Scientometrics (2018) 115:351–368
https://doi.org/10.1007/s11192-018-2668-z

Growth of the hepatitis literature over the period 1976–2015: What can the relative priority index teach us?

S. L. Sangam[1] · Uma B. Arali[2] · C. G. Patil[3] · Ronald Rousseau[4,5]

Received: 28 August 2017 / Published online: 12 February 2018

Abstract This study has a double purpose: a fact-finding one and a methodological one. The fact-finding part consists of two elements. First we study the growth of the hepatitis literature in the World and in particular in India over the latest 30 years (1986–2015). Second we determine the priority given to hepatitis studies in a number of leading countries, distributed over five continents. This part compares the periods (1976–1995) and (1996–2015). The methodological part studies the stability of the relative priority or activity index by calculating it with respect to different databases. Data are collected from PubMed, the Web of Science (WoS) and the WoS, restricted to the area Life Sciences Biomedicine, denoted as WoS (LS). Data collection in PubMed is more difficult than in the WoS, leading to possibly approximate results. Attention spent to publishing hepatitis research is measured with the Relative Priority Index (RPI). Initially the rise in the number of Indian publications is slow. Yet, the number of publications on hepatitis has substantially increased over time so that since 2008 India's cumulative number of publications has more than doubled. As to the methodological problem we found that different databases lead to different results, showing that results based on the RPI must be interpreted with

✉ Ronald Rousseau
ronald.rousseau@kuleuven.be

S. L. Sangam
slsangam@gmail.com

Uma B. Arali
umaba1210@gmail.com

C. G. Patil
patil.cg1970@gmail.com

1 Department of Library and Information Science, Karnatak University, Dharwad 580 003, India

2 Government First Grade College, Haliyal 581329, Karnataka, India

3 Department of Botany, Karnatak Science College, Dharwad 580 001, India

4 Facultair Onderzoekscentrum ECOOM, KU Leuven, Naamsestraat 61, 3000 Louvain, Belgium

5 Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

care. We, moreover, draw the reader's attention to hepatitis itself, including information about it and its deadly consequences.

Keywords Hepatitis · India · Bibliometrics · Growth rate · Relative priority index

Introduction

Hepatitis is an inflammation of the liver. The word is derived from the Greek hêpar, meaning "liver", and -its meaning "inflammation". Hepatitis may be temporary (acute) or long term (chronic). Acute hepatitis can sometimes heal on its own, progress to chronic hepatitis, or result in liver failure. There are five forms of viral hepatitis, referred to as types A, B, C, D and E. Types D and E are less common than types A, B and C.

In 2015, hepatitis A occurred in about 114 million people, chronic hepatitis B affected about 343 million people and chronic hepatitis C about 142 million people (GBD2015 2016). Worldwide about 325 million (data from 2015) are carriers of hepatitis B or C virus infections. Each year, 1.75 million people acquire a hepatitis C virus infection. These persons are at risk of a slow progression to severe liver disease and death, unless they receive testing and treatment in time (Global Hepatitis Report, 2017). Each year World Hepatitis Day is observed on July 28. This special day aims to raise global awareness on hepatitis and encourage prevention, diagnosis and treatment. In May 2016 the member states of the World Health Organization (WHO) committed to eliminating viral hepatitis as a public health threat by 2030.

Citing from (Global Hepatitis Report 2017) we note that in 2015, viral hepatitis led to 1.34 million deaths, a death toll on a par with that due to tuberculosis and higher than that due to HIV or malaria. Of these deaths, 96% were the result of complications of chronic hepatitis B (66%) and C (30%) infections, while hepatitis A and hepatitis E accounted for approximately 0.8 and 3.3% of the deaths, respectively. A proportion of the deaths attributed to Hepatitis B Viral infection (in short: HBV infection) may be explained by a superinfection with Hepatitis D Viral infection (HDV in short). Among the long-term complications of HBV and HCV infections, cirrhosis (720,000 deaths) accounts for more deaths than hepatocellular carcinoma (470,000 deaths). Although viral hepatitis affects persons in all countries and from all income levels, those living in low- and middle-income countries account for the highest burden.

India has over 40 million hepatitis B infected patients, second only to China. According to the WHO, 2–5% of the population in India is affected by a hepatitis causing virus. In India, each year 100,000 patients die of viral hepatitis. Apart from viruses, other infections, toxic substances (e.g. alcohol, some medications, illicit drugs), and autoimmune diseases can cause hepatitis. Acute infection may occur with limited or no symptoms, or may include symptoms such as jaundice (yellowing of the skin and eyes), dark urine, extreme fatigue, nausea, vomiting and abdominal pain. As with most diseases a healthy lifestyle, including physical activity and an appropriate diet may prevent or help curing hepatitis.

A short description of the five types of viral hepatitis, taken from the fact sheets of the WHO, is provided in "Appendix A".

A huge literature exists on hepatitis, reaching nowadays more than 10,000 publications a year. In spite of this we are not aware of a recent bibliometric study dealing with this topic. In the present study, scientometric/bibliometric methods are used to study scientific research dealing with the topic 'hepatitis'. A preliminary report related to this article has been presented at the 13th International Conference on Webometrics, Informetrics and Scientometrics (WIS) 7 18th COLLNET meeting, held at the University of Kent (Canterbury Campus), 9–11 July 2017.

