Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Revealing the trends in the academic landscape of the health care system using contextual topic modeling

Muhammad Inaam ul haq [1], Qianmu Li [1]

[1]School of computer science and Engineering, Nanjing University of Science and Technology

Corresponding authors: Muhammad Inaam ul Haq (Minaamulhaq1@hotmail.com) and Qianmu Li (qianmu@njust.edu.cn)

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Revealing the trends in the academic landscape of the health care system using contextual topic modeling

Abstract

The health care system encompasses the participation of individuals, groups, agencies, and resources that offer services to address the requirements of the person, community, and population in terms of health. Parallel to the rising debates on the healthcare systems in relation to diseases, treatments, interventions, medication, and clinical practice guidelines, the world is currently discussing the healthcare industry, technology perspectives, and healthcare costs. To gain a comprehensive understanding of the healthcare systems research paradigm, we offered a novel contextual topic modeling approach that links up the CombinedTM model with our healthcare Bert to discover the contextual topics in the domain of healthcare. This research work discovered 60 contextual topics among them fifteen topics are the hottest which include smart medical monitoring systems, causes, and effects of stress and anxiety, and healthcare cost estimation and twelve topics are the coldest. Moreover, thirty-three topics are showing in-significant trends. We further investigated various clusters and correlations among the topics exploring inter-topic distance maps which add depth to the understanding of the research structure of this scientific domain. The current study enhances the prior topic modeling methodologies that examine the healthcare literature from a particular disciplinary perspective. It further extends the existing topic modeling approaches that do not incorporate contextual information in the topic discovery process adding contextual information by creating sentence embedding vectors through transformers-based models. We also utilized corpus tuning, the mean pooling technique, and the hugging face tool. Our method gives a higher coherence score as compared to the state-of-the-art models (LSA, LDA, and Ber Topic).

Keywords: Contextual topic modeling, health care Bert, content analysis, health care system

1. Introduction

Healthcare systems are intricate. These are composed of various interrelated parts. The World Health Organization (WHO) says a health system is "all institutions, people, and actions whose primary aim is to promote, restore, or maintain health." They usually include both rural and urban areas, public and private systems, formal/allopathic, and informal/traditional methods of providing healthcare, as well as being at the national level, giving them a tremendous scope. (Martins *et al.*, 2021).

There are many more functions that health systems do in society in addition to providing healthcare and other treatments to maintain or improve health. They aid in preventing financial fallout from illness and medical costs for homes. It's imperative to keep in mind that health systems contribute to society's economy. (Sachs, 2001). The health system is a sector of the economy that provides employment, revenue, and business opportunities for many health workers and enterprises. For instance, there is some research that suggests that a population's health may have an impact on economic productivity. A wider range of societal norms and values are established through health systems, which are also social and cultural institutions. (Gilson, 2003). Health systems and the broader environment interact dynamically. Because of their diffuse nature and usually porous borders, health systems must include the social, political, and economic environment while assessing their structure and efficacy.

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Moreover, the health systems are locations where actors with various wants and desires compete and argue. Setting health priorities, funding health systems, and allocating resources within the system are all contentious issues. The place of the state and the market within health systems, as well as the function that a health system should serve in society, are frequently the subject of ideological and political disagreement. These various facets of the complexity of health systems are rarely addressed simultaneously and are transdisciplinary. The fact is that numerous disciplinary perspectives, such as those of history, economics, medicine, epidemiology, politics, law, ethics, anthropology, and sociology, are necessary to fully study and comprehend health systems (Martins *et al.*, 2021).

The researchers in different studies applied bibliometric methods to analyze the healthcare literature (Jalali *et, al.*,2019; Rejeb *et. al*, 2021) or extract latent patterns from the scientific literature on various subjects of the healthcare system (Ali & Kannan, 2022; Mustakim *et al.*,2021; Dantu, Dissanayake & Nerur, 2021). For example, Ali et al. describe a mapping of the research on healthcare operations and supply chain management. Dantu et al. states the exploratory analysis of research on IoT in healthcare. These studies are limited to a specific disciplinary viewpoint and do not show the holistic picture of healthcare research. Moreover, these approaches do not capture the context of the discussion. In the current study through the application of computational methods and advanced topic modeling tools, we capture the context of the research so that the topics are more semantically understandable. This goal is achieved by utilizing a contextual word embedding-based topic modeling method. It uses sentences as the elementary unit of analysis for creating embeddings. Combining computational methods with qualitative data analysis, we provided highly objective, coherent, superior, and meta-analytical insight into current research on healthcare systems. This study's overall technical and theoretical contributions can be illustrated as follows:

(i) We developed a novel contextual topic modeling approach that incorporates corpus tuning and mean pooling techniques to design healthcare Bert which we link up with CMT(CombinedTM) to generate the contextual topic modeling in the domain of healthcare care.

