final collection. The main difference is that many of the highly cited articles in our source population were relevant to the agricultural biotechnology story but were more about the context of technology development than about the impacts of adoption. Our voters stayed true to the task and overwhelmingly selected articles that were directly related to specific applications of biotechnology in agriculture rather than those that simply investigated potential ways of examining technological change.

A third way to select articles is to mine the data from a subject area. There would be two ways to do this. The first would be to start from key words and concepts and build a dataset by mining databases of articles (e.g. Thomson Reuters World of Science, AgEcon-Search, Google Scholar and RePEc). A second approach would be start with a dataset like our 421 articles. One could then map and analyse the scholarly network encompassed by these works to identify those individuals who are at the centre of knowledge discovery and to choose the articles based on the mostly prominently placed individuals. This method is biased away from sole-authored works because network analysis depends on co-authorship linkages, but has the potential to identify critical articles in which leading actors strengthen, deepen or broaden scholarly investigations. To assess this approach, we constructed the social network for the authors related to the 421 articles in our source population. This collection of 621 authors formed a loosely connected network (density <0.6%) that was anchored on a subset of 48 relatively highly-connected authors (density >8%). These authors exhibited either high network connectivity (i.e. degree centrality ≥3 SD higher than the population average) or significant reach (i.e. betweenness centrality ≥2 SD higher than the population average). We tried a number of methods of selecting articles using this data. When we picked articles with three or more highly connected authors, we identified 34 articles, with an average of about 8.8 years, an average of 8.4 cites per year but only six of those articles were similar to those voted into the collection. We landed on picking the top-cited article of each highly ranked author, which delivered a core collection of 23 articles, only six which made it into the voted collection.

Looking at all the methods together, we can say with some confidence that the Delphi process of expert opinion formation delivered a qualitatively different and in many ways more robust collection of articles that capture the state of the art.

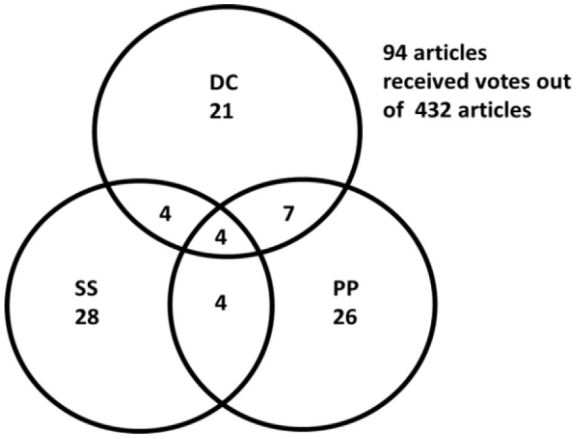

Figure 1. The editors' picks.

2.2. Structural and content analysis

One structural way to assess the body of work selected through the Delphi is as a knowledge network. To understand it better, we undertook a social network analysis (SNA) of the articles both individually and as a part of the larger population of articles. The density of the system (measuring the interlocking co-authorship of the works) was 6.2 per cent, which while modest, is large enough to suggest that there are some network effects among the actors.

A range of centrality measures—i.e., degree, betweenness, and Eigenvector—were employed through UCINET and Netdraw software to extract the key authors from the total of 96 authors engaged in knowledge production of the 51 high impact journal articles/book chapters. We undertook two specific analyses. The primary review included all of the 96 authors for the 51 papers, which yielded a tightly-linked group of 18 authors who co-authored a single highly valued and cited article in *Science* (Hutchison et al., 2010) as highly centrally placed; only four other authors emerged as relatively centrally placed (Figure 2). While this provides one view of the literature, it is somewhat misleading as that group of 18 had no links with the rest of the 78 authors; in contrast there were extensive linkages among the other 78 that warranted further investigation. Thus, a second analysis replaced the group of 18 with a single proxy (Hutchison), which had the effect of moving that cluster to the status of an isolate. The remaining population revealed a more nuanced set of leaders and connectors.

