Scientometrics (2020) 125:1733–1761
https://doi.org/10.1007/s11192-020-03572-9

Do national funding organizations properly address the diseases with the highest burden?: Observations from China and the UK

Lin Zhang[1,2] · Wenjing Zhao[1] · Jianhua Liu[3] · Gunnar Sivertsen[4] · Ying Huang[1,2]

Received: 25 September 2019 / Published online: 25 June 2020

Abstract
Recent years have witnessed an incipient shift in science policy from a focus mainly on academic excellence to a focus that also takes into account "societal impact". This shift raises the question as to whether medical research has given proper attention to the diseases imposing the greatest burden on society. Therefore, with the aim of identifying correlations between research funding priorities and public demand in health, we examine grants issued by the major medical research funding bodies of China and the UK during 2006–2017 and compare the focus of their funded projects with the diseases that carry the highest burden of death, risk, or loss of health. The results indicate that the funding decisions of both nations do correspond to the illnesses with the highest health impact on their citizens. For both regions, the greatest health concerns surround non-communicable diseases, and neoplasms and cardiovascular disease in particular. In China, national health priorities have remained focused on these illnesses for the benefit of its own population, whereas the UK has funded a wider variety of research, extending to projects with impacts outside its borders to some developing countries. Additionally, despite an increased incidence of mental illness and HIV/AIDs in China, there is evidence that less priority has been given to these conditions. Both of these health areas seem to require more attention from China's national funding agencies and the society in general. Methodologically, this study can serve as an example of how to conduct analyses related to public health issues by combining informetric methods and data with the tools and data from other fields, thereby inspiring other scientometrics studies.

Keywords Funding · Burden of disease · NSFC · MRC · Data visualization

✉ Ying Huang
huangying_work@126.com

[1] School of Information Management, Wuhan University, Wuhan, China

[2] Department of MSI, Centre for R&D Monitoring (ECOOM), KU Leuven, Leuven, Belgium

[3] Beijing Wanfang Data Ltd., Beijing, China

[4] Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

Introduction

Scientific research can be described as a social practice—that is, a complex, collaborative, goal-oriented and socially organized activity (Hicks and Stapleford 2016) that requires a large investment of time and money. Nobel Prize winner Patrick Blackett believes the curiosity of researchers should be the primary driver of advances in science (Anderson 1999; Blackett 1971). Meanwhile, Hicks et al. (2018) assert that scholarly research should be accompanied by both inward- and outward-facing goals motived by social practices and the broader impact of science. Moreover, some researchers have developed a framework for responsible innovation to address social and ethical concerns and to underpin a practical and systematic approach to governance (Stilgoe et al. 2013). Clearly, diverse motivations are involved in allocating scientific funding and are a significant part of science policy (Viergever et al. 2010). Decision-making may consider not only the general criteria of the scientific quality of projects and teams, and the potential for scientific advancement in the topic, but also societal demands or needs related to a given issue (Ciarli and Ràfols 2019), such as economic growth and altruistic goals for the betterment of one's citizenry (Klavans and Boyack 2017).

In recognition that scientific development may bring benefits beyond science itself, governments are increasingly being asked to make more specific and more substantial contributions to the health and wellbeing of their constituents (Cassi et al. 2017). Meanwhile, funding agencies and their funded projects and outputs are also being asked to be able to guide local decision-making and benefit populations (Mutapi 2019). Likewise, science policy is gradually shifting toward providing solutions to societal problems and grand challenges. For example, the US National Institutes of Health (NIH) created a working group on priority setting in 1997, which then confirmed social health needs as one of the criteria for research allocation (Gross et al. 1999). The US National Science Foundation (NSF) also included "achieving societal goals" as one of the review principles for application proposals. Researchers were also required to identify how a potential project "encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes" (National Science Foundation 2018). In a similar vein, the "2019 NSFC Reform Initiatives" issued by the National Natural Science Foundation of China (NSFC) explicitly states that critical national demands should be one of the primary sources of scientific problems (National Natural Science Foundation of China 2019).

Assessing the levels of research effort required to address complex global problems or societal demands, such as climate change, food security, poverty reduction, or the burden of various diseases, has been drawing increasing attention in both research and science policy (Cassi et al. 2017). The Global Observatory on Health R&D—an initiative of the World Health Organization that aims to help identify health R&D priorities based on public health needs—has tried to bring together information and statistics on health issues and research (World Health Organization (WHO) 2017). Notably, 54 papers on the 2019 Altmetrics.com list of the top 100 most-mentioned scholarly articles related to medical and health science.[1] Rising general concerns over health issues have also led to increased interest by researchers and a growing need to allocate more research funding to health-related projects in line with public demand (Atala et al. 2018; Røttingen et al. 2013).

[1] https://www.altmetric.com/top100/2019/.

In practice, while there has been a strong emphasis on assessing the scientific quality of publications and funding projects to foster "excellence", relatively less attention has been given to assessing if research efforts address social needs (Ràfols and Yegros 2017). Previous bibliometric research of scientific funding has usually focused on the final output of allocations, such as studying the impact or effectiveness of grants using publications resulting from funding projects (Gao et al. 2019; Wang et al. 2012; Zhao et al. 2009). Other measures included mapping the research of a field of study using funding information (Zhou and Tian 2014) and investigating the inequality in funding allocations and publication distribution between institutions (Halffman and Leydesdorff 2010; Shibayama 2011).

The most common representation of public demand in the medical and health fields is the concept of the burden of disease. And the most widely used indicator for measuring the burden of disease is disability-adjusted life years (DALYs), which attempts to quantify the risk of death and the impact of loss of quality of life for individuals affected by disease or disability (Prüss-Üstün et al. 2003). As a science aimed at preventing and treating diseases that cause illness and death in humans, medical research is supposed to reduce the burden of disease (Hagenaars et al. 2019). Therefore, it would be reasonable to expect that more research effort and higher funding investment should be directed toward diseases with a relatively high burden.

Several previous studies have adopted bibliometric methods that rely on publication data to analyze research on disease and its effects on society in various countries/regions (Agarwal and Searls 2009; Begum et al. 2016; Yegros et al. 2019). For example, Begum et al. (2018) mapped research activity using publications and compared the disease burden of different cancers in 29 countries over a 10-year period (2007–2016). Another study using publication data from 2002–2013, which considered outputs and funding related to European non-communicable respiratory disease and its disease burden, indicated that this was a severely under-researched health condition (Begum et al. 2016). Kalita et al. (2015) used bibliometric analysis to describe the focus and distribution of public health research output in India, finding marked inequities concerning the burden of disease and the geographic distribution of research. The inequality in the global disparity of health research has also attracted some interest in recent years. Evans et al. (2014) linked the burden of disease with MEDLINE articles for 111 conditions to assess the influence of disease burden on health research in both the global and national contexts. The results indicate that many of the principal health needs in less developed countries do not attract attention among researchers in developed countries. In contrast, local health needs within developed countries are drawing increased attention.

In terms of comparing disease burden with scientific funding, most research is based on univariate or multivariable analyses with condition-specific funding amounts. Gross et al. (1999) examined the relationship between disease burden and NIH disease-specific research funding, concluding that levels and the amount of financing moderately correlate with US disease burden. Since the publication of this landmark study in 1999, several relevant studies have emerged with a similar methodology along with data on the amount of funding allocated to specific health conditions (Gillum et al. 2011; Kinge et al. 2014). However, not all funding agencies have established a disease classification system to allocate their investments. More research and new methods are needed to help assess both the knowledge production side of funded projects, and the articulation of research agendas to meet societal needs.

While there have been several studies on the associations between research funding and burdens of disease at the country level, such as the US (Gross et al. 1999; Gillum et al. 2011), Norway (Kinge et al. 2014) and Australia (Mitchell et al. 2009), it is necessary to

steer more research towards the needs of developing countries. Early in 2012, the WHO Consultative Expert Working Group published a report about strengthening global financing and coordination on health needs in developing countries (World Health Organization (WHO) 2012b). The report suggested that the level of publicly-funded research available has not met the health needs of developing countries. As the largest developing country and one of the most significant contributors to science and technology in the world (Xie et al. 2014), China held a central position in this nexus bringing historical insights into whether the diseases with the highest burden had received corresponding attention from government-funded research. Further analysis of China's response to the problems identified in the WHO report may provide more detailed empirical evidence.

This study mainly focuses on public funding instead of private funding due to the significant role played by government funding in optimizing scientific research resources and improving the efficiency of knowledge production in the nation. Government funding bodies are also considered more representative of the orientation and priorities at the national level than private funding mechanisms, especially in China where a single funding agency predominates (Wang et al. 2012). The National Natural Science Foundation of China (NSFC), a public funding agency, is the primary source for scientific research, especially for basic research. Moreover, another practical reason is that only the data from government-funded projects were available for examination.

As mentioned above, the NSFC is the largest funding agency for basic research in China. The NSFC has funded more than 300,000 projects in support of around 1 million researchers since it was established in 1986 (Gao et al. 2019). The NSFC has made great strides in promoting basic research into natural science, especially in medical and health-related fields. According to statistics released from NSFC Annual Report, in 2010–2018,[2] more than 22% of funding for General Program research (one of the most fundamental project types) was invested in the Department of Health Sciences among the eight scientific departments. This is the main reason we selected the NSFC as the leading national funding organization for China.

Two main trends have emerged from previous studies of several countries. The first is that public funding tends to be correlated with a country's burden of disease (Gross et al. 1999; Kinge et al. 2014). The second is that research effort in terms of publication tends to respond to local needs instead of global needs (Evans et al. 2014). To conduct a comparative analysis and observe the different funding priorities in terms of the global and national health burden, we selected the Medical Research Council of the UK (MRC) within the UK Research and Innovation (UKRI) as a second research object. UKRI is the national funding agency sponsored by the Department for Business, Energy and Industrial Strategy[3] in which the MRC is a sub-agency with a particular focus on coordinating and funding research into medical and health science.

Assessing the efficiency and sufficiency of public-funded research is a valuable undertaking for governments, funding organizations, and academia. A combined perspective on which diseases should be receiving research funding alongside those that already are funded can give decision-makers a clear overview of their current funding strategy. Such insights may suggest continued support of current endeavors, modifications to existing

[2] https://www.nsfc.gov.cn/publish/portal0/tab505/.

[3] https://www.gov.uk/government/organisations/uk-research-and-innovation.

policies, or demand new strategies to improve scientific development and make research activity more relevant and more effective (Ebadi and Schiffauerova 2016).