Materials and methods

Basic queries and databases

This work studies the scientific literature on hepatitis in the World and in a number of selected countries with special attention to India. Publication data on hepatitis for the period 1976–2015 (40 years) has been collected. In a first step we used the basic query "hepatitis". Our assumption is that adding other query terms, e.g. names of medication specifically aiming at hepatitis, will certainly increase recall but this at the cost of decreasing precision. As this study is aimed at information scientists and not at medical specialists we do not strive for higher detail. In a second step we used the queries "hepatitis X", with X = A, B and C. We performed searches in PubMed, a database maintained by the National Center for Biotechnology Information (NCBI), (available at: https://www.ncbi.nlm.nih.gov/pubmed/) and in the Web of Science (WoS), nowadays part of Clarivate Analytics, to investigate possible differences.

Indicators

Scientific performance is a multidimensional concept, which cannot be measured by a single universal indicator. One may use a number of partial indicators, each representing a different aspect of research performance. Nonetheless, the number of publications in refereed scientific journals constitutes the most important basic indicator of research performance. Careful analysis of scientific output in the form of publications can provide meaningful insights for making inter-institution, inter-field and international comparisons of research interest and performance. In this contribution two main indicators are used: the growth rate and the relative priority index (RPI), also known as activity index, relative specialization index or Balassa index. Indicators are defined as follows.

(a) The growth rate, denoted as $R(t)$, used in our investigation is the so-called logarithmized cumulative annual growth rate, defined as

$$R(t) = \ln[W(t\ +\ 1)/W(t)] = \ln(W(t + 1) - \ln(W(t)), \quad (\log\ \text{base} = e).$$

where $W(t)$ is the cumulative number of publications in year t. We recall that if $W(t) = K{\cdot}e^{at}$ (exponential growth), $R(t)$ is equal to the constant a; while if $W(t) = ct + d$ ($c, d > 0$), i.e. linear growth, $R(t) = \ln(1 + c/(ct + d))$ which tends to zero for increasing t. This growth rate, be it with $\log_{10}$ instead of ln, has been used e.g., in (Amaral et al. 2001; Liang et al. 2006).

We note that when the growth rate is known one can easily derive the doubling time, denoted as $D(t)$ as it is calculated by the following formula:

$$D(t) = \ln(2)/R(t)$$

It is obvious that an increase in growth rate leads to a decrease in doubling time and vice versa. Doubling time, being easily derived from the growth rate, will not be discussed further.

(b) Raw counts alone do not convey much information as these figures are confounded by the size of the country and the size of the subject specialties. Hence, an index called Relative Priority Index (RPI) is computed for cross national comparison using the following formula (Nagpaul and Sharma 1995):
RPI $= (a/b)/(c/d)$ where:

$a =$ number of publications of country C in field F;
$b =$ number of publications of country C in all fields;
$c =$ number of publications of all countries in field F;
$d =$ number of publications of all countries in all fields.

The RPI was first suggested by Frame (1977), under the name of activity index, and later studied by Schubert and Braun (1986), among others. The RPI can be understood as "the country's share in the World's publication output in the given field F" divided by "the country's share in the World's publication output in all science fields", or, equivalently: "the given field's share in the country's publication output" divided by "the given field's share in the World's publication output" (Schubert and Braun 1986; Rousseau and Yang 2012).

Sometimes the RPI is multiplied by 100 for easy comparison, leading to an indicator which we denote by RIP_0. In this case the value $\mathrm{RPI}_0 = 100$ indicates that the research priority of a country for a given field corresponds precisely to the average of all countries i.e. average priority; $\mathrm{RPI}_0 > 100$ reflects higher-than-average priority, and $\mathrm{RPI}_0 < 100$ lower-than-average priority. We recall that, because of its definition, no country can have a high priority ($\mathrm{RPI} > 1$ or $\mathrm{RIP}_0 > 100$) in all fields.

Rousseau and Yang warned that in some cases this index may yield misleading results (Rousseau and Yang 2012; Rousseau 2012) as activities outside the field and the country under study may influence its behaviour. Aksnes et al. (2014) studied the effect on the Netherlands of the Relative Specialization Index (RSI), a renormalization of the RPI, as the result of the increased representation of China in the WoS. We note that in this investigation two dynamic aspects are at play: the huge growth of China in terms of publications (described as 'booming') and the change of the WoS over time (possibly influenced by China). Aksnes et al. (2014) concluded however that, although the influence of China is visible in the RSI for the Netherlands, and this especially in the latest decade and in fields where the countries have opposite specializations, the basic research profile of the Netherlands as measured by the RSI remained the same.

In another approach to stability we will compare results obtained from PubMed with results obtained from the WoS (with respect to all publications and with respect to medical publications).

Values of the RPI will be used in this investigation to compare

1. Research priorities of countries for different forms of viral hepatitis in a given time span;
2. Research priorities of countries for a given field at different time spans.

We will calculate the relative priority index to study cross-national assessment of hepatitis A, B and C (the most important cases).

Further queries

As we will calculate priority indices in three databases, namely PubMed, the whole WoS and the WoS, restricted to the area Life Sciences Biomedicine, denoted as WoS(LS), this will not only inform us about research priorities of countries, but also about the interpretation of values of the priority index. Concretely: is the priority index reasonably stable or is it highly dependent on the used database?

