doi.org/10.26398/asaproc.0056

Health differences across population subgroups: Exploring inequalities through bibliometric techniques

Anna Paterno [a], Maria Gabriella Grassia [b], Thaís García Pereiro [a], Rocco Mazza [a]
[a] Department of Political Science, Aldo Moro University of Bari, Bari, Italy
[b] Department of Social Science, Federico II University of Naples, Naples, Italy

1. Introduction and research aims

Several studies have identified the presence of a socio-economic gradient on both health outcomes and health care access in almost every country of the world. This gradient is acknowledged as a downward slope on differences in health outcomes or health care access across population subgroups and defined as health inequalities. Health inequalities can be guessed as outcomes of a complex combination of individuals' characteristics and their chances to access resources. The former regards the conditions and experiences in which individuals are born, live and work—such as their gender, education, income, social networks, migratory status, and social relations. The latter, instead, deals with their real chances to access resources and decision-making processes, in terms of access to social protection (such as affordable child services or housing, sickness and unemployment protection; etc.), access to quality health services and prevention measures; access to healthy housing and settlement, or access to financial and non-financial services, among others.

Discrimination most often affects women, older people, people with disability, or are based on ethnicity or sexual identity. Discrimination has often a personal basis but may also be caused by social bodies, and this could imply that whole population receive inferior services or has difficulties in accessing health services, which, in turn, could prevent them from enjoying healthier lives.

Previous bibliometric analyses regarding the scientific production on health inequalities have been focused either on a specific region or country or groups of countries (Benach de Rovira, 1995; Almeida-Filho et al., 2003; Ritz et al., 2010), or on health systems or reforms (Macias-Chapula, 2002; Yao et al., 2014). Other recent literature focused on citation practices, together with the most productive authors and journals in health inequalities (Bouchard et al., 2015); north-south gaps in research and international collaborations (Cash-Gibson et al., 2018); and citation space and roles of several factors on health structure (Collyer and Smith, 2020).

As health is a fundamental human right, identifying health inequalities and its main drivers remains essential to achieve health equity. Health equity is achieved when everyone can attain their full potential for health and wellbeing. Research on health inequality is one of the principal sources of knowledge for policy and planning in aged and multicultural modern societies (WHO, 2019).

The main purposes of this paper are, first, to describe the temporal evolution of the amount of academic production focused on health inequalities/disparities/equity during an extended period 1991-2022 and, second, to identify its main research topics and map the specific roles of these topics within the academic production.

2. Data collection and methodology

In this paper we employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model proposed by Moher et al. (2009) to retrieve scientific publications (Figure 1). PRISMA offers a comprehensive framework that outlines the criteria for selecting articles in a systematic literature review, ensuring transparent and reproducible selection processes through three distinct phases: the identification of the queries for the search of the documents, the screening of the texts collected through specific filters and the application of these and then the inclusion of the papers in the corpus to be analyzed.

To retrieve the articles, we accessed the Web of Science (WoS) indexing database, specifically the expanded Science Citation Index (SCI) and the Social Science Citation Index (SSCI) WoS Core Collections maintained by Clarivate Analytics. The query used for document search in the WoS database was the following: health AND inequalit* OR health AND inequit* OR health AND disparit*. The tag searches for the query terms in the titles, abstracts, and keywords fields of indexed documents. Quotation marks were used to retrieve records with exact term sequences, while asterisks served as wildcards to capture term variations.

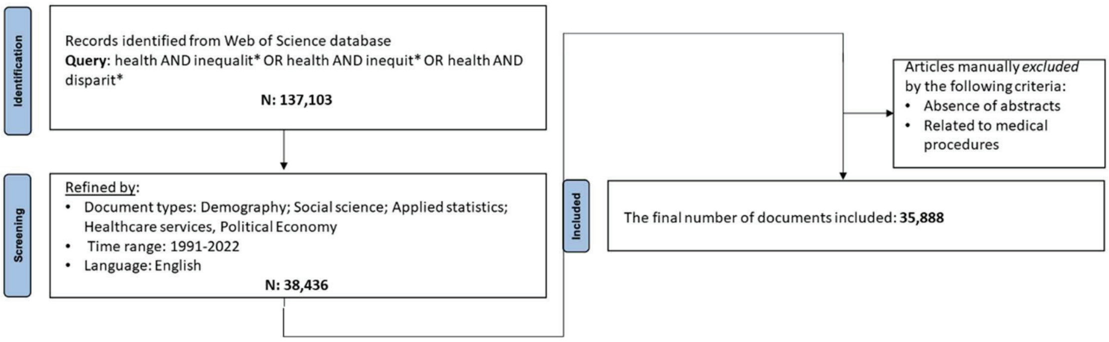
Data were collected in early July 2023. We refined the search by selecting only articles, proceedings papers, review articles, and book chapters published in English between 1991 and 2022, based on the relevance of their content. We collected bibliographic data, including titles, abstracts, author names, keywords, and cited references. The documents were exported to PlainText format and screened by two selectors to include only relevant and coherent documents. After excluding records without abstracts and those focusing on detailed medical procedures or practices, a total of 35,888 documents were retrieved.