While investigating the question of whether health research funding organizations have paid corresponding attention to the diseases with the highest burden in China and the UK, we directly used titles and abstracts of funding projects as the research object, and adopted the Medical Text Indexer (MTI) produced by the National Library of Medicine (NLM).[4] This tool was used to extract MeSH terms[5] for each project so as to conduct a topic analysis of the NSFC and MRC grants. There is potential for further application of MTI as a tool to extract standardized subject terms from arbitrary medical texts, which may then be used to identify essential topics from different types of documents in the health field. Beyond identifying how research into the hottest currently funded topics has evolved, this study mainly explores the corresponding relationship between the diseases of severe concern and high burden with further extraction of disease-related terms according to MeSH terms identified by MTI. In addition to the above, we have also attempted to provide a more detailed analysis and discussion about three specific diseases. Thus, this research was designed to answer the following research questions:

1. What are the differences and similarities between the research priorities funded by the Health Science Department in the NSFC and the MRC?
2. What kinds of diseases receive close attention from public funding agencies? Is this attention consistent with the structure of disease burden in China and the UK from 2006 to 2017?
3. What is the relationship between funding levels and some diseases with specific features on the burden index, such as extreme burden or high growth rates?

This paper has three key highlights. Firstly, titles and abstracts of funded projects act as direct data sources for this study to assess both the topic structure and prior funded topics. This step was followed by a further comparison between high-burden diseases and the above data sources. Second, in terms of methods, our study presents an introduction to and an example of how to use MTI to explore the funded topics and the diseases of most concern for research projects. With the expansion of different types of research objects in scientometric studies, such as publications, patents and policy documents, we suggest that MTI can be further applied to diversified objects with the capability of extracting subject terms based on the MeSH vocabulary of arbitrary medical texts. Unlike previous studies that mostly focus on one single country, we regard China, the largest developing country, and the UK, a highly developed country, as the main research object for a national comparative analysis. Beyond comparing the funding priorities and health needs in these two nations, we also observe the different funding priorities of each country in terms of the global and national health burden, which might lead to some policy implications.

The paper unfolds as follows. The next section presents the data, methods and tools used in this analysis, which includes the background of how funding is awarded and the definition and data used to analyze the burden of disease. The main results and their interpretation are described in the third section. The last section contains a discussion of this research, including its limitations, plus our intended directions of future work.

[4] https://ii.nlm.nih.gov/MTI/.

[5] https://ii.nlm.nih.gov/MTI/index.shtml.

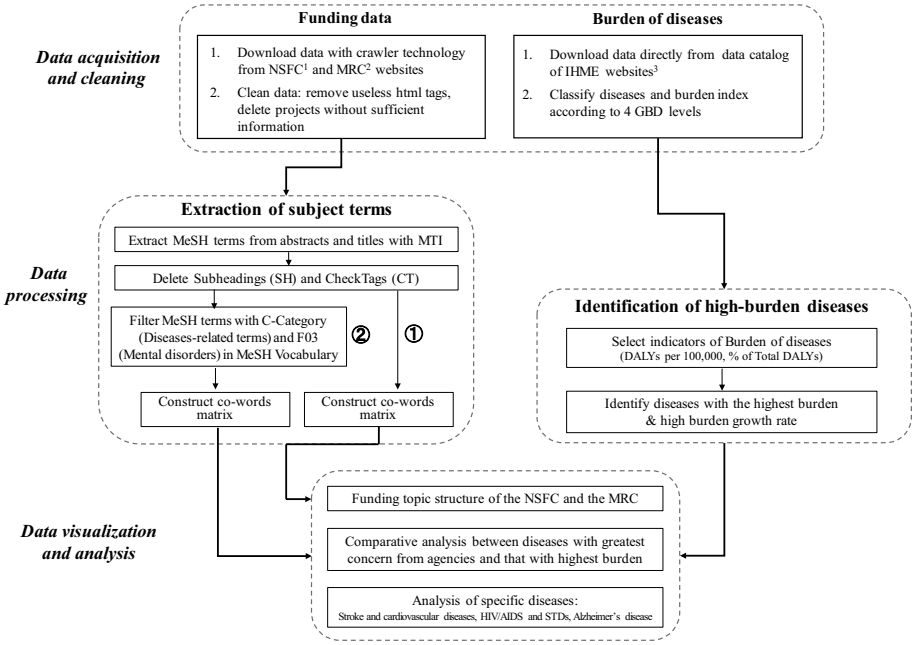

Fig. 1 Research framework. *Notes*: ① indicates the analytical process which outlines the topic skeletons of funded research. ② indicates the procedure used to undertake the comparative analysis of diseases with high concern and burden

Data and methodology

Figure 1 illustrates several important procedures used to conduct this research. This section describes the procedures showed in Fig. 1.

Term interpretation: MeSH stands for Medical Subject Headings, which is a biomedical indexing vocabulary maintained by the US National Library of Medicine (NLM); MTI stands for Medical Text Indexer, which provides indexing recommendations based on MeSH; MeSH terms here mean the recommended keywords produced by MTI from the projects' abstracts and titles; Checktag refers to a particular type of MeSH term required in recommended MeSH list of each biomedical text to designate species, sex, historical periods, and various kinds of research support.

1. https://npd.nsfc.gov.cn/; https://isisn.nsfc.gov.cn/egrantindex/funcindex/prjsearch-list
2. https://gtr.ukri.org/search/project?term=*
3. https://ghdx.healthdata.org/gbd-results-tool

Table 1 Basic statistics of NSFC and MRC-funded projects

	Projects with abstracts	Projects with titles only	Projects without abstracts and titles (removed)	Total
NSFC	33,499	4715	0	38,214
MRC	4006	165	70	4241

Data acquisition and cleaning

Funding data

Funding analysis seeks insights into how and where the most significant public financial resources in science are used. Most previous bibliometric studies derive funding information from publications (Morillo 2019; Wang et al. 2012; Zhou and Tian 2014). Nevertheless, this may cause problems with incomplete and inaccurate data that may skew the study's results (Tang et al. 2017). To ensure the reliability of the data, we acquired data about funding projects directly from official funding organizations (as shown in the 'Data acquisition and cleaning' of Fig. 1). This study could be regarded as a complementary study of previous work using publication data (Begum et al. 2016, 2018; Kalita et al. 2015). The two national organizations selected were the NSFC in China and the MRC in the UK.

The NSFC is tasked with administrating the Chinese Central Government's National Natural Science Fund and falls directly under the jurisdiction of the State Council. The funding system established by NSFC provides for two categories of predominantly domestic research—Research Programs and Talent Training Programs—and a further category of sponsorship for cooperative international research called Research Support Programs. Eight scientific departments share the management of the various research disciplines. Falling within the Research Programs category, we chose to study "General Programs" funded by the Department of Health Sciences. General Programs are considered the most centrally important projects and receive the lion's share of total funding. The Department of Health Sciences is the department responsible for managing health and medical research funding. (Note that before 2010, the Department of Life Sciences managed such projects, but the project codes used as the basis for inclusion in the corpus were consistent between the two departments across the whole period of study).

The MRC is one of nine councils that comprise UK Research and Innovation (UKRI) and is the council responsible for co-coordinating and funding medical and health research. UKRI is a non-departmental government funding agency that directs research and innovation funding, which brings together the seven Research Councils, Innovate UK, and Research England. The MRC is the main funding body for basic research in the field of health. Like the NSFC's General Program, the Research Grant is the most basic type of project in the MRC and is the research object of this study.

We collected raw data from the official NSFC[6] and UKRI[7] websites using Python web crawler technology on 29 July 2019. The last time the NSFC website was updated was 1 July 2019, and the MRC website was updated irregularly in sync with the UKRI. The

[6] https://npd.nsfc.gov.cn/; https://isisn.nsfc.gov.cn/egrantindex/funcindex/prjsearch-list

[7] https://gtr.ukri.org/search/project?term=*

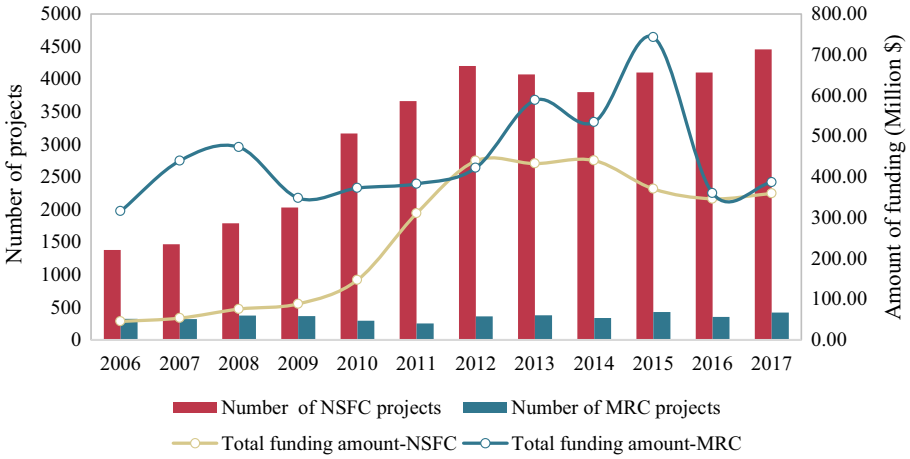

Fig. 2 The number of projects and funding amount granted by the NSFC and the MRC during 2006–2017

dataset comprised project titles, abstracts, principal investigators, recipient organizations, funding amount, project category (General Programs for the NSFC and Research Grant for the MRC) and grant dates for the period 2006–2017. The raw data were preprocessed by removing useless HTML tags and deleting projects without sufficient information. After excluding 70 MRC projects due to lack of both abstracts and titles, the final dataset contained a total of 38,214 NSFC-funded projects and 4171 MRC-funded projects. As noted in Table 1, 4715 NSFC projects and 165 MRC projects only had titles. Therefore, the titles were used as substitutes for the missing abstracts with these projects. Figure 2 shows the number of funded projects and the amount of funding annually for both funding organizations from 2006 to 2017. The amount of funding for each project in each year of the NSFC and the MRC was converted to the United States Dollars (USD) at the exchange rate of the apply year for the project.

Burden of disease

The burden of disease data used in the comparative analysis was sourced from the Global Burden of Disease (GBD),[8] which is a comprehensive regional and global research program of disease burden that assesses mortality and disability from major diseases, injuries, and risk factors. GBD is based at the Institute for Health Metrics and Evaluation (IHME), University of Washington, and also institutionalized at the World Health Organization (WHO). With the aim of measuring disability and death from a multitude of causes worldwide, the GBD study attributes each death to a single underlying cause that began the series of events that ultimately led to death. The GBD structures named causes of death in a four-level hierarchy, but the names and nature of each cause accord with the International Classification of Diseases (ICD) (GBD 2017 Causes of Death Collaborators 2018). The disease burden at the highest level (Level 1), divides into three mutually exclusive and collectively exhaustive categories: communicable, maternal, neonatal, and nutritional

[8] https://ghdx.healthdata.org/gbd-results-tool.