We choose the following countries in different continents:

Africa: Egypt
America: Brazil, Canada, USA
Asia: China, India, Japan
Europe: Belgium, France, Germany, UK
Oceania: Australia

Eleven of these are well-known leading research countries in their continents. We just added Belgium as the home country of one of the authors. Some country affiliations are difficult to search for, especially in PubMed (see "Appendix B" for some examples). Data for Germany in the period 1976–1995 were obtained as: CU = (GERMANY OR "FED REP GER" OR "DEUTSCH DEM REP" OR "GER DEM REP" OR "DDR" OR "BUNDES REP*") in the WoS and similarly in PubMed. Queries for China in PubMed retrieved also articles originating from Taiwan (Republic of China in the address field). We explain in the appendix how this problem was solved. In PubMed we searched for UK or United Kingdom, while in the WoS we searched for CU = (England OR Scotland OR Wales OR North Ireland). Finally articles originating from the USA were retrieved in PubMed by a search for (USA OR US OR United States OR America). As some records in PubMed only have the name of the state, e.g. NY or CA, this misses an unknown number of contributions from the USA. As PubMed has no country field, only an address field (affiliation) other errors occur when the name of the locality or company includes one the countries studied here (see Appendix B for examples). As a result of our queries numbers given further on refer to publications in which at least one author belongs to the corresponding country. Articles resulting from international collaboration between these twelve countries are hence counted more than once. Data from the WoS refer to publications of "article' type and to publications of "journal article" type in PubMed.

Consequently most absolute numbers retrieved from PubMed are incorrect, especially for the USA. Yet, as we are interested in ratios, and the same type of queries were performed when considering hepatitis A, B and C we do not expect (but, of course, cannot prove) that ratios are heavily affected by these imprecise retrieval strategies.

Results and discussions

Comparison between the World and India

In the first decade (1976–1985) we could hardly find hepatitis articles in PubMed written by Indian scholars, although the yearly average of Indian publications of article type in the WoS is about ten. For this reason we restrict the comparison between India and the World to the 30-year period (1986–2015). Further analyses, however, will cover the complete 40-year period (1976–2015) announced in the title and the previous section. The total publication output for hepatitis for the World and India as obtained from PubMed is shown in Table 1, along with the growth rate for the period (1986–2015). Corresponding results from the WoS are shown in Table 2. The growth rate of hepatitis publications in the World, according to PubMed, decreases gradually from 0.72 to 0.07, but seems to be constant from 2006 to 2015. The mean relative growth rate for the World is 0.15 (median = 0.09). The growth rate of Indian research output on hepatitis also decreases gradually from 1.79 to 0.10. The average growth rate for India is 0.27 (median value: 0.14). Results for the World according to the WoS are very similar (see Table 2). India, according to the WoS, starts from a lower growth rate (0.92) than according to PubMed.

Figure 1 illustrates the growth in number of publications about hepatitis deposited in PubMed from 1986 to 2015 for the World and for India. Initially the rise in the number of

Table 1 World versus India: growth rates of hepatitis publications as reflected in PubMed (1986-2015)

Year	World			India		
	Pub.	Cum	$R(t)$	Pub	Cum	$R(t)$
1986	1855	1855	0.72	1	1	1.79
1987	1947	3802	0.41	5	6	1.25
1988	1921	5723	0.32	15	21	0.54
1989	2197	7920	0.28	15	36	0.35
1990	2516	10,436	0.23	15	51	0.21
1991	2690	13,126	0.20	12	63	0.50
1992	2847	15,973	0.17	41	104	0.25
1993	3002	18,975	0.16	29	133	0.22
1994	3184	22,159	0.15	33	166	0.14
1995	3666	25,825	0.12	25	191	0.12
1996	3382	29,207	0.12	25	216	0.11
1997	3639	32,846	0.11	24	240	0.13
1998	3679	36,525	0.10	34	274	0.15
1999	4025	40,550	0.10	45	319	0.16
2000	4253	44,803	0.09	56	375	0.13
2001	4338	49,141	0.09	52	427	0.16
2002	4742	53,883	0.09	73	500	0.13
2003	4967	58,850	0.09	69	569	0.14
2004	5337	64,187	0.09	85	654	0.12
2005	5803	69,990	0.08	86	740	0.14
2006	5903	75,893	0.08	108	848	0.12
2007	6270	82,163	0.07	110	958	0.14
2008	6399	88,562	0.07	141	1099	0.12
2009	6659	95,221	0.07	145	1244	0.12
2010	7074	102,295	0.07	158	1402	0.13
2011	7862	110,157	0.07	195	1597	0.13
2012	8219	118,376	0.07	226	1823	0.13
2013	8957	127,333	0.07	245	2068	0.13
2014	9443	136,776	0.07	289	2357	0.10
2015	9563	146,339		252	2609	
Average			0.15			0.27
Median			0.09			0.14

Table 2 World versus India: growth rates of hepatitis publications as reflected in the WoS (1986–2015)

Year	World			India		
	Pub.	Cum	$R(t)$	Pub	Cum	$R(t)$
1986	863	863	0.72	8	8	0.92
1987	909	1772	0.38	12	20	0.53
1988	817	2589	0.29	14	34	0.34
1989	869	3458	0.27	14	48	0.27
1990	1088	4546	0.38	15	63	0.29
1991	2073	6619	0.29	21	84	0.31
1992	2191	8810	0.23	31	115	0.20
1993	2304	11,114	0.19	25	140	0.21
1994	2375	13,489	0.18	33	173	0.15
1995	2613	16,102	0.17	28	201	0.10
1996	3026	19,128	0.16	22	223	0.09
1997	3371	22,499	0.14	21	244	0.16
1998	3388	25,887	0.13	41	285	0.15
1999	3648	29,535	0.12	46	331	0.16
2000	3732	33,267	0.11	57	388	0.11
2001	3795	37,062	0.10	47	435	0.16
2002	3980	41,042	0.10	73	508	0.12
2003	4274	45,316	0.09	65	573	0.14
2004	4365	49,681	0.09	83	656	0.10
2005	4918	54,599	0.09	70	726	0.14
2006	4934	59,533	0.09	105	831	0.12
2007	5500	65,033	0.08	109	940	0.14
2008	5614	70,647	0.08	141	1081	0.13
2009	5681	76,328	0.08	148	1229	0.11
2010	6067	82,395	0.08	147	1376	0.12
2011	6531	88,926	0.07	172	1548	0.10
2012	6805	95,731	0.07	161	1709	0.10
2013	7299	103,030	0.07	175	1884	0.10
2014	7539	110,569	0.07	196	2080	0.08
2015	7636	118,205		167	2247	
Average			0.17			0.19
Median			0.11			0.14