To conduct the analysis, we employed the bibliometrix R open-source package (Aria and Cuccurullo, 2017; Aria et al. 2022), which facilitates quantitative research in scientometrics and bibliometrics. We utilized bibliometric analysis to examine the conceptual structure of publications within a specific scientific field, enabling the generation of clusters that provide a comprehensive overview of the research in the field (Borner et al., 2003). To explore the conceptual structure, we performed two complementary analyses: co-occurrence network analysis and thematic mapping. These approaches facilitated the identification of relationships among terms, key research themes, and their development. The degree of similarity between publications was determined by the extent of shared keywords, indicating their association within the same research field. Co-occurrence network analysis (Wang et al., 2019) specifically captured themes represented by sets of terms extracted from documents, such as author and journal keywords. This technique quantified the frequency of term co-occurrence in the document collection and normalized the results using the association index proposed by Van Eck and Waltman (2009). The resulting co-occurrence matrix was represented as an undirected weighted network.

Community detection, performed using the Walktrap algorithm (Pons and Latapy, 2006), identified strongly linked groups of terms sharing common characteristics or playing similar roles within the network. We employed thematic mapping, a two-dimensional representation of network findings proposed by Cobo et al. (2011), to visualize the identified themes. The x-axis represented Callon centrality, indicating the level of significance that a theme holds within a research field, while the y-axis represented Callon density, reflecting the degree of theme development (Callon et al., 1983). This combination of measures facilitated the identification of four types of topics based on their location on the map. The first quadrant of the map represents highly significant and well-developed motor themes. The second quadrant encompass isolated or niche themes with limited external links, resulting in low centrality and limited importance for the broader research field. The third quadrant captures emerging or declining themes, indicating weak or marginal development. The fourth quadrant identifies basic and transversal themes that cut across different research areas.

Each theme was represented as a network cluster on the map, with the bubble name indicating the word with the highest occurrence within the cluster, and the bubble size representing the proportion of word occurrences within the cluster. This way, our study effectively mapped the conceptual structure of the collected scientific documents, thereby revealing significant research topics and trends in the field of health inequalities.

Figure 1: Data retrieval. PRISMA chart for health query.



3. Preliminary results

The information extracted from the utilized library allowed us to outline an overview of the analyzed corpus. There are 4,640 documents, a relatively high number considering the applied filters which might be suggesting a high number of journals are interested in the theme. The average number of citations per document amounts to approximately 28, indicating a prominent activity in the community about this field. The total number of authors is 79,534, with 4,089 documents being single authored.

Figure 2 displays the growth of scientific production over the examined time interval for this analysis, with an overall annual growth rate of 20.46%. The graph shows that the growth in scientific production has significantly growth recently. In fact, until 2008 the number of published papers was 5,007 and, since 2009, in the last 14 years, more than 30 thousand papers on this subject have been published.