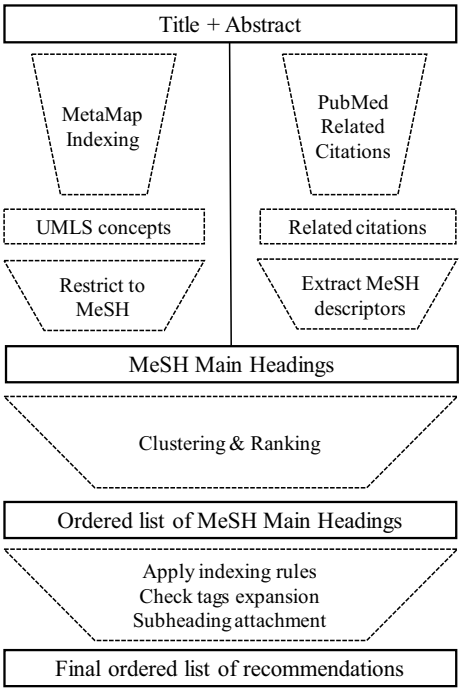

Fig. 3 MTI process flow diagram (Mork et al., 2013)

(CMNN); non-communicable diseases (NCDs); and injuries. Level 2 distinguishes these Level 1 categories into 22 cause groups, such as cardiovascular diseases, neoplasms, and neurological disorders. Level 3 disaggregates these causes further.

In most cases, this disaggregation represents the finest level of detail by cause, such as stroke, lung cancer, and Alzheimer's disease. Level 4 further disaggregates some causes into even more detailed classifications. The above four hierarchies provided the classification for disease burden data. This study mostly focuses on the NCD and CMNN categories for further analysis.

Data processing

Extraction of subject terms

As indicated in Fig. 1, we used the National Library of Medicine (NLM) Medical Text Indexer (MTI)[9] to extract the corresponding MeSH terms from the abstracts and titles of the projects. Produced and maintained by the NLM as a controlled and hierarchically organized lexicon, the primary use of the MeSH vocabulary is to index, catalog, and search for biomedical and health-related information.

The MTI is the main product of the NLM's Indexing Initiative project and has been producing indexing recommendations based on the MeSH vocabulary since 2002. The design of the MTI uses the title and abstract of MEDLINE citations to extract MeSH terms. It is

[9] https://ii.nlm.nih.gov/MTI/index.shtml.

also capable of processing arbitrary biomedical texts to provide an ordered list of MeSH terms for use as the keywords of the funded projects. Figure 3 depicts the processing flow of the various components of the MTI system (Mork et al. 2013). Conceived as a means of indicating the characterizing power or "aboutness" of a given concept for a piece of text, the Metamap Indexing (MMI) operates as a ranking function. It is the product of a frequency factor and a relevance factor, which is essentially measured by MeSH Tree depth. The following steps are used to map concepts to MeSH terms.

The principle underlying PubMed Related Citations (PRC) organization is that the neighbors of a document are those documents in the database that are the most similar to it. The PRC algorithm considers term frequency (modeled as a Poisson distribution), inverse document frequency and document length when computing the similarity between documents. MTI currently uses two different methods for determining PubMed Related Citations (PRC) for the different types of text it is processing. If MTI is working with a MEDLINE citation and there are enough indexed PRC defined by the PubMed system, MTI uses that list of PRC. For a free form text and the situation of an insufficient number of indexed PRC, MTI will default to using the in-house TexTool implementation of PRC.[10]

After clustering and ranking, MTI provides an ordered list of MeSH main headings, subheadings, and check tags as a final result. We only kept the main headings, deleting the check tags and subheadings for two reasons. First, the main headings are the main descriptors or headings from the MeSH vocabulary and the most appropriate subject term for each funding project. Subheadings are only used to qualify the main headings, and check tags are a special type of the main heading that must be included for each article to indicate the species, sex, and age groups of the research. Second, limitations and errors may exist in any purely machine-generated data. Hence, we removed the other headings to reduce noise as much as possible. As the main focus of this study is on terms related to diseases, we further filtered the MeSH terms by their C-category and F-category/F03 branch, shown as process ② in Fig. 1. The C-category segregates diseases but does not include mental disorders; F03 is the branch that focuses on mental disorders.

Additionally, there were two chief reasons we chose to use the MeSH terms as recommended by MTI rather than the keywords provided by applicants. The first was that project keywords provided by applicants tend to be subjective, and they do not always accurately reflect the research area, especially for specific diseases. Secondly, the keywords are not standardized, and it is common for applicants to use different expressions for the same disease. In a co-word analysis, the bias this creates is problematic. MTI is the tool used at NLM for pre-processing all of the MEDLINE citations that are then indexed by human indexing staff in NIH. It has been around now for about 20 years and has steadily improved throughout the years. The MTI statistics for 2014 show that MTI's consistency with human indexers is comparable to the available studies on indexer consistency (Mork et al. 2017). MTI offers a basic standard for extracting MeSH terms from abstracts and titles, which then form a more precise and accurate basis of analysis.

Identification of high-burden diseases

Highly reliable and transparent health statistics are of great significance to policymakers and other stakeholders concerned with the development of public health care. The first step

[10] https://www.ncbinlmnih.gov/CBBresearch/Wilbur/IRET/TexTool/.

in identifying high-burden diseases is to select an appropriate index. A wide range of indicators has been developed to monitor and manage health initiatives, including life expectancy, infant mortality, prevalence rates for specific diseases, and many more indexes that reflect general health conditions (Murray 2007; Murray and Frenk 2008). This framework reflects the fact that health is indeed a complex notion and comprises several dimensions: fatalities, disabilities and quality of life, the prevalence of disease, the severity of diseases, etc. Moreover, it is a widely held belief that extending both the length and the quality of life are important goals (Gross et al. 1999). Hence, one of the main problems with constructing a composite measure of disease burden is ascertaining the appropriate balance between the duration of life and the quality of that life. One answer to this problem is "healthy-year equivalents" indicators and similar.

The WHO has promoted the Global Burden of Disease (GBD) concept for over a decade under its express mandate to report on health information. The GBD reports its results using a composite measure of morbidity, disability, and mortality that was developed in 1993 by the World Bank and the WHO (The World Bank 1993). The indicator, called disability-adjusted life years (DALYs), distills these three factors into a single number as a measure of the burden attached to that disease (GBD 2017 Causes of Death Collaborators 2018). This measurement is reached by calculating the time lost through premature death and the time lived in a state of less than optimal health, loosely referred to as "disability". One DALY may be considered the equivalent of one lost year of "healthy" life. A DALY can be calculated for a specific cause of death or disability as follows:

$$\text{DALY} = (\text{deaths} \times \varepsilon) + (\text{cases} \times \omega \times \text{duration})$$

where ε is the standard life expectancy at the age of death in years and ω is a disease-specific disability weight. In simple terms, a DALY is an expression of the number of deaths times the average years of life lost (YLL) added to the number of incidents multiplied by the average years lost to disability (YLD). The sum of DALYs across the population represents the burden of disease as a measurement of the gap between current health status and an ideal health situation where the entire community lives to an advanced age, free of disease and disability.

In this research, we mainly adopt the DALYs per 100,000 population as the standard value by which to identify the diseases with the highest burden and to calculate the burden growth rate of the UK and China. Also, % of total DALYs is adopted as the indicator of "level of burden". This measure represents the proportion of one specific disease's burden in relation to all diseases' burden in the GBD list for a country/region.

Data visualization and analysis

To address our three research questions, we divided our analysis into three parts: the analysis on funding topic of two public funding agencies, the comparative analysis between diseases with severe concern and high burden, and an analysis of specific diseases.

As the most common method to explore knowledge with subject terms in the scientific literature, co-word analysis has the merit of extending the object of analysis to patents, articles, newspapers (Ding et al. 2001) or, in this case, research funding applications. Hence, co-word analysis was used to extract funding topics in both the first two analyses. To outline the topic structure of funded research over the whole period required the development of a co-occurrence matrix of funded projects during 2006–2017. This action enabled a general analysis of the frequency of MeSH terms extracted directly from MTI,

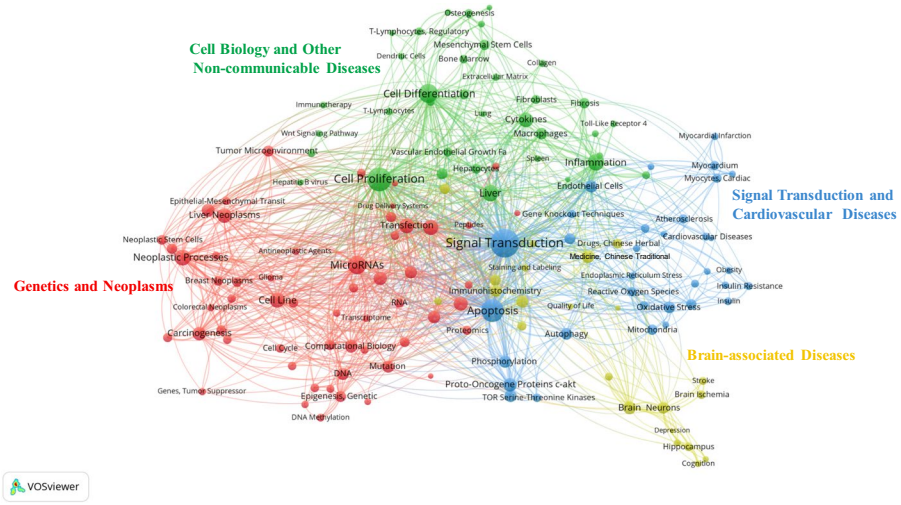

Fig. 4 The occurrence map of MeSH terms in NSFC-funded research during 2006–2017

which corresponded to process ① of the funding data. For the comparative analysis, we used terms filtered by C-category and F-category/F03 branch in process ②, as mentioned above. To present a relatively comprehensive and well-structured visualization, we considered the 150 most frequently used MeSH terms constructed as a symmetrical co-word matrix and maps. The Derwent Data Analyzer and VOSviewer were used to generate network maps and visualizations.

Various indicators were employed to provide a more in-depth analysis of the relationship between funding and burden level and to conduct further analysis of three specific diseases.

Results and analysis

Following the above research flow, this section presents three parts of the analysis required to answer our research questions.

Analysis of funding topics

Figure 4 shows four evident clusters with different research focuses formed by the top 150 most frequently funded MeSH terms between 2006 and 2017 by the NSFC. MeSH terms related to genetics (e.g., micro RNA, DNA) and neoplasms (e.g., carcinogenesis, neoplastic processes) constitute the largest cluster, which reflected the great attention paid by the NSFC to these two research fields and the strong connection between neoplasms with genetic research. The other two large clusters both concern basic cytology and cell biology, which indicated the high attention paid by the NSFC to fundamental research. Moreover, signal transduction, which means the transmission of molecular signals from a cell's exterior to its interior, has been the most-funded topic for the NSFC for the last 12 years. As

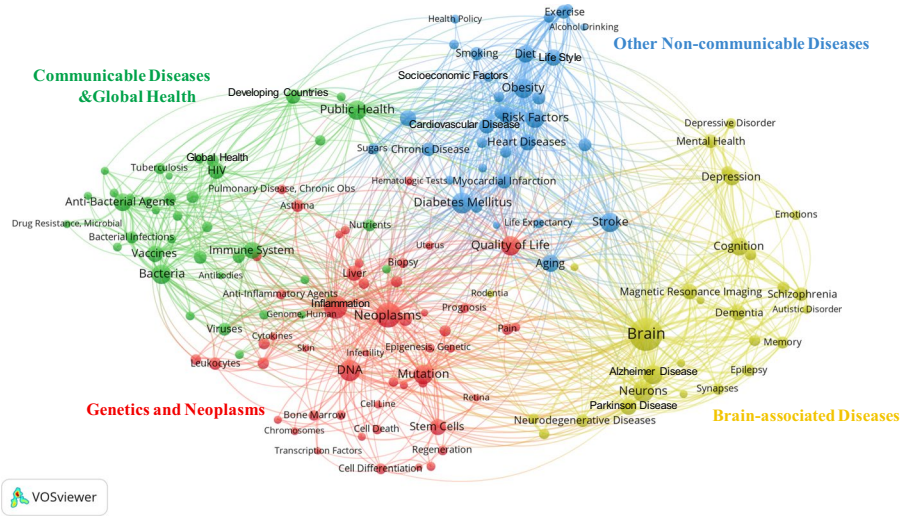

Fig. 5 The occurrence map of MeSH terms in MRC-funded research during 2006–2017

the most fundamental activity of cells, there is a close association between signal transduction and various diseases.