Interestingly, there appears to be a series of clusters of scholars who are intensively engaged in the development of this knowledge base (see Figure 3). The single largest cluster is largely based in the US, involving a range of scholars attached to or trained at UC Berkeley anchored on David Zilberman, professor and former head of Agricultural and Resource Economics, and linked through Greg Traxler, formerly of Auburn and CIMMYT and now at Washington State University, and Falck-Zepeda, his PhD student and subsequently with ISNAR and now IFPRI. This group has done much of the heavily cited work on adaptation, adoption and the alternate strategies for optimizing impacts. A second network of US and Asian scholars is intensively interested in uptake and market effects in Asia, and any related effects on supply chains. A number of smaller clusters emerge. A cluster of US-based scholars is focused on the gains to research while a team in Canada has concentrated on trade and market effects. Beyond that, there are a few triads and many dyads of authors. While not directly obvious, most of these subnetworks are actually linked at one remove, as many have undertaken other work that extends the network. One significant link between these authors is the International Consortium on Agricultural and Bioeconomy Research

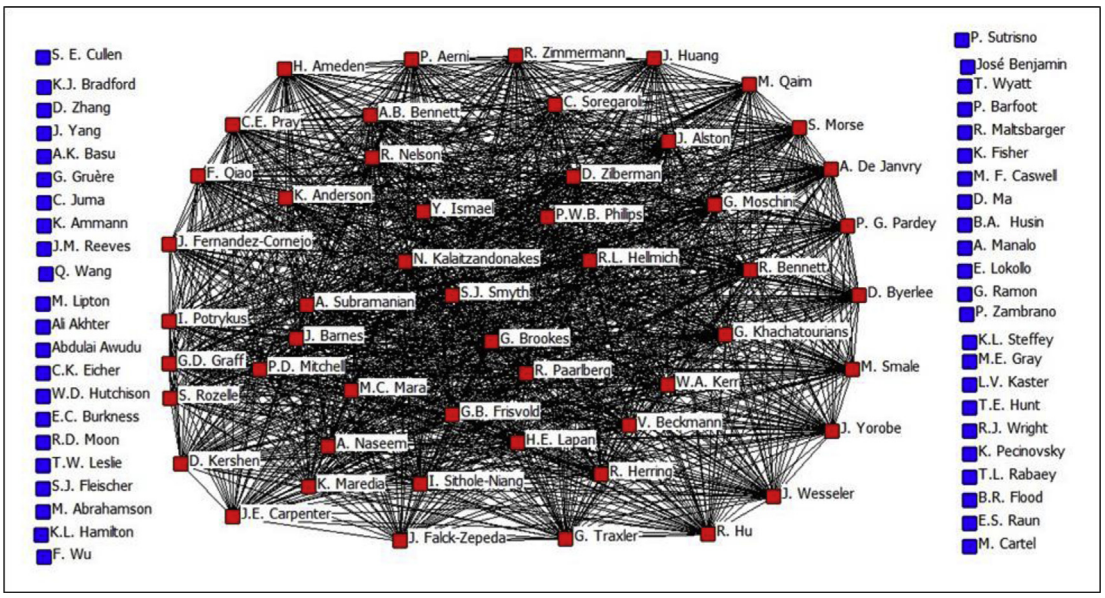

Figure 2. Co-authorship ties among the authors of the Collection of 51.