Indian publications is slow. One of the possible reasons may be the lack of awareness of the severity and lethality of hepatitis. However, the number of publications about hepatitis has substantially increased over time so that since 2008 the cumulative number of Indian publications has more than doubled, according to both databases (see Tables 1, 2). This observation reflects not only the interest of Indian scientists towards research and development in the field, but also the higher funding by governmental and non-governmental organizations, such as funding for training by the Bill and Melinda Gates Children's Vaccine Programme (PATH/CVP) (WHO 2002; Government of India 2011; Kumar and Sarin 2014).

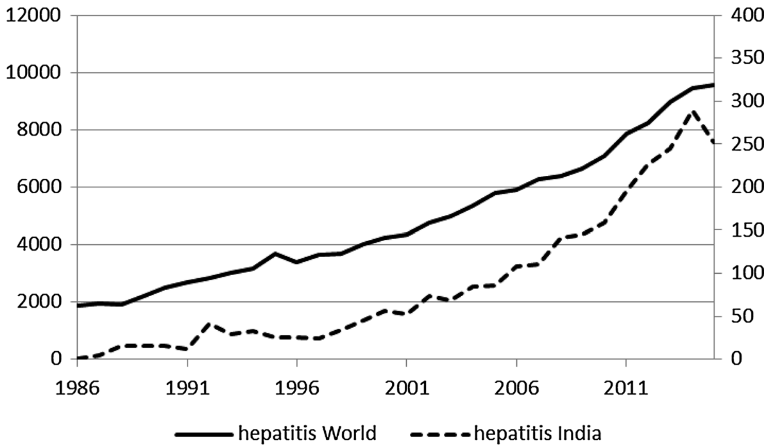

Fig. 1 Yearly number of publications on hepatitis (all types) in the World (left axis) and in India (right axis) as reflected in PubMed (1986–2015)

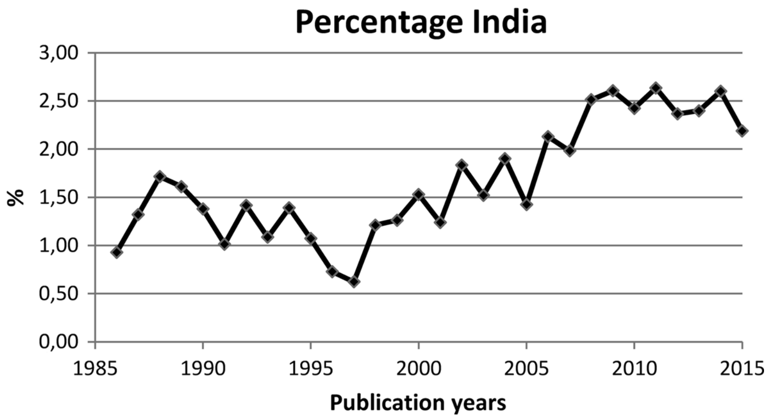

Fig. 2 Yearly percentage of Indian publications to the World's literature on hepatitis according to the WoS

Figure 2 shows the yearly relative contribution of India to the World's literature on hepatitis (article type only) based on WoS data. This percentage has a decreasing trend from 1988 till 1997, an increasing trend from 1997 till 2009 and stays more or less stable around 2.5% since then (as usual, the last year may be incomplete).

According to both databases, the growth rate of Indian research output on hepatitis is higher than that of the World output. There are probably several reasons for this: one is that India starts from the lowest possible value (one publication in 1986 in PubMed) hence it is easy to show a large increase. Another reason may be that hepatitis is a well-known disease that affects developing countries more than leading scientific countries, who might prefer to concentrate biomedical research on the latest developments such as e.g., CRISPR. Since India is a developing country and the incidence of hepatitis is high, emphasis is given for research activities in fields such as hepatitis research. In the recent past, the Government of India has been keen in eradicating contagious diseases like hepatitis, polio, AIDS, cholera etc., by supporting research projects and by providing sufficient funds to carry out research activities in these fields (Government of India 2011; Kumar and Sarin 2014). This might have encouraged young academicians and scientists to take up research projects. International funding agencies such as the already mentioned Bill and Melinda Gates Foundation, are also coming forward to fund research on hepatitis and related branches of study.

Finally we note that we found, over the period 1986–2015, 146,339 hepatitis publications in PubMed and 118,205 in the WoS. Hence, 24% more in PubMed than in the WoS. This illustrates the well-known fact that PubMed covers the medical literature better than the WoS. For India the numbers are respectively 2609 and 2247 or 16% more in PubMed. This leads to the question: why is World coverage in PubMed relative higher than the coverage of Indian research? Is this just the case for hepatitis research? Is this the case for most countries other than the USA? We leave these questions as open problems for further research.