Another prominent cluster points to brain-associated diseases related to neurons and the hippocampus. Two particular items in this cluster that illustrate the application of traditional Chinese medicine treatments to brain-related diseases are "Medicine, Chinese Traditional" and "Drug, Chinese Herbal". In addition, the links between these two Chinese medical terms with MeSH terms related to genetics and cytology are quite strong. Therefore, we surmised that the presence of these terms indicates the NSFC has placed some importance on funding the traditional Chinese medical system, and this is likely due to a desire by the State Council to integrate Chinese medicine with Western medical protocols.

Figure 5 illustrates the topic structure of the MRC-funded projects. This shows some similarities and differences with the NSFC structure. The largest cluster of MRC projects is similar to that of the NSFC, which is mainly composed of MeSH terms related to genetics (e.g., DNA, mutation) and neoplasms. Another cluster akin to the NSFC range is the yellow one in Fig. 5, which also points to brain-associated diseases. The inclusion of this yellow cluster in Fig. 5 not only indicates the higher attention given to neurological and mental disorders by the MRC than the NSFC model. It also includes more specific diseases compared to that of the NSFC.

Some terms for risk factors (e.g., obesity, alcohol drinking, socioeconomic factors) and various diseases, such as cardiovascular disease (e.g., heart disease, myocardial infarction), diabetes (e.g., diabetes mellitus) form the blue cluster. This arrangement shows the MRC appears to be greatly concerned with lifestyle and societal factors that cause diseases, which is quite different from the NSFC approach. Another significant term of note in this cluster is "aging". According to the World Bank (2017), while the universally recognized standard for an aging society begins at around 7%, in 2006, more than 16% of the UK's population was over the age of 65. In comparison with the high percentage of an aged population in the UK, China, with an aged population of 7.84% in 2006, was only in the early stages of grappling with this problem. This factor is the most likely explanation for

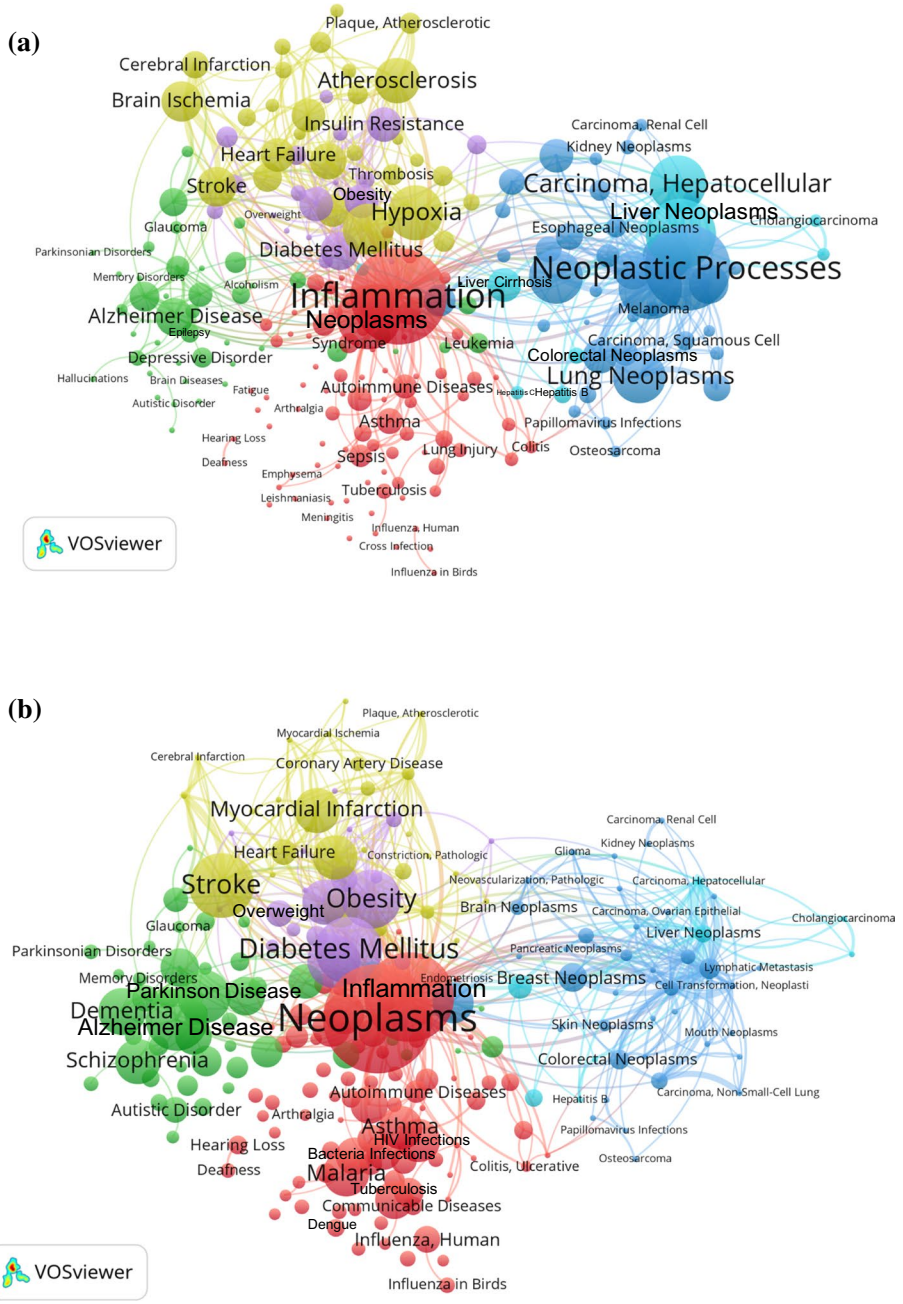

Fig. 6 Major funded diseases of the **a** NSFC Department of Health, **b** MRC

the close attention MRC has paid to diseases that coincide with age, as illustrated in the red cluster in Fig. 5.

Other differences between the funding models were evident. Unlike the NFSC funding clusters, the most visible MRC cluster focused on research relevant to infectious diseases caused by bacteria or viruses. The MRC also concentrated on global public health issues, especially various infections in developing countries.

Genetics and neoplasms are common interests for both the NSFC and MRC. The funding topics used by each agency also reveal other issues of importance for each country; traditional medicine for China and aging issues for the UK. Therefore, since the two agencies' strategic goals and development models differ, the emphasis in funding also differs. The primary focus for NSFC was on basic research, such as cell biology, while the MRC concentrated on disease-oriented research and related risk factors. Finally, the most profound difference between the topics funded by these two agencies was global health and communicable diseases, which do not feature on the NSFC map.

Comparative analysis between diseases of severe concern and high burden

Our comparative analysis used two steps to examine how the funding for significant disease categories correlated to diseases identified with either high burden or high burden growth characteristics. Initially, we presented major funded categories of the disease using visualizations of MeSH terms filtered by the C-category and F-category/F03 branch. The next step provided a comparative analysis of diseases with severe concern and high burden and high growth rate.

Diseases with high burden

Figure 6 shows the major funded disease categories used by the two main health funding agencies in China and the UK. Apparently, neoplasms, cardiovascular diseases, diabetes and neurological diseases are the common concerns of the NSFC and MRC. However, there are clear distinctions in that NSFC concentrates more on various types of neoplasms, while the MRC shows more emphasis on more diversified categories of diseases, including chronic respiratory diseases and communicable diseases.

In this section, we combined the top 5 categories of disease with the highest burden and major- diseases to conduct comparative analysis. Figure 7 illustrates the structure of the disease burden in China and the UK. The colored arc slices display the proportional average % of DALYs for diseases from 2006 to 2017 at each level. The internal circle identifies the top five categories of diseases classified at Level 2 of the GBD causes of death list, and the external circle indicates the Level 3 diseases disaggregated by their level 2 categories.

The top five level 2 categories of diseases, for the most part, were relatively similar for both countries, except for the chronic respiratory disease common to China and the neurological disorders suffered in the UK. Neoplasms, as the first and second category of diseases with the highest burden in the UK and China (Fig. 7), have taken a significant position in the funding allocations of both the NSFC and MRC (Fig. 6). What calls for special attention is the extremely intensive concern from the NSFC on different types of neoplasms showed in Fig. 6, such as lung and liver neoplasms—the two major neoplasms with the highest burden in China. Additionally, the NSFC paid special focus on different types of liver-related diseases, such as liver cirrhosis, hepatitis B and hepatitis C.

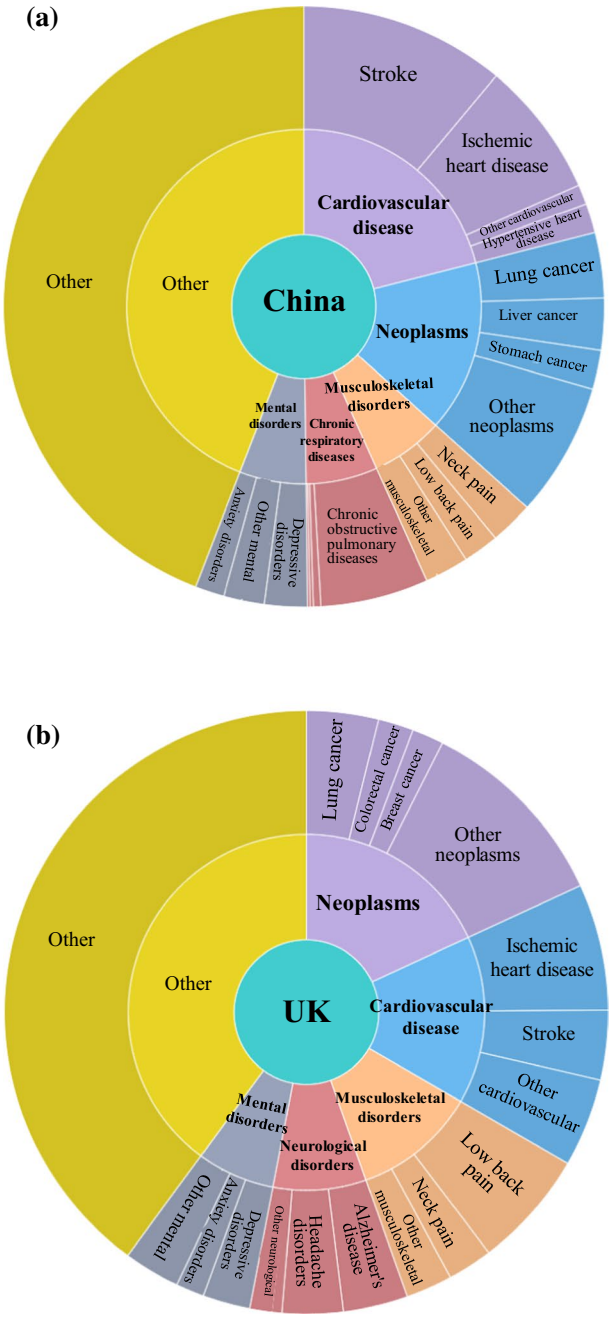

Fig. 7 Disease burden in **a** China, **b** the UK

Consistent with the high attention received from both the NSFC and MRC organizations, cardiovascular disease received a top two ranking in the structures of disease burden utilized in both countries. Further analysis also identified that stroke was not only the main contributor to cardiovascular diseases in both countries but was also the leading cause of

death and DALYs at the national level in China for the past 12 years. Ischemic heart disease also consistently ranked high as a cause of death in both countries. In Fig. 6, the high prominence of these illnesses reflects the great concern of two health agencies about the level of impact that stroke and other types of cardiovascular diseases have on the general health of populations in both countries.