(ii) We compare our model with LSA, LDA, and Ber Topic. Our model achieves a maximum coherence score as compared to these state-of-the-art models.

(iii) This study performs the quantitative assessment of scientific literature in the domain of health care in terms of contextual topics and classified the topics into hot and cold categories on the basis of p-values. We also investigated topic clusters and correlations by exploring inter-topic distance maps.

(iv) We also elaborated the top cited studies qualitatively to cross-validate the topic themes which enhance the significance of our findings.

The rest of the article is divided into six sections. The first section describes the introduction. The related work is covered in the second section, and the materials and methods are covered in depth in the third section. Results and discussions are covered in the fourth section, followed by a conclusion in the fifth section.

2. Literature review

Here, we discuss (i) the top-cited works on the healthcare system as well as (ii) related work on topic modeling approaches.

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

2.1 Healthcare systems

This section covers some overviews of highly cited research on healthcare systems. Significant research generally deals with diseases, diagnoses, interventions, treatments, and other related subjects. In the context of diseases, the scholars focus on heart disease stroke, congenital heart disease, and other vascular diseases (Marelli et al., 2007; Roger *et al.*, 2012; Heidenreich *et al.*,2013; Mozaffarian *et al.*,2015). Regarding healthcare interventions, scholars emphasize various subjects such as the acceptability of the interventions, evaluation of interventions, behavioral interventions, and care transition interventions (Liberati *et al.*,2009; Shea *et al.*,2014). The most cited research also analyzes medication effectiveness (Kakkar *et al.*, 2008) and stigma as the cause of health inequality, service utilization of lifetime mental disorder, cultural competence in the delivery of healthcare services, and patient perception of hospital care (Merikangas *et. al.*, 2011; Hatzenbuehlere *et. al*,2013).

The healthcare industry is undoubtedly the most significant of all the sectors that have profited from technological adoption. As a result, it eventually raised the standard of living and contributed to several life savings. The research scholars developed various tools and techniques to automate the various operations and tasks from a healthcare perspective. For example, a 3D slicer, a clinical research tool similar to a radiology workstation (Fedorov *et al*, 2012). WSN technologies have various applications in the health sector like sensor-integrated devices which can monitor human activity such as pressure, temperature, and strain. It also provides monitoring facilities through contextual information that minimizes the caregiver's needs (Alemdar & Ersoy, 2010; Trung & Lee, 2016). Additionally, the IoT-enabling solutions based on a WSN, RFID, and mobile technology can monitor patients, personnel, and healthcare devices are another application of technology for healthcare services (Catarinucci *et al.*,2015). Material sciences also have a significant contribution to the healthcare industry. For example, electrospun nanofibers can be used for membrane preparation (Ramakrishna *et al.*,2006). Silk-molded electronic skin can monitor the psychological signals of human beings (Wang *et al.*, 2014). From a healthcare perspective, a large amount of data is generated that can be processed using machine learning and deep learning techniques. It advances healthcare research and improves human health (Ludvigsson *et al.*, 2009; Miotto *et al.*,2018). Beyond these areas, the physician acceptance of telemedicine technology and cell phone-based interventions (voice, text messages) are being evaluated as an alternative to ordinary business settings (Chau & Hu, 2002; Krishna, Boren & Balas, 2009). Additionally, the usage of technology in the healthcare system positively impacts hospital revenue and the quality of services delivered to patients (Devaraj & Kohli, 2003). Video conferencing technology has also been applied in healthcare to train primary care providers to treat complex diseases like HCV infection, which increase the patient-treatment ratio (Arora *et al.*,2011).

Regarding healthcare costs and related subjects, researchers pointed out various factors that increase or decrease healthcare costs. For example, medication adherence, higher medical adherence decreases patient hospitalization (Ho, Bryson, & Rumsfeld, 2009; Sokol *et al.*, 2005). The implication of patient follow-up intervention (Jack *et al.*,2009) and fall-related injuries can overcome re-hospitalization risk and expenditure (Stevens *et al.*, 2006; Taylor, 2017; Reginster & Burlet, 2006). Surgical site infections (SSIs) are one of the major

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

contributors to healthcare-associated infections and contribute significantly to the damage in medical care through the over-length of stay at hospitals (Zhan, & Miller, 2003; Hidron *et al.*,2008; Zimlichman *et al.*,2013). Infections such as; Clostridium difficile infection and antibiotic-resistant bacteria threaten the healthcare system as these are the cause of various deaths (Angus *et al.*,2004; Klevens *et al.*, 2007; Lessa *et al.*,2015; Cassini *et al.*,2019). Therefore, the deployment of a surveillance system can provide an estimate of the burden of the infection. Additionally, prevention activities along with surveillance can avoid infections and overcome the burden of extra costs in the healthcare system (Magill *et al.*,2014). Cancer imposes various health and economic burdens in terms of its treatment which reflects a substantial increase that highlights the importance of cancer prevention efforts, which may result in future savings to the healthcare system. Therefore, research recommends early cancer detection and treatment for effective cancer control (Guy Jr *et al.*,2015; Siegel *et al.*, 2019).