Major countries and their priority indices for three types of viral hepatitis

We recall that we choose 12 countries spread over five continents to study their priority profiles in the fields of hepatitis A, B and C over two periods: 1976–1995 and 1996–2015. Tables 3 and 4 show the results based on PubMed data, while Tables 5 and 6 show the corresponding data based on the whole WoS. Finally, Tables 7 and 8 show the results based on a biomedical sub-database of the WoS (the query used to delineate this sub-database is shown in "Appendix C"). Considering this biomedical sub-database was done to obtain results which are more comparable to the PubMed results.

In the first period 1976–1995, the United States published the most articles on hepatitis, followed by Japan and France (PubMed). In the WoS (complete and restricted to Life Sciences), Germany is third. In the second period, however, China has becomes second (in

Table 3 Publication output on hepatitis in major countries according to PubMed (in Title/Abstract; journal articles); period 1976–1995

PubMed data	# Articles	Numbers of articles			Priority index		
1976–1995	All fields	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis A	Hepatitis B	Hepatitis C
Australia	44,642	20	101	45	88	87	136
Belgium	22,951	28	90	36	240	151	211
Brazil	18,809	10	51	16	105	104	114
Canada	88,874	26	119	48	58	51	73
China	12,755	14	105	17	216	316	179
Egypt	3066	9	47	17	579	589	746
France	90,679	52	407	244	113	172	362
Germany	76,183	101	336	111	261	169	196
India	21,866	17	97	84	153	170	517
Japan	179,698	53	707	723	58	151	541
UK	102,067	48	281	143	93	106	188
USA	958,785	261	1972	476	54	79	67
World	6488,840	3292	16,890	4823	100	100	100

Table 4 Publication output on hepatitis in major countries according to PubMed (in Title/Abstract; journal articles); period 1996–2015

PubMed data 1996-2015	# Articles All fields	Numbers of articles Hepatitis A	Hepatitis B	Hepatitis C	Priority index Hepatitis A	Hepatitis B	Hepatitis C
Australia	306,884	101	780	1191	75	79	99
Belgium	118,931	94	378	396	179	99	85
Brazil	210,210	170	684	941	183	102	115
Canada	415,087	141	671	1242	77	50	77
China	807,546	270	7600	2077	76	294	66
Egypt	31,745	41	238	855	293	234	690
France	432,610	199	1578	2926	104	114	173
Germany	615,792	214	1602	2180	79	81	91
India	273,108	211	1049	627	175	120	59
Japan	762,146	175	2357	4838	52	97	163
UK	790,798	194	1398	2237	56	55	72
USA	3506,196	1109	6352	12,040	72	57	88
World	13,099,005	5781	41,955	51,158	100	100	100

Table 5 Publication output on hepatitis in major countries according to the WoS (publications of article type); period 1976–1995

WoS data 1976–1995	# articles All fields	Numbers of articles Hepatitis A	Hepatitis B	Hepatitis C	Relative Priority index Hepatitis A	Hepatitis B	Hepatitis C
Australia	220,354	57	210	78	145	100	104
Belgium	97,070	64	157	53	370	169	160
Brazil	53,293	18	68	15	190	134	83
Canada	468,061	37	219	32	44	49	20
China	88,022	23	171	27	147	203	90
Egypt	27,698	8	43	24	162	162	254
France	546,731	126	786	321	129	150	172
Germany	750,198	180	890	201	135	124	79
India	220,080	17	133	21	43	63	28
Japan	662,955	56	880	727	47	139	322
UK	833,280	138	871	248	93	109	87
USA	3,732,558	574	2859	832	86	80	65
World	1,0361,238	1845	9902	3528	100	100	100

Table 6 Publication output on hepatitis in major countries according to the WoS (publications of article type); period 1996–2015

WoS data	# articles	Numbers of publications			Relative priority index		
1996–2015	All fields	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis A	Hepatitis B	Hepatitis C
Australia	634,010	140	1057	1584	86	97	105
Belgium	275,008	195	705	792	276	149	121
Brazil	439,234	181	705	1092	161	94	104
Canada	894,819	211	1014	1911	92	66	90
China	1990,737	232	6699	2484	45	196	52
Egypt	86,287	55	291	975	249	197	474
France	1110,731	371	2150	4130	130	113	156
Germany	1544,807	400	2359	3462	101	89	94
India	622,152	185	880	578	116	82	39
Japan	1460,945	259	2629	5820	69	105	167
UK	1605,290	279	1926	3157	68	70	83
USA	5954,403	1464	8389	15,494	96	82	109
World	20,448,768	5245	35,070	48,703	100	100	100

Table 7 Publication output on hepatitis in major countries according to the biomedical sub-database of the WoS (publications of article type); period 1976–1995

WoS data	# articles	Numbers of articles			Relative priority index		
1976–1995	medical fields	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis A	Hepatitis B	Hepatitis C
Australia	127,522	57	208	76	123	85	86
Belgium	57,670	63	157	52	302	142	129
Brazil	29,750	18	68	15	167	119	72
Canada	242,699	35	214	32	40	46	19
China	18,224	18	143	26	273	410	205
Egypt	10,536	8	43	24	210	213	327
France	287,608	124	755	320	119	137	160
Germany	370,114	180	872	200	134	123	78
India	87,086	16	129	20	51	77	33
Japan	306,049	56	854	714	51	146	335
UK	457,530	135	850	247	81	97	78
USA	1942,494	557	2685	796	79	72	59
World	4995,021	1809	9570	3478	100	100	100

Table 8 Publication output on hepatitis in major countries according to the biomedical sub-database of the WoS (publications of article type); period 1996–2015