Despite ranking third in both countries, another notable health concern, musculoskeletal disorders, received less attention from the NSFC and MRC than other illnesses. Within this category, the top two diseases specified as most relevant to the burden of diseases were neck pain and lower back pain. Further observation of the burden data of musculoskeletal disorders showed that the major contributor to the high rank of DALYs is the years lost due to disability (YLD) instead of the years of life lost (YLL). As noted earlier, DALYs are the sum of YLD and YLL. Although the musculoskeletal disorder category ranked lower on YLL, it occupied the first position on YLD in the full-time period for both China and the UK. Moreover, both neck pain and lower back pain ranked in the top five in terms of YLD at the national level. The smaller degree of attention received by musculoskeletal disorders may be due to their relatively high YLD instead of high death rates.

Another significant disease burden category, mental disorders, received a top-five ranking in both countries, and again as shown in Fig. 7, this area received more attention from the MRC than the NSFC. This category formed a single cluster with subsections for depressive disorders and schizophrenia and various types of neurological disorders. Regarding the status of specific ailments, neurological disorders were among the five diseases with the highest burden in the UK, and Alzheimer's disease, Parkinson Disease, dementia and schizophrenia were prominent in the MRC major funded diseases. Moreover, Alzheimer's disease and seizures were also received relatively closer attention from the NSFC even though neurological disorders were not on the top five diseases list in China.

"Inflammation" is another condition that has received sustained attention from both agencies. It has strong connections with many common chronic diseases, especially neoplasms. There is mounting evidence that some common chronic conditions are indeed triggered by low-grade, long-term inflammation (Shaw 2019). These disorders include Alzheimer's disease, cancer, arthritis, asthma, Parkinson's disease, diabetes, and depression, among others. News reports from Sciencedaily.com indicate that chronic inflammation is associated with up to 25% of all cancers. For example, the long-term and chronic infection of helicobacter pylori may increase the risk of stomach neoplasms, while chronic hepatitis may increase the risk of liver neoplasms (ScienceDaily 2011).

Diseases with high burden growth rate

In the next stage of our analysis, we explored trends and dynamics. As previously noted, Level 1 on the GBD cause of death list is divided into communicable, maternal, neonatal, and nutritional diseases (CMNN); non-communicable diseases (NCDs); and injuries. Because the injuries category mainly refers to transportation and intentional/unintentional injuries rather than specific diseases, we did not include it in this study. Level 2 further divides the Level 1 categories into 22 cause groups with 7 in the CMNN category and 12 in the NCD category. (The remaining three groups pertain to injuries.)

Figure 8 charts the percentage change in the burden of various diseases over time. Generally speaking, the trends for China, UK, and the world are similar. However, the changes in China are somewhat magnified, i.e., an increase in a disease burden saw a more significant degree in China, and likewise, the reverse applied for any decrease. During the study

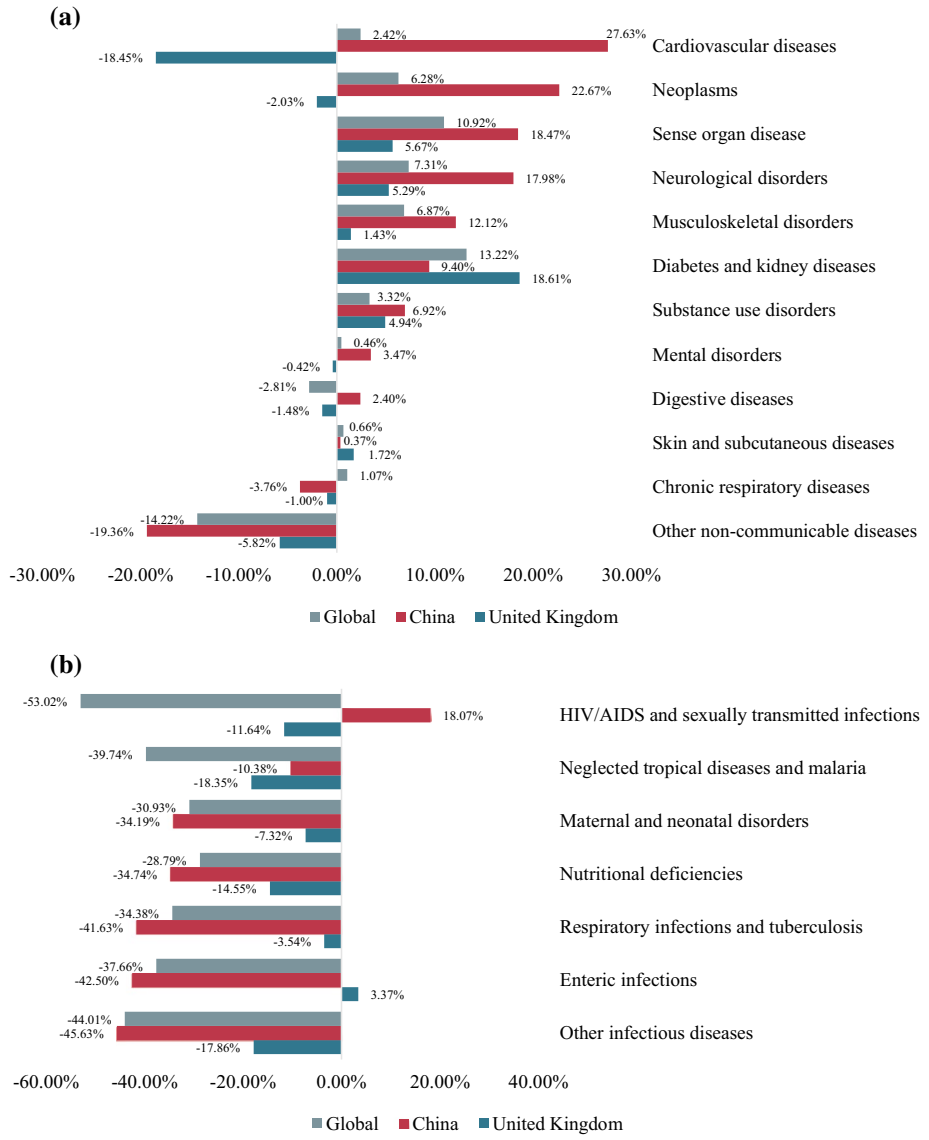

Fig. 8 **a** Growth rate of NCD causes' burden from 2006 to 2017 (DALYs per 100,000, both sexes and all ages). **b** Growth rate CMNN causes' burden from 2006 to 2017 (DALYs per 100,000, both sexes and all ages)

period, there was a substantial decrease in the burden of CMNN causes, while the burden generated by NCD factors generally increased. Stroke and ischemic heart disease were the leading causes of all-age DALYs in 2017 in China, overtaking lower respiratory infections and neonatal disorders in 1990.

From the standpoint of a macroscopic perspective, some authors have found rapid and sustained economic growth and increasing levels of educational attainment have likely

Table 2 The funding and burden level of cardiovascular diseases and neoplasms

	Cardiovascular disease (H02) (%)	Neoplasms (H16) (%)
% of funded projects (NSFC)	6.32	18.80
% of amount funding (NSFC)	6.40	18.98
Average % of total DALYs in China	21.03	15.63

contributed to the lower burden of communicable diseases in China (Gakidou et al. 2010; Zhu 2012). A range of national programs implemented to target interventions may also have contributed to the changes in the structure of China's disease burden (Wang et al. 2016). However, there has been an increase in the burden of cardiovascular diseases, neoplasms, and musculoskeletal diseases categorized as NCD elements with the rapid development of society and medicine's unceasing progress, especially in China. Transforming social structures and lifestyles mean increases in the number of work hours spent in sedentary occupations, which is a significant contributor to musculoskeletal and mental problems. However, neither musculoskeletal nor mental issues are the most pressing concerns of the NSFC.

In terms of funded research relevant to disease, both the NSFC and MRC paid much more attention to NCD rather than CMNN causes, which is consistent with the variations in burden. Diabetes is a prime example of a non-communicable disease that has increased its profile in both burden and research attention since 2006. Both health agencies in China and the UK consider the disease as closely linked to "obesity" and that the growing incidence of both conditions is related to the changing living and dietary habits of people. Diabetes is one of the most common research concerns for both the NSFC and the MRC (Fig. 6).

During the same period, CMNN causes related to the global burden of disease, and the national burdens of China and the UK, saw a significant decrease, except for HIV/AIDS and STDs in China and enteric infections in the world. According to Fig. 6, the communicable disease category is one of the most apparent differences between the funding focuses of the NSFC and MRC. The MRC has placed heightened attention on a variety of infections, including HIV/AIDs, malaria, tuberculosis, Dengue fever, leishmaniasis and Chagas disease, which have all attracted more concern year by year. The Commission on Macroeconomics and Health divides diseases into three types according to the national income level and the burden of disease (World Health Organization (WHO) 2012a). Type I includes diseases of incidence in both developed and developing countries. Type II also includes diseases that affect both development categories but substantially more so in developing countries. Type III refers to diseases that are overwhelmingly or exclusively present in developing countries. For example, tuberculosis and diarrhea are considered to be Type II diseases, while malaria is a Type III disease. Our findings show that the MRC placed more emphasis on funding disease research in the Type II and III categories than the NSFC. Correspondingly, the global burden of communicable causes of death has decreased over the last decades. The largest contributors to this decrease include reduced DALY rates for HIV/AIDS, tuberculosis, diarrhea, and malaria (GBD 2017 Causes of Death Collaborators 2018), all of which have been consistent topics of funding for the MRC.