Figure 2: Frequency of the scientific production about health, 1991-2022

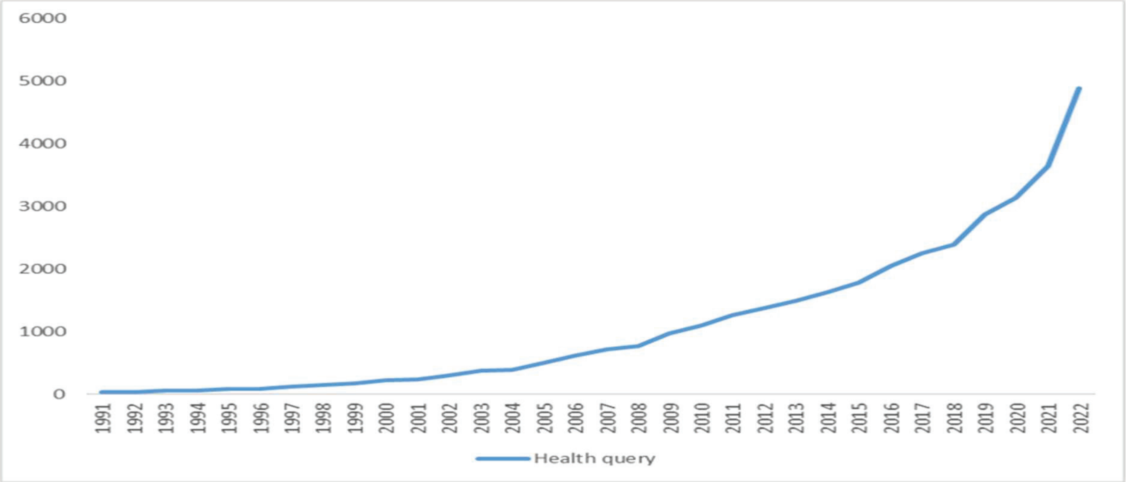


Figure 3: Thematic map of the scientific production on health

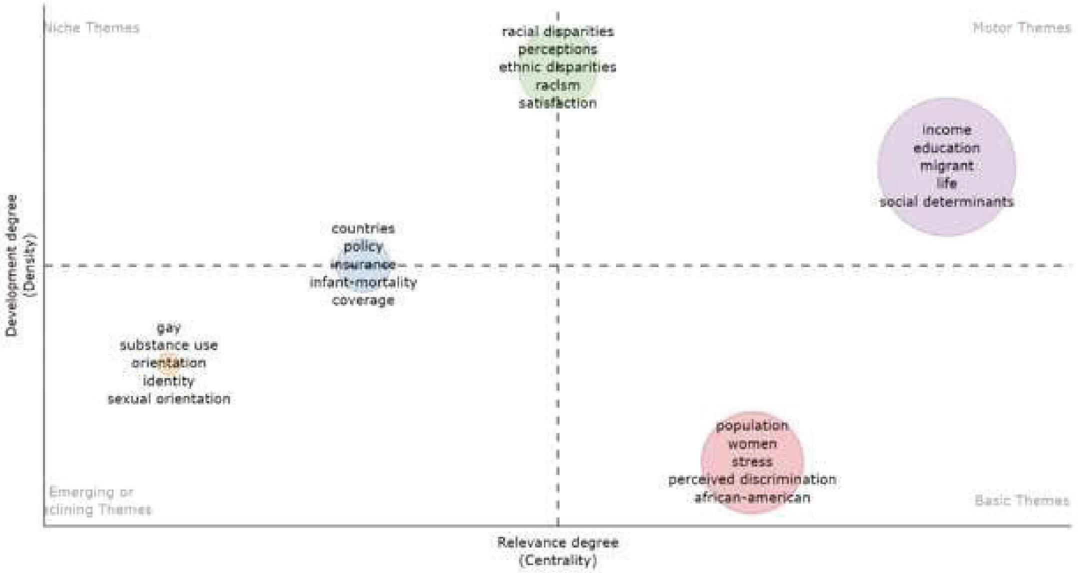


Figure 3 shows the five themes extracted from the papers; the area of a cluster represents the frequency of the cluster. To support the interpretation of results we included Table 1, which displays the values describing network measures that are calculated by this method. The cluster with the highest frequency (35%) is labelled with the keywords “income, education, migrant, life, social determinant”. It encompasses motor themes, with high density and centrality. In the second cluster (23%) we find the labels “population, women, stress, perceived discrimination, African-American”, which represent basic themes with a transversal dictionary, low density but high centrality. The third cluster (18%) with the labels “racial disparities, perceptions, ethnic disparities, satisfaction” is difficult to classify because it is between motor and niche themes, it is very high in density but is ranked third by degree. The same difficulty is encountered when analyzing the fourth cluster (14%) labelled “countries, policy, insurance, infant-mortality, coverage”. This cluster is between niche and emerging or declining themes. The last cluster (10%) with the keywords “gay, substance abuse, orientation, identity, sexual orientation” seems to represent emerging or declining themes. To clearly identify what the cluster is actually representing, it would be necessary to analyze the time trend through a longitudinal analysis.

Table 1: Network measures for each cluster.