WoS data	# articles	Numbers of articles			Relative priority index		
1996–2015	All medical fields	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis A	Hepatitis B	Hepatitis C
Australia	355,861	136	1019	1501	75	84	89
Belgium	149,693	187	695	754	245	137	106
Brazil	272,968	172	677	1052	123	73	81
Canada	484,440	198	958	1800	80	58	78
China	600,410	207	5995	2186	68	294	77
Egypt	37,276	49	269	943	258	213	532
France	504,591	364	2080	3985	141	121	166
Germany	742,046	381	2261	3294	101	90	93
India	235,138	180	832	532	150	104	48
Japan	708,823	253	2552	5642	70	106	167
UK	831,786	264	1840	3016	62	65	76
USA	3267,134	1397	7918	14,648	84	71	94
World	9786,928	4996	33,236	46,521	100	100	100

PubMed and WoS, but third when restricted to the Life Sciences), after the United States and before Japan (Japan is second in the WoS life sciences sub-database).

The relative priority index differs sometimes considerably between databases and between periods. This will be discussed in the next section. Here we have a look at results that are stable with respect to databases. Canada has always a low priority index for all databases, all periods and all types of viral hepatitis, while Egypt always has a high, even very high priority index. Also Belgium and France mostly have a high priority index (in all cases). India has most of the time a high priority index in PubMed, but has low values in the WoS. China has high values in the first period (1976–1995) in the medical databases (PubMed and WoS(LS)). It further has high values in all databases for hepatitis B research. The USA has mostly low values as does the UK. Japan has high values (in all databases and all periods) for hepatitis C and low values for hepatitis A. Generally there seem to be more extreme values (high or low) for the first period than for the second. This can be explained by the fact that all raw data are smaller in the first period than in the second one, leading to larger fluctuations in ratios.

Discussions related to the relative priority index (RPI)

It is clear that data differ strongly between PubMed and the WoS. Also taking the whole database or just the Life Sciences Biomedicine (LS) yields different results. We calculated the corresponding Pearson correlation coefficients, shown in Table 9. Correlations are always calculated with respect to the same type of hepatitis research. It makes no sense to calculate a calculation for the empty areas as they refer to different periods AND different databases.

Table 9 Pearson correlation coefficients between priority indices; data for (hepatitis A/hepatitis B/hepatitis C) in that order

	PubMed Period 1	PubMed Period 2	WoS (all) Period 1	WoS (all) Period 2	WoS (LS) Period 1	WoS (LS) Period 2
PubMed Period 1	1	0.77/0.82/0.73	0.37/0.57/0.68		0.57/0.62/0.71	
PubMed Period 2		1		0.84/0.90/0.99		0.89/0.97/0.99
WoS (all) Period 1			1	0.86/0.83/0.92	0.74/0.83/0.66	
WoS (all) Period 2				1		0.96/0.94/0.99
WoS (LS) Period 1					1	0.54/0.95/0.72
WoS (LS) Period 2						1

Correlations between the same database in different periods are generally high (0.70) to very high (0.95) with the exception of hepatitis A in WoS(LS). For period 1 we see that the correlation between PubMed and the WoS databases is moderate to high (0.37–0.71), but these correlations are much higher for the second period (0.84–0.99). A somewhat similar phenomenon occurs, be it with higher correlation values, for the two WoS databases. For period 1 correlations are between 0.66 and 0.83, while for period 2 these correlations lie between 0.94 and 0.99. Finally, concerning the types of hepatitis, we see that correlations involving hepatitis A are never the highest and often (5 out of 9 cases) the lowest. Correlations involving hepatitis C are in 6 cases out of 9 the highest.

Conclusion

We found that, initially the rise in the number of Indian publications is slow. Yet, the number of publications on hepatitis has substantially increased over time so that since 2008 India's cumulative number of publications has more than doubled. In recent years the percentage contribution of India in research on hepatitis is about 2.5%.

This investigation confirms, see (Rousseau 2012; Rousseau and Yang 2012) that, because the activity or relative priority index is a relative index, it is often difficult to find a proper interpretation. The reasons for this are that, besides being dependent on the used universe, i.e. database or sub database, it includes a factor, namely d (number of publications of all countries in all field) which is influenced by activities outside the country or countries and the field(s) of interest. A deeper study on this phenomenon and a suggestion for an index replacing the activity index can be found in (Rousseau 2018).

Final remarks on hepatitis itself

Viral hepatitis is a worldwide leading cause of mortality. Indian high quality research in this field is largely missing and requires strategic planning, investment and more support. There is also a need to improve the existing medical education system, which should foster research culture. Keeping in view the severity of liver disorders in India, there is a need to increase the investment in R & D to bring about improvements in supportive care. For this reason the authors, as information scientists, stress the need to set up a Hepatitis Registry. Efforts must be made to develop state/union government-supported prevention and control strategies, including conducting mass awareness programs, formulation of universal guidelines for immunization, promotion of partnerships, development of an evidence-based policy, and development of programs for prevention of transmission, along with better financial and social support initiatives. To tackle this deadly disease, the Government of India said that it had launched a national programme for prevention and control of viral Hepatitis during the 12th Five Year Plan period, but there is little activity visible at the ground level.

Acknowledgements The authors thank Raf Guns (Univ. Antwerp) for useful remarks leading to a better manuscript. We also thank the reviewers for useful comments.