In short, both the NSFC and the MRC are greatly concerned with neoplasms and cardiovascular diseases, which correspond to the top two families of diseases with the highest burden. Musculoskeletal disorders, especially neck pain and low back pain, have also

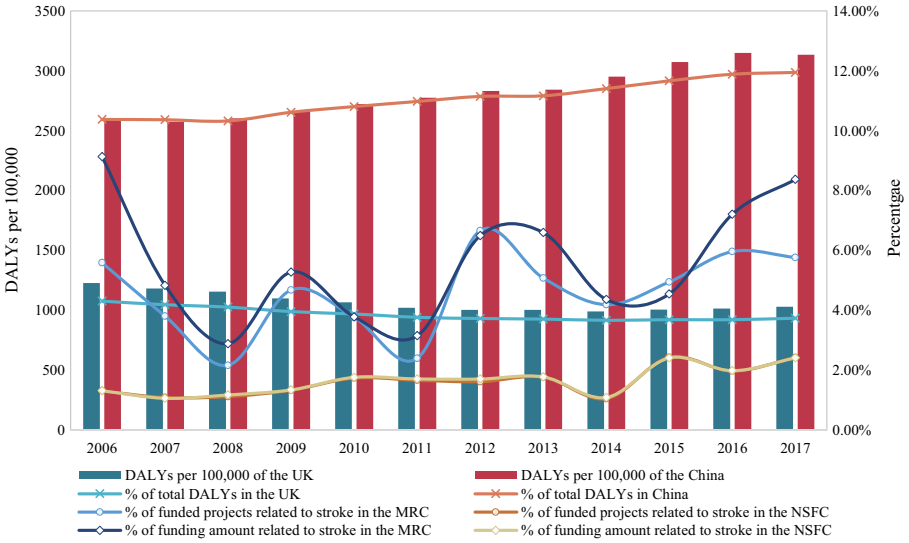

Fig. 9 The funding and burden level of stroke in China and the UK

scored with a high burden for both China and the UK in the last 12 years, but have so far received scant attention from either agency. Although it is a disease category ranked last in the top five list of Level 2 diseases, mental disorders have received relatively more attention from the MRC than the NSFC. Alzheimer's disease has still attracted considerable funding from health agencies in both countries, even though neurological disorders only appear among the five diseases with the highest burden in the UK. One possible explanation is the growth in the aging population that both countries are experiencing.

Analysis of specific diseases

The previous analysis was completed mainly from an aggregate perspective to draw comparisons between the funded topics and various families of diseases with high ranks or growth rates of burden. In this section, we provide a more in-depth analysis of three specific diseases in an attempt to investigate the relationship between levels of funding and disease burden.

Stroke and cardiovascular diseases

The stroke is a disease, which has seen a significant increase in burden since 1990 and has the highest DALY burden for all ages. Stroke is classified as a type of cardiovascular disease by the GBD. This category has the highest burden index and the highest burden growth rate in China (see Fig. 7a and Fig. 8a). In the previous analysis, we found that the NSFC had paid considerable attention to various cardiovascular diseases. However, from further calculations, we established that the number of funded projects and funding amounts invested in cardiovascular disease research over the last 12 years is relatively small compared to that of neoplasms (see Table 2). More specifically, three times more projects have been funded under the Department of Health Sciences' H16

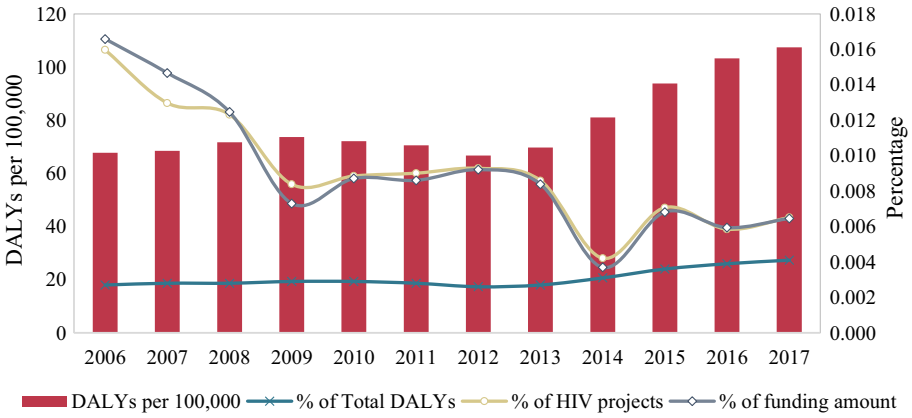

Fig. 10 The funding and burden level of HIV/AIDs in China

category of neoplasms than the H02 category of circulatory systems (which mainly refer to cardiovascular diseases). Yet the % of total DALYs for cardiovascular disease is 5.4% higher than neoplasms.

Figure 9 shows the percentage of projects and funding amount related to stroke-funded research by two agencies and the burden of stroke for two nations. The number of research projects related to stroke was calculated by counting the number of projects containing "stroke" in its recommended MeSH terms by MTI (as explained in the term interpretation of Fig. 1). All the terms containing "stroke" were verified manually to confirm that each term was relevant to the disease of stroke. The percentage of projects was calculated by dividing stroke-related projects with the number of total funded projects for each year. Also, the stroke-related amount of funding of each year was established by summing the funding amount of each stroke-related project. Similarly, the percentage of funding amount was determined by dividing the stroke-related amounts with the total funding amount for each year. Here we use the % of total DALYs and percentage of projects and funding amount to represent the level of burden and funding, respectively.

The bars and lines with cross in Fig. 9 represent DALYs per 100,000 and % of DALYs of stroke, respectively. It is quite obvious by the two burden indexes that the population of China suffers far more from stroke than the UK. As for levels of research funding in the UK, the proportion of MRC stroke research funding has fluctuated greatly but has stayed at a level higher than the % of DALYs of stroke. The percentage of funding for stroke research of the MRC shares a similar trend with the rate of projects. For the NSFC, the two lines of the percentage of projects and funding amount coincide, which actually conforms with the fact that General Programs in the NSFC granted similar funding amounts for each project per year, and the total investment over the years has increased proportionately. There is an overall upward trend in the percentage of stroke-related projects and funding amount granted by the NSFC, but this was only a third of the level of the MRC's number and far lower than the % of DALYs of stroke in China. However, it is worth noting that the term "stroke" appeared in the NSFC major funding diseases (Fig. 6a), which indicates the NSFC has paid a certain level of attention to this disease. As the leading cause of DALYs in China since 2006, our finding suggests that the NSFC should probably be placing a higher priority on research into strokes.

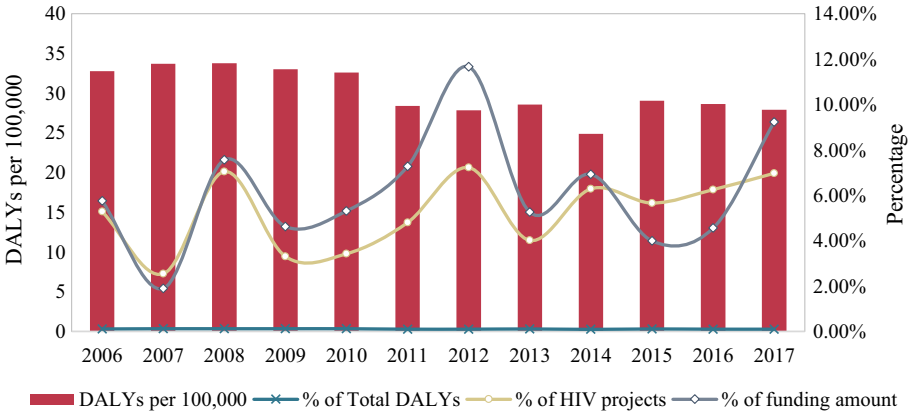

Fig. 11 The funding and burden level of HIV/AIDs in the UK

HIV/AIDS and STDs

HIV/AIDs and sexually transmitted diseases (STDs) comprise the only family of communicable diseases that shows an increasing burden in China, while the burden tends to be decreasing in the UK and the world, according to Fig. 8b. It is striking that neither the topic structure map nor the major disease funding map of the NSFC expressed severe concerns about HIV/AIDS and STDs or even communicable diseases. Yet, the maps for the MRC did.

Multiple factors may influence the change in burden for HIV/AIDs and STDs, and it is difficult to argue for any causality between changes in burden and the number of relevant funded projects. Moreover, there is no direct match between HIV/AIDs or STDs and a specific funding category set by the Department of Health Sciences of the NSFC, except for one called "sexually transmitted infections" (H1910).[11] At 0.08%, there were few NSFC-funded projects found in this category (32 of 38,214). This scenario somewhat reflects a lack of consideration toward STDs by the NSFC and possibly even medical academia. However, as this category only refers to "sexually transmitted infections", it means that some research projects related to HIV/AIDs may not be included in H1910.

To further investigate this possibility, we conducted a detailed search of the funded projects relating to HIV/AIDs. Similar to the method used above, the number of funded projects and funding related to HIV/AIDs was calculated based on the extracted MeSH terms for each project with further manual confirmation. As shown in Fig. 10, both the percentage of projects funded and the proportion of funding from the NSFC has seen an alarming downward trend with obvious drops between 2008 to 2009 and 2013 to 2014. There was a slight increase in funding from 2009 to 2012, but the DALYs per 100,000 for these diseases decreased during this period. Compared to the decreasing burden of these diseases across the world and in the UK, China has seen a sharp increase in DALYs per 100,000 caused by HIV/AIDs since 2013. In China, the three main subpopulations affected are drug users (Wang et al. 2015), female sex workers (Wang et al. 2014), and men who have sex with men (Chow et al. 2014). More than

[11] https://www.nsfc.gov.cn/publish/portal0/tab558/.