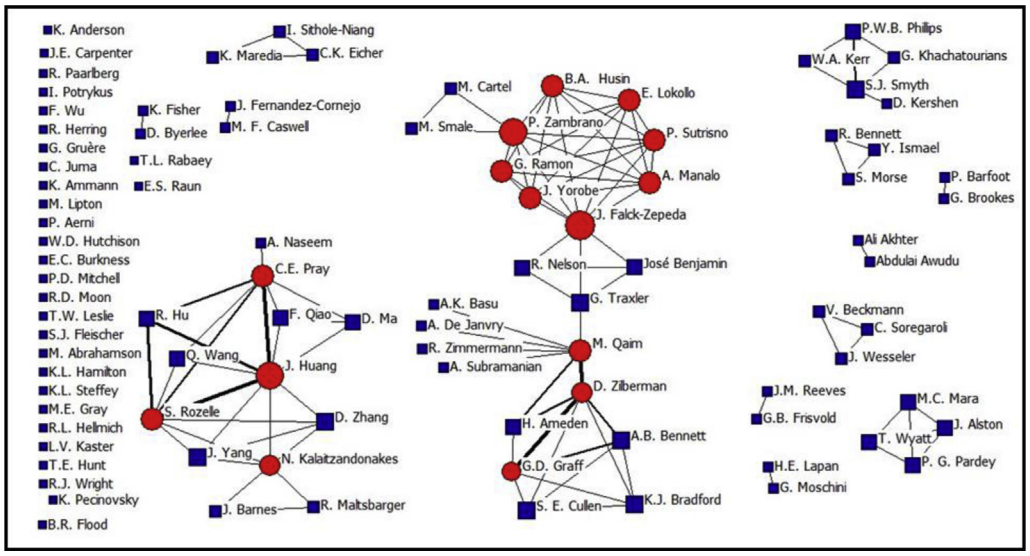

Figure 3. The network of 96 authors from the collection of 51. *Red signifies > 2STD and blue signifies <2STD; node size based on degree centrality scores and tie width depicts tie strength. **Hutchison et al. is incuded as isolate.

Table 2. Analysis of content of different methods.

	All picked articles	Delphi vote	Editor Majority picks	Editor Sample	Highly cited	SNA
N =	51	23	156	19	93	55
Economic impact	43%	***55%***	32%	38%	42%	***59%***
LDCs and development	42%	***51%***	37%	37%	38%	***64%***
Policy analysis	40%	39%	***47%***	***49%***	33%	36%
Trade and markets	24%	12%	***26%***	***30%***	20%	***32%***
IPRs	16%	***20%***	11%	***20%***	7%	9%
Consumer response	14%	2%	0%	13%	***20%***	9%
Environment	12%	12%	***16%***	10%	13%	9%
Scientific Studies	6%	4%	5%	5%	***13%***	0%

Note: Bold and italicized numbers represent over representations of subject relative to all picked articles.

(ICABR), an international research network established in 1996 by Professors Bob Evenson (Yale), Vittorio Santaniello (Tor Vergta, Italy) and David Zilberman (UC Berkeley). Since 1997 this group has hosted at least one international symposium or conference annually (either in Rome or Ravello in Italy or internationally), which at one time or another has attracted as presenter at least one of the authors of 47 of the 51 articles in the collection.

In order to understand the impact of different methods on content, we reduced the active population to the 156 articles that were chosen by one of the five methods out of the 423 source population and then coded them into eight major theme areas. Our goal was to determine if and how the various methods might amplify some topics and downgrade others (Table 2). We found that the full collection of 156 articles strongly represented the literature on economic impact, development (especially in LDCs) and policy issues, but trade and markets, IPRs, consumer response, environ.

2.3. Impact of method on content

Interestingly, while the four methods identified 156 candidate articles, only four articles met the selection criteria for all four methods. Three explicitly addressed economic impacts, with a focus on LDCs, and one addressed questions of market access in the face of differentiated tolerances for GM seeds. Another five articles met three of the four selection criteria (only one was produced by a highly-linked author as defined in our SNA method). Once again the focus was on impacts, especially for LDCs, but now including inclusion of an article about IPRs. Each of these nine articles had an empirical base with a policy commentary.

The Delphi method offered a set of 51 articles that significantly overrepresented the economic impact and LDC/development literature. Interestingly, it also elevated the literature on IPRs, which in the eyes of many are inextricably linked to the question of who will benefit from biotechnology. Surprisingly, the Delphi included only one article that focused primarily on the consumer response. After some consideration, the authors realized that in most cases the candidate articles on consumer response focused on first world problems that were not replicated in developing countries. Given the focus on the role of biotechnology in agriculture and development, this explained the underweighting.

The editors using their own judgement identified a subset of 93 articles and converged on a core set of 19 pieces that differentially overemphasized policy analysis, trade and markets and IPRs, which undoubtedly reflected their professional expertise.

The quick and dirty method of choosing only highly cited articles (>100 Google Scholar Citations) produced a list of 55 articles which over-emphasized the consumer response and scientific matters, at the expense of policy, trade and development.

The SNA method revealed that the scholarly communities are differentially more connected and led by key scholars in the economics area, especially among those working on assessing the economic impacts, the implications for development and the role of trade and markets. As noted above, this may be partly because of the efforts of some to create a new epistemic community based at the ICABR in Italy.