Appendix A: Short descriptions of the five forms of viral hepatitis, taken from the Fact Sheets of the WHO

Hepatitis A

- Hepatitis A is a viral liver disease that can cause mild to severe illness.
- The hepatitis A virus (HAV) is transmitted through ingestion of contaminated food and water or through direct contact with an infectious person.
- Almost everyone recovers fully from hepatitis A with a lifelong immunity. However, a very small proportion of people infected with hepatitis A could die from fulminant hepatitis.
- The risk of hepatitis A infection is associated with a lack of safe water, and poor sanitation and hygiene (such as dirty hands).
- Epidemics can be explosive and cause substantial economic loss.
- A safe and effective vaccine is available to prevent hepatitis A.
- Safe water supply, food safety, improved sanitation, hand washing and the hepatitis A vaccine are the most effective ways to combat the disease.

Hepatitis B

- Hepatitis B is a viral infection that attacks the liver and can cause both acute and chronic disease.
- The virus is transmitted through contact with the blood or other body fluids of an infected person.
- An estimated 257 million people are living with hepatitis B virus infection (defined as hepatitis B surface antigen positive).
- In 2015, hepatitis B resulted in 887,000 deaths, mostly from complications (including cirrhosis and hepatocellular carcinoma).

- Hepatitis B is an important occupational hazard for health workers.
- However, it can be prevented by a currently available safe and effective vaccine.

Hepatitis C

- Hepatitis C is a liver disease caused by the hepatitis C virus: the virus can cause both acute and chronic hepatitis, ranging in severity from a mild illness lasting a few weeks to a serious, lifelong illness.
- The hepatitis C virus is a blood borne virus and the most common modes of infection are through exposure to small quantities of blood. This may happen through injection drug use, unsafe injection practices, unsafe health care, and the transfusion of unscreened blood and blood products.
- Globally, an estimated 71 million people have chronic hepatitis C infection.
- A significant number of those who are chronically infected will develop cirrhosis or liver cancer.
- Approximately 399,000 people die each year from hepatitis C, mostly from cirrhosis and hepatocellular carcinoma.
- Antiviral medicines can cure more than 95% of persons with hepatitis C infection, thereby reducing the risk of death from liver cancer and cirrhosis, but access to diagnosis and treatment is low.
- There is currently no vaccine for hepatitis C

Hepatitis D

- Hepatitis D virus (HDV) is a ribonucleic acid (RNA) virus that requires hepatitis B virus (HBV) for its replication. HDV infection occurs only simultaneously or as superinfection with HBV.
- The virus is transmitted through contact with the blood or other body fluids of an infected person.
- Approximately 15 million people across the World are chronically co-infected with HDV and HBV.
- Currently there is no effective antiviral treatment for hepatitis D.
- Hepatitis D infection can be prevented by hepatitis B immunization

Hepatitis E

- Hepatitis E is a liver disease caused by infection with a virus known as hepatitis E virus (HEV).
- Every year, there are an estimated 20 million HEV infections Worldwide, leading to an estimated 3.3 million symptomatic cases of hepatitis E, and 56,600 hepatitis E related deaths.
- Hepatitis E is usually self-limiting but some cases may develop into fulminant hepatitis (acute liver failure).
- The virus is transmitted via the faecal-oral route, principally via contaminated water.
- Hepatitis E is found Worldwide, but the prevalence is highest in East and South Asia.
- A vaccine to prevent hepatitis E virus infection has been developed and is licensed in China, but is not yet available elsewhere.

Appendix B: Retrieval problems in PubMed

We include these examples as warnings as well as suggestions for (partial) solutions.

Problem 1. American records that do not show the words US, USA or United States in the address field (Author information).

An example:

STEP Perspect. 1995 Summer; 7(2):13–15.
Liver function and HIV-1 infection.
Hernandez V.
Author information: Direct AIDS Alternative Information Resources, New York, NY.

We checked that if, indeed, this record cannot be retrieved with an American affiliation:

Search (((((“1976/01/01”[Date - Publication]: “1995/12/31”[Date - Publication])) AND hepatitis[Title/Abstract])) AND Hernandez V[Author]) AND (USA [Affiliation] OR US[Affiliation]).

This search yields no records.

A partial solution

The term “America” occurred sometimes in the name of an institution that otherwise only provided state information. So, by adding “America” in the search we retrieved quite some records that had no USA country name. An example is the following record.

J Perinat Educ. 2005 Winter; 14(1):46–49.
Toward more evidence-based practice.
Hotelling BA.
Author information: BARBARA HOTELLING is an independent childbirth educator and doula in Rochester Hills, Michigan. She has served as president of Lamaze International, president of Doulas of North America (DONA), and chair of the Coalition for Improving Maternity Services (CIMS).

Similarly, this happened with the term US. By adding US in the search we retrieved records that had no USA country name, and not “America’ in the affiliation field. An example is the following record.

Annu Rev Public Health. 1988; 9:203–221.
Viral vaccines and antivirals: current use and future prospects.
Parkman PD., Hopps HE.
Author information: Center for Biologics Evaluation and Research, US Food and Drug Administration, Bethesda, Maryland 20892.

However, this may have led to new errors, such as in the following record, which is not an American contribution (in our sense). Yet, we decided that the gain (correct unique records) resulting from including US in the query surpassed the loss (adding incorrect records). An example:

Trans R Soc Trop Med Hyg. 1994 Jan–Feb; 88(1):57.
Hepatitis E virus in Indonesia.
Jennings GB., Lubis I., Listiyaningsih E., Burans JP., Hyams KC.
Author information: US Naval Medical Research Unit No. 2, Jakarta, Indonesia.

Problem 2. Country names occurring in the address field of another country.