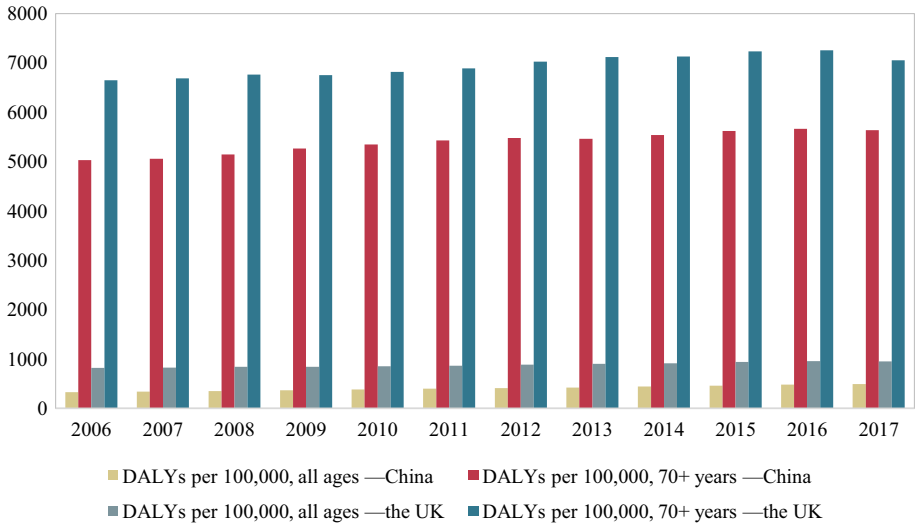

Fig. 12 The burden of Alzheimer's disease of different age groups in China and the UK

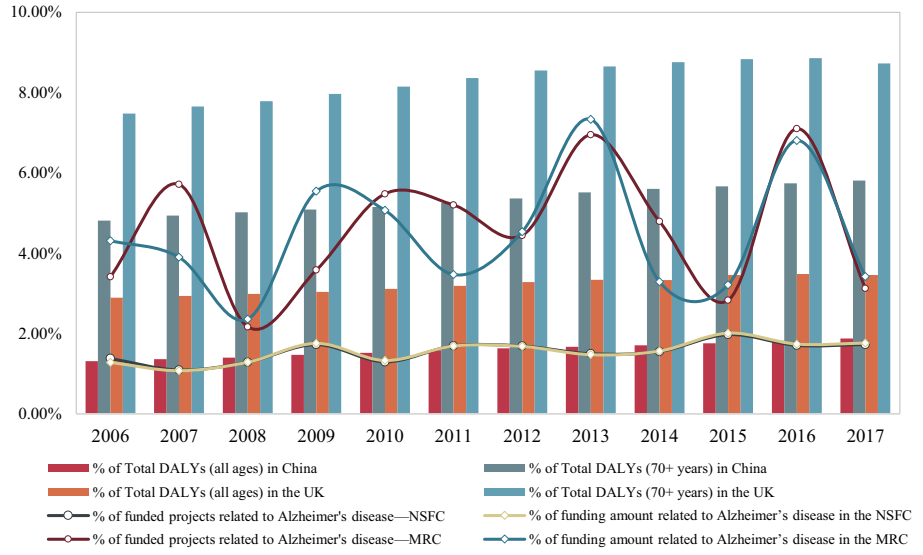

Fig. 13 The funding and burden level of Alzheimer's disease in China and the UK

90% of new HIV/AIDs infection incidences were transmitted through sex from January to October in 2014, according to a news report by Chinanews.com and data from the National Center for AIDS/STD Control and Prevention of China (China News Service (CNS) 2014). The risk of contracting HIV/AIDs through same-sex relations between men in China is exceptionally high, with a prevalence rate that rose from 5.73% in 2010 to 7.75% in 2014 (Cui et al. 2016).

The above analysis may lead to the conclusion that the NSFC has given too little attention to HIV/AIDs. However, it seems like another story from the perspective of % of total DALYs. Although the % of total DALYs also indicates a rising trend, especially from 2012, the overall % of total DALYs is lower than the average percentage of NSFC-funded projects and funding amount related to HIV/AIDs. There may also be a time lag between funding allocations and when the annual directory of projects is published. Taken together, the NSFC has paid adequate attention to HIV/AIDs in terms of % of total DALYs of HIV/AIDs. However, the statistic of increasing burden and decreasing investment provided above should also sound some alarm bells to both the National Government and Chinese society.

As indicated by Fig. 11, both DALYs per 100,000 and % of total DALYs of HIV/AIDs in the UK were maintained at a relatively low level during 2006–2017. The level of funding (both percentage of projects and funding amount) related to HIV/AIDs in the MRC is much higher than the level of burden in the UK, which conforms with the results from the previous analysis. Although HIV/AIDs is not a disease group that imposes a significant burden on the UK, HIV and its related illnesses have attracted attention from the MRC.

Alzheimer's disease

Classified as a Level 3 disease by the GBD, Alzheimer's is a chronic neurodegenerative disease that afflicts the elderly. Although neurological disorders only appeared among the five diseases with the highest burden in the UK, Alzheimer's disease has still attracted considerable funding from both agencies. Figure 12 shows the distinct difference in the burden of Alzheimer's disease between different age groups in China and the UK. Overall, the burdens of Alzheimer's disease for these groups in both countries increased steadily from 2006 to 2016. Only in the last year has it shown a slight downward trend. Furthermore, Fig. 12 shows that China's burden is lower than that of the UK for both the "all ages" and 70+ years' groups.

As for the level of burden and funding, the % of total DALYs for Alzheimer's diseases in both China and the UK share similar features (see Fig. 13) with the trend of DALYs per 100,000 in Fig. 12. Interestingly, Fig. 13 indicates that the percentages of projects and funding related to Alzheimer's disease funded by the NSFC are quite consistent with the % of total DALYs for all ages group in China. In comparison, allocations by the MRC are somewhere in between the % of total DALYs for the all-ages group and the 70+ years group in the UK. Both China and the UK had more than a 7% aging population early in 2006 (the universally recognized standard for an aging society) (The World Bank 2017). The following 12 years witnessed a further consistent increase in the aging population ratio, reaching 10.64% and 18.52%, respectively, in the total populations of China and the UK (The World Bank 2017). This analysis indicates that the MRC funding level is higher than the NSFC's and that this situation is likely due to the greater severity of the aging problem facing the UK.

Further to the above points, although Alzheimer's disease did not appear on the lists of diseases with the highest burden in China, its burden increased more than 50% from 2006 to 2017. Moreover, Alzheimer's disease ranked second in the UK and fourth in China in 2017 for the population aged 70 and above. The relatively high attention paid to Alzheimer's disease by both the NSFC and the MRC illustrates that the socio-demographic structure is one of the most important factors for setting funding priorities in the health science field.

In summary, the above analysis is a first attempt at exploring whether funding agencies have paid enough attention to specific diseases based on the different types of burden indexes and the level of funding. It is difficult to reach an absolute judgment or conclusion since multiple factors may have had an impact on past funding decisions and policies. However, our analysis does raise a question about whether governments and funding agencies should concentrate more on diseases with high burdens or diseases with high burden growth rates. Further, some diseases are experiencing a rate of increase but a decreasing % of total DALYs, which raises another question about which type of index should be adopted as representative of the real disease burden to a country based on standard values or proportions. Both these questions need to be studied further with a more accurate measurement of statistical data.

Conclusion and discussion

In this work, we explored the focus of disease-related research projects supported by two national funding organizations and juxtaposed their relative levels of funding with the burden of disease. Our analysis compared China and the UK using co-word and network analysis on funded project data from China's NSFC and the UK's MRC, accompanied by an in-depth analysis of three specific diseases.

With these methods, we identified the following similarities and differences between the topic structures of the NSFC and MRC. The NSFC has focused more on basic research, such as cell biology and genetics, while the MRC has concentrated more on disease-oriented research, related risk factors and public health. Our results also indicated that the burden of diseases played a significant role in both funding schemes. Both the NSFC and the MRC have devoted funding to the top two diseases with the highest national burden—neoplasms and cardiovascular disease. Diabetes and kidney disease have also received substantial attention, which corresponded to an increasing burden associated with these conditions evident in both countries from 2006 to 2017. Moreover, funding allocations placed more emphasis on diseases with a higher mortality rate instead of those with higher morbidity and disability. For example, one of the top five diseases with the highest burden in both countries, musculoskeletal disorders, received much less attention from both the NSFC and MRC.

One of the major differences between the types of projects sponsored by the NSFC and the MRC is the variety of diseases. MRC-funded research topics span more kinds of diseases compared to the NSFC. Of particular interest is the fact that the MRC has placed heightened attention on a variety of infections, especially on HIV/AIDs and some neglected tropical diseases that bear high burdens in some developing countries, while the NSFC has put the majority of its efforts towards non-communicable diseases within China.

Possible explanations for this scenario could be the different funding requirements of each organization. Only Chinese scholars can apply for the NSFC General Program, while an MRC Research Grant can be awarded to scholars from all over the world. Another critical point is the steadily increasing burden of HIV/AIDs in China in contrast with the decreasing burden related to these conditions in the UK and the world. However, we feel the decline in attention on HIV/AIDs by the NSFC, along with an increasing incidence in STDs in the population, should sound alarm bells for both China's government and wider society.

Our findings also reveal that the results obtained from the comparative analysis were affected by the burden indexes used for the assessments. Moreover, other social factors, such as socio-demographic structure, may have affected funding priorities. Therefore, we have not presumed that a direct connection exists between disease burden reduction and research efforts. Beyond putting in more research effort, a disease burden can continue to be alleviated through implementing strategies for prevention, diagnosis, and treatment of a disease, especially for diseases where mature research already exists. Further studies on various categories of diseases with different research strategies are also required to provide more in-depth insights.

Assessing the societal impact of scientific research is a critical issue for both academia and policymakers. Societal impact is about making science useful to society, which raises the fundamental question—are we conducting the most suitable research given the existing societal needs? Specifically, are scholars doing the research that society really needs? And, are funding agencies supporting scientific projects that respond to societal demands? However, there are difficulties and challenges with working out how to accurately measure the relationship between research efforts and public needs. While our study shows that it is useful to use scientometric approaches and data to face such challenges, it will be important to further explore multi-source data and tools from other fields. This paper, then, provides an exploratory analysis with funding data and disease burden data and serves as an attempt to investigate the relation between societal needs and scientific investment. Indeed, our paper is more of an observational than a causal analysis. Follow-up research and more explicit deliberations on priority settings are needed to embrace a broad perspective of research governance that considers underlying drivers (Wallace and Ràfols 2018). With the improving performance and increasing robustness of MTI, the methods and tools used in this study may be further applied to analyze the allocation of research efforts on different diseases in a broader range.

Limitations and future research

This study contains three data-related limitations that need to be addressed. Our comparative analysis only included data from the division of scientific departments within the NSFC and the division of Research Councils in the UK, where medical and health research and bio-scientific research are divided and managed separately. We drew our data corpus from the Department of Health Sciences in the NSFC (China) and the MRC (the UK), which manage medical and health-related research in their respective countries. However, we acknowledge that some relevant research projects may have been assigned to the agencies responsible for bio-scientific research, i.e., the Biotechnology and Biological Sciences Research Council (BBSRC) in the UKRI (the UK) and the Department of Life Sciences in the NSFC (China).

The second data-related limitation of this study is that we did not link the MeSH vocabulary with the principles of the ICD. Therefore, the previous results cannot be analyzed in the context of a unified classification. Further collaborative research between medical science and informetrics would be necessary to produce such a system.

The third limitation of this study refers to the selection of research object-focused public funding agencies. Although private funding is not common in China, and relevant data is usually not available, it is recognized that private funding is a critical component of scientific funding in global terms. This form of funding plays an important role in supporting

medical research and may offset the cost of national health burdens. It is our view that further research is needed into the different focuses of public and private funding to provide a more comprehensive understanding of the relationship between these types of funding and societal demands.

Acknowledgements The present study is an extended version of an article presented at the 17th International Conference on Scientometrics and Informetrics, Rome (Italy), 2–5 September 2019 (Zhang et al. 2019). The authors would like to acknowledge support from the National Natural Science Foundation of China (Grant Nos. 71573085; 71974150), the National Social Science Foundation of China (Grant No. 18VSJ087) and the National Laboratory Center for Library and Information Science in Wuhan University. We thank Ismael Rafols (Ingenio (CSIC-UPV), Universitat Politècnica València) for inspiring discussions.