The actual process of developing a literature analysis tells us as much about the nature of the field as the articles themselves. The high degree of willingness to engage among the scholarly community was a pleasant surprise. Equally surprising was what they thought was important. The laser-like focus of our respondents on the development challenge implicit in agricultural biotechnology was intriguing. The spike in food prices after 2010 and increasing anxiety among governments and policy makers about the global ability to bridge the gap between the inexorable rise in demand for protein and fibre to meet our food, feed and fuel needs appears to have refocused interest on the role of new technology. The resulting choice of articles that focus on intellectual property (IP), industrial structure, policy and economic impact narrows the focus on a few measurable and manageable policy and institutional settings.

We, and some of our respondents, were somewhat surprised about what did not get ranked more highly. The leading articles related to consumer and citizen theory were offered as candidate articles but were generally ranked as less important. There has undoubtedly been lots of innovation in the field, with Lancaster's (1966) 'new' theory—that consumers search for and consume attributes of products, and not simply the product—providing the foundation for a wide array of willingness-to-pay and willingness-to-avoid analyses (e.g. Moon et al., 2007), contingent valuations (e.g. Moon and Balasubramanian, 2003) and experimental auctions (e.g. Huffman et al., 2003) to discern prospective plans and to quantify the impacts of new technologies on consumer welfare. Lusk et al. (2005) and Dannenberg (2009) have produced valuable and insightful meta-analyses of that body of work. Our best guess is that while these are important and interesting innovations, they are viewed by most scholars engaged in development as a nice refinement, but not necessarily appropriate for driving global policy related to food security. In effect, the optimal matching of consumer preferences about provenance to product attributes is a luxury that has little relevance to the majority of the world's consumers.

Second, as we engaged with the publisher of our collection we learned more about the economics and politics of scholarly publishing itself. Publishers of research collections need to negotiate rights to republish journal articles. As with all publishing, there is a budget constraint. When we got into the details of the negotiating process we learned that research collections that seek to republish may be seriously constrained by this. In our case, the publisher informed us that they were unable to negotiate terms with all the journals and still produce the volume within the budget. The initial indication was that the ten articles published by Nature Publishing Group (esp. *Nature*) and the American Association for the Advancement of Science (esp. *Science*) would be omitted as they fall into category of 'tried to negotiate with them but failed'. After some to-ing and fro-ing, the *Science* papers were included, but five papers from *Nature* were removed due to cost. One paper, co-authored by two of the editors, was allowed to be reproduced free of charge under copyright provisions. After further negotiation, a popular article from the *International Journal of Technology and Globalization* was also omitted on the grounds of cost. In the end, we were only able to include 43 of the 51 selected articles. We were encouraged by the publisher to select another set of articles, which in our case would have compromised our method. We did consider going one level deeper in our pool of articles (e.g. those with 8 or more votes in the Delphi) but upon investigation discovered that (a) the number of articles was too large and (b) the diversity of topics widened significantly at that level. In the end, we compromised by citing the excluded articles in the introductory essay but not including them in the collection. Ironically, all but one of the excluded articles is otherwise available on the Worldwide Web in an open source website or other database.

2.4. Implications

Without a doubt, more knowledge is being published today than at any previous point, but the question is whether the evidence generates any greater insight? The expansion of online journals, especially the predatory journals that will publish anything submitted as long as the authors pay a fee, certainly raise questions about the quality of evidence. Careful selection is becoming more challenging. From the standpoint of the production, accumulation, and organization of knowledge, those 'in the trade' have tacit and explicit knowledge of the literature, its overall structure as a field of inquiry and a sense of the dynamics of trending and declining themes. As discussed earlier, the epistemological and social study of knowledge in these systems has its own field of interest for centuries. The challenge is that some of the conventional norms about knowledge production, distribution and access have been disrupted by the sheer scale and increasing diversity in publication practices.