An example: Because of the local address of the institution (which includes US) this record is counted for the USA, but it is also counted for China.

IEEE Trans Neural Netw. 1992; 3(3):423–433.
The Mod 2 Neurocomputer system design.
Mumford ML., Andes DK., Kern LL.
Author information: US Naval Weapons Center, China Lake, CA.

Problem 3. China in PubMed

We searched for China OR Beijing (the capital city was included in order to retrieve articles with no country affiliation). This is set A.

Next we searched for “Republic of China”. The resulting set is set B. Most of these must be removed from the original search, except those that have “People’s Republic of China” in the affiliation field (set C). The required set of publications is those included in A\ (B\C). This is illustrated in Fig. 3.

Appendix C: Search for Life Sciences Biomedicine

SU = (Agriculture OR Allergy OR Anatomy & Morphology OR Anesthesiology OR Anthropology OR Behavioral Sciences OR Biochemistry & Molecular Biology OR Biodiversity & Conservation OR Biophysics OR Biotechnology & Applied Microbiology OR Cardiovascular System & Cardiology OR Cell Biology OR Critical Care Medicine OR Dentistry, Oral Surgery & Medicine OR Dermatology OR Developmental Biology OR Emergency Medicine OR Endocrinology & Metabolism OR Entomology OR Environmental Sciences & Ecology OR Evolutionary Biology OR Forestry OR Gastroenterology & Hepatology OR General & Internal Medicine OR Genetics & Heredity OR Geriatrics & Gerontology OR Health Care Sciences & Services OR Hematology OR Immunology OR Infectious Diseases OR Integrative & Complementary Medicine OR Legal Medicine OR Life Sciences Biomedicine Other Topics OR Marine & Freshwater Biology OR Mathematical & Computational Biology OR Medical Ethics OR Medical Informatics OR Medical Laboratory Technology OR Microbiology OR Mycology OR Neurosciences & Neurology OR Nursing OR Nutrition & Dietetics OR Obstetrics & Gynecology OR Oncology OR Ophthalmology OR Orthopedics OR Otorhinolaryngology OR Paleontology OR Parasitology OR Pathology OR Pediatrics OR Pharmacology & Pharmacy OR Physiology OR Plant Sciences OR Psychiatry OR Public, Environmental & Occupational Health OR Radiology, Nuclear Medicine & Medical Imaging OR Rehabilitation OR

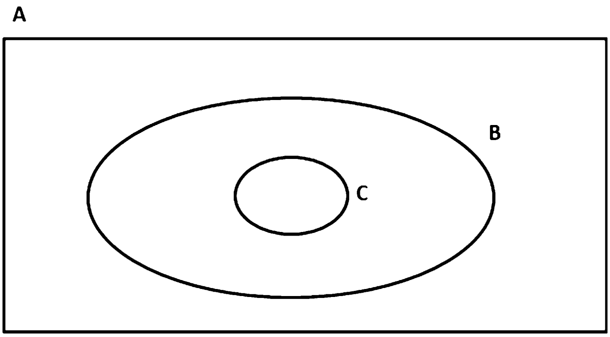

Fig. 3 Illustration of the methodology to obtain China’s contributions in PubMed

Reproductive Biology OR Research & Experimental Medicine OR Respiratory System OR Rheumatology OR Sport Sciences OR Substance Abuse OR Surgery OR Toxicology OR Transplantation OR Tropical Medicine OR Urology & Nephrology OR Veterinary Sciences OR Virology OR Zoology).

References

Aksnes, D. W., van Leeuwen, T. N., & Sivertsen, G. (2014). The effect of booming countries on changes in the relative specialization index (RSI) on country level. *Scientometrics, 101*(2), 1391–1401.

Amaral, L. A. N., Gopikrishnan, P., Matia, K., Plerou, V., & Stanley, H. E. (2001). Application of statistical physics methods and concepts to the study of science & technology systems. *Scientometrics, 51*(1), 9–36.

Frame, J. D. (1977). Mainstream research in Latin America and the Caribbean. *Interciencia, 2*(3), 143–148.

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*(10053), 1545–1602.

Global Hepatitis Report 2017 (2017). Geneva: World Health Organization.

Government of India. (2011). *Operational guidelines for hepatitis B vaccine introduction in Universal Immunization Programme*. New Delhi: Ministry of Health and Family Welfare, Government of India.

Kumar, M., & Sarin, S. K. (2014). Viral hepatitis eradication in India by 2080—Gaps, challenges & targets. *Indian Journal of Medical Research, 140,* 1–4.

Liang, L. M., Havemann, F., Heinz, M., & Wagner-Döbler, R. (2006). Structural similarities between science growth dynamics in China and in western countries. *Scientometrics, 66*(2), 311–325.

Nagpaul, P. S., & Sharma, L. (1995). Science in the eighties: A typology of countries based on inter-field priorities. *Scientometrics, 34*(2), 263–283.

Rousseau, R. (2012). Thoughts about the activity index and its formal analogues. *ISSI Newsletter, 8*(4), 73–75.

Rousseau, R. (2018). The F-measure for research priority. *Journal of Data and Information Science*, 3(1) [to appear].

Rousseau, R., & Yang, L. Y. (2012). Reflections on the Activity Index and Related Indicators. *Journal of Informetrics, 6*(3), 413–421.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5), 281–291.

WHO (2002). *Prevention of Hepatitis B in India*: *An overview.* World Health Organization, Regional Office for South-East Asia, New Delhi. SEA-Hepat.-5; SEA-EPI-141.