N.	Cluster	Callon Centrality	Callon Density	Rank Centrality	Rank Density	Cluster Frequency
1	population, women, stress, perceived discrimination, African-American	0.033	0.37	4	1	23%
2	countries, policy, insurance, infant-mortality, coverage	0.017	0.40	2	3	14%
3	racial disparities, perceptions, ethnic disparities, satisfaction	0.020	0.47	3	5	18%
4	income, education, migrant, life, social determinant	0.035	0.41	5	4	35%
5	gay, substance abuse, sexual orientation, identity	0.014	0.38	1	2	10%

4. Conclusion and future research

This study has confirmed that socioeconomic inequalities in health mostly has socio-demographic determinants (gender, education, income, migratory status, sexual orientation), but are also related to health care systems coverage. Also, it highlighted the mediation role of health systems, which can act as buffers for inequalities, while supporting the reduction not only of the gap observed across groups of the population but also of its consequences at both the private and public spheres.

Translating the results of these studies into policies remains a main challenge. It emerges the need to look for better and innovative forms of knowledge integration between researchers, policy makers and stakeholders. Once health inequalities were measured (identified, described and analyzed) and the corresponding actions and strategies developed by national and local governments and international organizations, then, it is time to measure its impact.

Further research on this subject -through a bibliometric approach- must deal not only with the identification of the main policy guidelines and interventions proposed and evaluated by the scientific production but also with its evolution over time and across countries.

References

Almeida-Filho, N., Kawachi, I., Filho, A.P., Dachs, J.N.W. (2003). Research on health inequalities in Latin America and the Caribbean: Bibliometric analysis (1971-2000) and descriptive content analysis (1971-1995). *American Journal of Public Health*, **93**(12), pp. 2037-2043.

Aria, M., Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, **11**(4), pp. 959-975.

Aria, M., Cuccurullo C., D’Aniello L., Misuraca M., Spano M. (2022). Thematic analysis as a new culturomic tool: The social media coverage on COVID-19 pandemic in Italy. *Sustainability*, **14**(6) pp. 36-43.

Benach De Rovira, J. (1995). Análisis bibliométrico de las desigualdades en salud en España (1980-1994). *Gaceta Sanitaria*, **9** pp. 251-264.

Borner, K., Chen, C., Boyack, K.W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, **37**(1), pp. 179-255.

Bouchard, L., Albertini, M., Batista, R., De Montigny, J. (2015). Research on health inequalities: a bibliometric analysis (1966-2014). *Social Science & Medicine*, **141**, pp.100-108.

Callon, M., Courtial J.P., Turner W.A., Bauin S. (1983). From translations to problematic networks:

An introduction to co-word analysis. *Social Science Information*, **22**(2), pp. 191- 235.

Cash-Gibson, L., Rojas-Gualdron, D.F., Pericas, J.M., Benach, J. (2018). Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015), *PloS one*, **13**(1): e0191901.

Cobo, M. J., Lopez-Herrera A. G., Herrera-Viedma E., Herrera F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, **62**(7), pp. 1382-1402.

Collyer, T.A., Smith, K.E. (2020). An atlas of health inequalities and health disparities research: ”How is this all getting done in silos, and why?”. *Social Science & Medicine*, **264**, pp. 113-330.

Macias-Chapula, C., Mijangos-Nolasco, A. (2002). Bibliometric analysis of AIDS literature in Central Africa. *Scientometrics*, **54**(2), pp. 309-317.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group (2015). Linee guida per il reporting di revisioni sistematiche e meta-analisi: il PRISMA Statement. *Evidence*, **7**(6): e1000114.

Pons, P., L atapy M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, **10**(2), pp. 191-218.

Van Eck, N., Waltman L. (2009). How to normalise co-occurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, **60**, pp. 1635-1651.

Wang, H., Zhao Y., Dang B., Han P., Shi X. (2019). Network centrality and innovation performance: The role of formal and informal institutions in emerging economies. *Journal of Business & Industrial Marketing*, **34**(6), pp. 1388-1400.

World Health Organization (2019). *Thirteenth General Programme of Work, 2019-2023: Promote Health, Keep the World Safe, Serve the Vulnerable (No. WHO/PRP/18.1)*. World Health Organization.

Yao, R., Zhang, W., Evans, R., Cao, G., Rui, T., Shen, L. (2022). Inequities in health care services caused by the adoption of digital health technologies: scoping review. *Journal of Medical Internet Research*, **24**(3): e34144.