There are two key implications resulting from our methodological analysis that leads to this observation. First, the vast explosion of

journals, legitimate and predatory alike, can and often does send mixed and confusing signals to policy makers, or even generate a sense of bewilderment about how to start to engage an unfamiliar field of research. Our review of the emergence of the current approach to academic publishing suggests that peer review is vital to scientific progress. The emergence of predatory journals is in some ways returning us to the world of alchemy that dominated before the scientific revolution. Many of these new pay-to-publish reviews eschew peer review; many of the resulting publications simply muddy the waters, contributing poor theory, method and evidence that makes finding solid evidence more difficulty. Differentiating between predatory and credible journals can be challenging. Both may require some payments to facilitate publishing, but only the established authorities sustain academic standards through rigorous peer review; once published, however, it is difficult to select credible from speculative work. While the impacts vary from country to country there is some evidence to suggest the impact is greater in developing countries. Lack of local capacity to consolidate and assess the validity and relevance of published work can have significant policy impacts (Colombo Page, 2014; European Trade Union Institute, 2019). Added to the question about scientific integrity and quality, the ease of transmission through the internet has exposed everyone to junk science that can distort policy.

Second, there is a serious risk that current trends in academic publishing may create an accessibility bias if what is available is taken to be what is known. The movement by mainstream, prestigious journals to charge open access fees for articles restricts policy makers' access to leading scientific discoveries. Journals now charge academics an open access fee that can be as high as US$6,000 to make their research free for public access. Many cannot afford this fee and find their work sequestered behind a pay-wall that impedes uptake and use. While this definitely impedes the public's access to research, it also often limits policymakers' access to cutting edge research. Few governments directly subscribe to journals, especially the most expensive (often most prestigious) journals. This is especially the case in developing countries where the allocation of fiscal resources is frequently more constrained than with industrial economy governments; moreover, most of the more important journals are published in English, which can limit their access. However, developing country governments are the ones that might benefit most from access to the information in the emerging literature.

3. Conclusion

Rounding up and making sense of what we know is undeniably an arduous task for any field of research. Figuring out what the theoretical, methodological or empirical issue is, what the candidate literature is, and how to select the most important articles from that population is difficult enough without the cross-cutting constraint of the for-profit academic publishing enterprise and predatory actors. Having undertaken this process, the authors are now much more sensitive and critical of scholarly surveys and much more aware and concerned about the potential bias that undoubtedly challenges regulators and policy-makers as they attempt to engage with evidence in support of policy making and policy assessment.

To return to our opening question of what is evidence, we are able to conclude that evidence is the result of robust science-based methodologies, rigorous and statistical data analysis and results that are supported by data. Outliers do, from time to time, get published, but by relying on these three core aspects, research can be conducted by other scientists to test whether the original finding are repeatable. Repeatability of findings is a fundamental component of evidence that possesses significant statistical confidence. Evidence that builds upon and contributes to existing stocks of knowledge serves as the basis for policy-makers, whereas one-off publications may merit further investigation, they rarely impact existing policy.

Given the importance of science-based evidence for policy-makers, our research highlights the importance of engaging multiple methodologies in the assessment of data in order to reduce unintended bias inherent in the methodologies. Employing several methodologies and comparing the results can reveal core knowledge that can provide policy-makers with high confidence for policy development. Where different methodologies provide a common core set of results, policy-makers should look seriously at the message. Where the results are more defuse, caution should be the order.

This exercise additionally highlights the need for a global evidence network for policy-makers. While policy-makers in industrial countries may have the luxury of engaging multiple methodologies and accessing all the primary evidence, those in developing countries often lack such capacity. Trust in science has to be global to be effective. Providing policy-makers in developing countries with full access to evidence from all jurisdictions should help to increase confidence in regulatory and policy oversight and reduce the cost and delays in commercialization of new technologies.

Declarations

Author contribution statement

Stuart Smyth, Peter Phillips, David Castle: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2020.e04519.

References

Björk, B.-C., Roos, A., Lauri, M., 2009. Scientific journal publishing: yearly volume and open access availability. Inf. Res. 14 (1), 391.

Colombo Page, 2014. Sri Lanka Bans Sale of Glyphosate Weedicide Responsible for Kidney Disease accessed. http://www.colombopage.com/archive_14A/Mar12_1394634963CH.php. (Accessed 19 October 2019).

Dannenberg, A., 2009. The dispersion and development of consumer preferences for genetically modified food: a meta-analysis. Ecol. Econ. 68 (8-9), 2182–2192.