References

Agarwal, P., & Searls, D. B. (2009). Can literature analysis identify innovation drivers in drug discovery? *Nature Reviews Drug Discovery, 8*, 865–878.

Anderson, R. S. (1999). Patrick Blackett in India: Military consultant and scientific intervenor, 1947–72. Part two. *Notes and Records of the Royal Society, 53*(3), 345–359.

Atala, I., Trinquart, L., Ravaud, P., et al. (2018). A mapping of 115,000 randomized trials revealed a mismatch between research effort and health needs in non-high. *Journal of Clinical Epidemiology, 98*, 123–132.

Begum, M., Lewison, G., Jassem, J., et al. (2018). Mapping cancer research across Central and Eastern Europe, the Russian Federation and Central Asia: Implications for future national cancer control planning. *European Journal of Cancer, 104*, 127–136.

Begum, M., Lewison, G., Wright, J. S. F., et al. (2016). European non-communicable respiratory disease research, 2002–13: Bibliometric study of outputs and funding. *PLoS ONE, 11*(4), e0154197.

Blackett, P. M. S. (1971). *Report to the Leverhulme Trust*.

Cassi, L., Lahatte, A., Rafols, I., et al. (2017). Improving fitness: Mapping research priorities against societal needs on obesity. *Journal of Informetrics, 11*(4), 1095–1113.

China News Service (CNS) (Producer). (2014). *87,000 new HIV/AIDs cases were report from January to October in 2014 in China, more than 90% of new HIV/AIDs infections incidences were transmitted through sex* (**in Chinese**). Retrieved August 2, 2019, from https://www.chinanews.com/gn/2014/12-01/6832929.shtml.

Chow, E. P. F., Lau, J. T. F., Zhuang, X., et al. (2014). HIV prevalence trends, risky behaviours, and governmental and community responses to the epidemic among men who have sex with men in China. *BioMed Research International, 2014*, 1–19.

Ciarli, T., & Ràfols, I. (2019). The relation between research priorities and societal demands: The case of rice. *Research Policy, 48*(8), 949–967.

Cui, Y., Guo, W., Li, D., et al. (2016). Estimating HIV incidence among key affected populations in China from serial cross-sectional surveys in 2010–2014. *Journal of the International AIDS Society, 19*(1), 20609.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management, 37*(6), 817–842.

Ebadi, A., & Schiffauerova, A. (2016). How to boost scientific production? A statistical analysis of research funding and other influencing factors. *Scientometrics, 106*(3), 1093–1116.

Evans, J. A., Shim, J.-M., & Ioannidis, J. P. A. (2014). Attention to local health burden and the global disparity of health research. *PLoS ONE, 9*(4), e90147.

Gakidou, E., Cowling, K., Lozano, R., et al. (2010). Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. *Lancet, 376*(9745), 959–974.

Gao, J., Su, C., Wang, H., et al. (2019). Research fund evaluation based on academic publication output analysis: The case of Chinese research fund evaluation. *Scientometrics, 119*(2), 959–972.

GBD 2017 Causes of Death Collaborators. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet, 392*(10159), 1736–1788.

Gillum, L. A., Gouveia, C., Dorsey, E. R., et al. (2011). NIH disease funding levels and burden of disease. *PLoS ONE, 6*(2), e16837.

Gross, C. P., Anderson, G. F., & Powe, N. R. (1999). The relation between funding by the National Institutes of Health and the burden of disease. *New England Journal of Medicine, 340*(24), 1881–1887.
Hagenaars, N., de Kruif, T., van de Laar, L., et al. (2019). The relationship between publication volume of biomedical research and burden of disease. *OSF Preprint*. https://doi.org/10.31219/osf.io/jeuar.
Halffman, W., & Leydesdorff, L. (2010). Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. *Minerva, 48*(1), 55–72.
Hicks, D. J., Stahmer, C., & Smith, M. (2018). Impacting capabilities: A conceptual framework for the social value of research. *Frontiers in Research Metrics and Analytics, 3*, 24.
Hicks, D. J., & Stapleford, T. A. (2016). The virtues of scientific practice: MacIntyre, virtue ethics, and the historiography of science. *Isis, 107*(3), 449–472.
Kalita, A., Shinde, S., & Patel, V. (2015). Public health research in India in the new millennium: A bibliometric analysis. *Global Health Action, 8*(1), 27576.
Kinge, J. M., Roxrud, I., Vollset, S. E., et al. (2014). Are the Norwegian health research investments in line with the disease burden? *Health Research Policy and Systems, 12*(64), 1–8.
Klavans, R., & Boyack, K. W. (2017). The research focus of nations: Economic vs. altruistic motivations. *PLoS ONE, 12*(1), e0169383.
Mitchell, R. J., McClure, R. J., Olivier, J., et al. (2009). Rational allocation of Australia's research dollars: Does the distribution of NHMRC funding by National Health Priority Area reflect actual disease burden? *Medical Journal of Australia, 191*(11–12), 648–652.
Morillo, F. (2019). Collaboration and impact of research in different disciplines with international funding (from the EU and other foreign sources). *Scientometrics, 120*(2), 807–823.
Mork, J. G., Aronson, A., & Demner-Fushman, D. (2017). 12 years on—Is the NLM medical text indexer still useful and relevant? *Journal of Biomedical Semantics, 8*(1), 1–10. https://doi.org/10.1186/s13326-017-0113-5.
Mork, J. G., Yepes, A. J. J., & Aronson, A. R. (2013). The NLM Medical Text Indexer System for indexing biomedical literature. *BioASQ@ CLEF*.
Murray, C. J. L. (2007). Good practices for health statistics: Lessons from the Millennium Development Goal health indicators. *Lancet, 369*(9564), 826–873.
Murray, C. J. L., & Frenk, J. (2008). Health metrics and evaluation: Strengthening the science. *Lancet, 371*(9619), 1191–1199.
Mutapi, F. (2019). Africa should set its own health-research agenda. *Nature, 575*(7784), 567.
National Natural Science Foundation of China. (2019). *National natural science fund guide to programs 2019*. Retrieved August 27, 2019, from https://www.nsfc.gov.cn/english/site_1/pdf/NationalNaturalScienceFundGuidetoPrograms2019.pdf.
National Science Foundation. (2018). *Proposal & award policies & procedures guide*. Retrieved Feburary 21, 2020, from https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_3.jsp#IIIA2.
Prüss-Üstün, A., Mathers, C., Corvalán, C., et al. (2003) Introduction and methods: Assessing the environmental burden of disease at national and local levels. In *Environmental burden of disease series No. 1*. Geneva: World Health Organization (WHO).
Ràfols, I., & Yegros, A. (2017). Is research responding to health needs? Available at SSRN. https://doi.org/10.2139/ssrn.3106713.
Røttingen, J.-A., Regmi, S., Eide, M., et al. (2013). Mapping of available health research and development data: What's there, what's missing, and what role is there for a global observatory? *Lancet, 382*(9900), 1286–1307.
ScienceDaily (Producer). (2011). How inflammation can lead to cancer. *Science News*. Retrieved July 7, 2019, from https://www.sciencedaily.com/releases/2011/04/110419091159.htm.
Shaw, J. (2019). Raw and Red-Hot: Could inflammation be the cause of myriad chronic conditions? *Harvard Magazine*. Retrieved July 13, 2019, from https://harvardmagazine.com/2019/05/inflammation-disease-diet.
Shibayama, S. (2011). Distribution of academic research funds: A case of Japanese national research grant. *Scientometrics, 88*(1), 43–60.
Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*(9), 1568–1580.
Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology, 68*(3), 790–794.
The World Bank (Ed.). (1993). *World development report, 1993. Investing in health*. New York: Oxford University Press for the World Bank.
The World Bank. (2017). *Population ages 65 and above (% of total population)*. Retrieved June 16, 2019, from https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?end=2017&locations=CNGB&start=1960&view=chart.

Viergever, R. F., Olifson, S., Ghaffar, A., et al. (2010). A checklist for health research priority setting: Nine common themes of good practice. *Health Research Policy and Systems, 8*(1), 36.

Wallace, M. L., & Ràfols, I. (2018). Institutional shaping of research priorities: A case study on avian influenza. *Research Policy, 47*(10), 1975–1989.

Wang, L., Guo, W., Li, D., et al. (2015). HIV epidemic among drug users in China: 1995–2011. *Addiction, 110*(01), 20–28.

Wang, L., Tang, W., Wang, L., et al. (2014). The HIV, syphilis, and HCV epidemics among female sex workers in China: Results from a serial cross-sectional study between 2008 and 2012. *Clinical Infectious Diseases, 59*(1), 1–9.

Wang, X., Liu, D., Ding, K., et al. (2012). Science funding and research output: A study on 10 countries. *Scientometrics, 91*(2), 591–599.

Wang, Y., Li, X., Zhou, M., et al. (2016). Under-5 mortality in 2851 Chinese counties, 1996–2012: A subnational assessment of achieving MDG 4 goals in China. *Lancet, 387*(10015), 273–283.

World Health Organization (WHO). (2012a). *Background document provided by the WHO Secretariat-Defining disease types I, II and III*. Retrieved June 16, 2019, from https://www.who.int/phi/3-background_cewg_agenda_item5_disease_types_final.pdf.

World Health Organization (WHO). (2012b). *Research and development to meet health needs in developing countries: Strengthening global financing and coordination*. Retrieved July 23, 2019, from https://www.who.int/phi/cewg_report/en/.

World Health Organization (WHO). (2017). *WHO global observatory on health research and development (R&D)—Terms and conditions of use.* Retrieved April 14, 2019, from https://www.who.int/research-observatory/portal/Global_RD_Observatory_Conditions_of_Use.pdf?ua=1.

Xie, Y., Zhang, C., & Lai, Q. (2014). China's rise as a major contributor to science and technology. *Proceedings of the National Academy of Sciences of the United States of America, 111*(26), 9437–9442.

Yegros, A., Van de Klippe, W., Abad-Garcia, M. F., et al. (2019). *Exploring why global health needs are unmet by research efforts: The potential influences of geography, industry, and publication incentives*. Available at SSRN: https://ssrn.com/abstract=3459230. https://doi.org/10.2139/ssrn.3459230.

Zhang, L., Zhao, W., Liu, J., et al. (2019). Do national funding organizations address the diseases with the highest burden adequately?—Observations from China and the UK. In *Proceedings of ISSI 2019—The 17th international conference on scientometrics and informetrics* (pp. 826–837). Rome: Sapienza University.

Zhao, X., Gao, X., & Pe, H. (2009). The h-index of science funding: Comprehensive characterization of the quantity and impact of funded papers. *Bulletin National Nature Science Foundation of China, 23*(1), 15–18.

Zhou, P., & Tian, H. (2014). Funded collaboration research in mathematics in China. *Scientometrics, 99*(3), 695–715.

Zhu, X. (2012). Understanding China's growth: Past, present, and future. *Journal of Economic Perspectives, 26*(4), 103–124.