European Trade Union Institute, 2019. Vietnam Bans Glyphosate accessed. https://www.etui.org/About-Etui/News/Vietnam-bans-glyphosate. (Accessed 19 October 2019).

Fraley, R.T., Rogers, S.G., Horsch, R.B., 1983. Use of a chimeric gene to confer antibiotic resistance to plant cells. In: Downey, K., Voellmy, R.W., Ahmad, F., Schultz, J. (Eds.), Advances in Gene Technology: Molecular Genetics of Plants and Animals. Academic Press, New York, pp. 211–221.

Framond, A.J., Bevan, M.W., Barton, K.A., Flavell, R., Chilton, M.D., 1983. Mini-ti plasmid and a chimeric gene construct: new approaches to plant gene vector construction. In: Downey, K., Voellmy, R.W., Ahmad, F., Schultz, J. (Eds.), Advances in Gene Technology: Molecular Genetics of Plants and Animals. Academic Press, New York, pp. 159–170.

Huffman, W.E., Shogren, J.F., Rousu, M., Tegene, A., 2003. Consumer willingness to pay for genetically modified food labels in a market with diverse information: evidence from experimental auctions. J. Agric. Resour. Econ. 28, 481–502.

Hutchison, W.D., Burkness, E.C., Mitchell, P.D., Moon, R.D., Leslie, T.W., Fleischer, S.J., Abrahamson, M., Hamilton, K.L., Steffey, K.L., Gray, M.E., Hellmich, R.L., Kaster, L.V., Hunt, T.E., Wright, R.J., Pecinovsky, K., Rabaey, T.L., Flood, B.R., Raun, E.S., 2010. Area wide suppression of European Corn Borer with Bt maize reaps savings to non-Bt maize growers. Science 330, 222–225.

James, C., 2017. Global status of commercialized biotech/GM crops: 2016. ISAAA Brief 52 accessed. http://www.isaaa.org/resources/publications/briefs/52/pressrelease/default.asp. (Accessed 23 October 2017).

Lancaster, K., 1966. The new approach to consumer theory. J. Polit. Econ. 74 (2), 132–157.

Lusk, J., Jamal, M., Kurlander, K., Roucan, M., Taulman, L., 2005. A meta-analysis of genetically modified food valuation studies. J. Agric. Resour. Econ. 30 (1), 28–44.
Moon, W., Balasubramanian, S., 2003. Is there a market for genetically modified foods in Europe? Contingent valuation of GM and non-GM breakfast cereals in the UK. AgBioforum 6 (3), 128–133.
Moon, W., Balasubramanian, S., Rimal, A., 2007. Willingness-to-Pay (WTP) a premium for non-GM foods versus Willingness-to-Accept (WTA) a discount for GM foods. J. Agric. Resour. Econ. 32 (2), 363–382.
National Science Board, 2014. Science and Engineering Indicators. Academic Research and Development. Chapter 5. http://www.nsf.gov/statistics/seind14/index.cfm/chapter-5/c5h.htm. (Accessed 21 February 2019).
Phillips, P.W.B., Smyth, S.J., Castle, D. (Eds.), 2015. Biotechnology, Agriculture and Development: an Elgar Research Collection. Edward Elgar, Cheltenham, UK.
Schell, J., Van Montagu, M., Holsters, M., Zambryski, P., Joos, H., Inze, D., Herrera-Estrella, L., Depicker, A., de Block, M., Caplan, A., Dhaese, P., Van Haute, E., Hernalsteens, J.-P., de Greve, H., Leemans, J., Deblaere, R., Willmitzer, L., Schroder, J., Otten, L., 1983. Ti plasmids as experimental gene vectors for plants. In: Downey, K., Voellmy, R.W., Ahmad, F., Schultz, J. (Eds.), Advances in Gene Technology: Molecular Genetics of Plants and Animals. Academic Press, New York, pp. 191–220.
Shapin, S., Schaffer, S., 1985. Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life. Princeton University Press, Princeton, NJ.
Smyth, S.J., Phillips, P.W.B., Castle, D. (Eds.), 2014. Handbook on Agriculture, Biotechnology and Development. Edward Elgar, Cheltenham, UK, p. 890.