The most cited themes also highlight the clinical practice guidelines frequently used as recommendations (Harris *et al.*,2001) such as guidelines related to diagnosis, prevention, intervention, treatment, and patient safety or care (Kimiko & Lowenstein,1985; Boyce, & Pittet, 2002; Chinn & Sehulster, 2003; Tablan *et al.*,2004; Jensen *et al.*,2005; Barlow *et al.*, 2007; Erasmus *et al.*,2010; McKhann *et al.*, 2011; McAlindon *et al.*, 2014; Muraro *et al.*,2014).

2.2 Topic modeling

The group of algorithmic machine learning techniques used in the field of text mining is called probabilistic topic models. These models look for structural patterns within a corpus to extract semantic data. The topic templates create word clusters representing the major subjects in a given corpus. These methods offer an automatic method of locating common topics in the papers that are being displayed in this manner. Topic modeling can be performed using various approaches that employ algorithms like NMF, LSA LDA, and clustering employing the K-means or Ward's method used for hierarchical clustering (Gurcan *et al.*, 2020; Principe *et. al*, 2021).

A variety of statistical and probabilistic approaches are used in language modeling (LM) to estimate the likelihood that a given string of words will appear in a sentence. Language models examine the corpora of text data to provide a foundation for their word predictions. These models are more efficient as compared to other approaches as they take into account the meaning and semantics of words and sentences as well as the relationship between words. Additionally, by using this strategy, we were able to achieve the highest level of semantic integrity within each topic, enhancing the topics' relevance and differentiating them from one another (Koroteev,2021). Moreover, google created a transformer-based machine learning method for pre-training natural language processing called Bidirectional Encoder Representations from Transformers Known as BERT. It was developed and released by Google employees Jacob Devlin and his team in 2018 (Devlin *et al.*,2018).

In the literature, we found various approaches that focused on the analysis of scientific discourses. The existing studies mostly concentrated on using bibliometric techniques (Jalali *et al.*,2019; Rejeb *et al.*, 2021), Latent Dirichlet Allocation (LDA) based topic modeling techniques (Principe *et al*, 2021; Cho,2019), or qualitative content analysis (Kim *et al*, 2021). The other qualitative methods also applied by the researchers in the scientific trajectory analysis include systematic mapping reviews, critical reviews, and narrative reviews (Pare &

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Kitsiou 2017). Researchers primarily developed keywords-based analysis techniques that do not usually capture the context of the discussion. The applications of traditional methods such as Latent Dirichlet Allocation and Probabilistic Latent Semantic Analysis are difficult given the high dimensionality of massive data. The problem also exists of unclear topics caused by the sparse distribution of topics (Shi *et al.,* 2019). Saheb et al. (2022) propose a context-based topic modeling approach. It integrates the general Bert model LDA and K-means clustering to contextually analyze research articles. The generalized Bert model is less efficient in detecting cluster quality and Kmeans is inefficient in generating clusters of data which has outliers. To solve the above challenges, we applied contextual topic modeling (Basmatkar & Maurya, 2022) in the domain of health care. We tuned the specialized Bert model (Medical-Bio-Bert2) on the corpus of healthcare research articles and then added a mean pooling layer on top of it giving us a novel HealthcareBert which we link with the CombinedTM (CTM) model to develop a novel contextual topic model. This strategy improves the coherence score, giving more accurate embeddings and resulting contextualized topics. We also compared our model with the state of art LSA, LDA, and Ber Topic models. Our model outperforms the existing models. We further investigated the topic trends and correlations among the contextualized topics to add more dimension to understanding the healthcare landscape.

3. Materials & Methods

In this section, we discuss the data collection, its pre-processing, and the methodology for the contextual topic modeling procedure in detail.

3.1 Data collection

We selected Scopus as a data source and the following query: TITLE-ABS-KEY ((healthcare system* OR health care system* OR Health-care system* OR healthcare* OR health care* OR Health-care*)) AND (LIMIT-TO (SRCTYPE, “j”)) AND LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (LANGUAGE, “English”)) is executed in this database on 30th November 2022. As a result, we obtained 29600 records having publication dates from 2000 to 2022. We removed duplicates and empty abstract articles from the dataset; the rest have 28036 records.

3.2 Pre-processing

After the data collection process was finished, pre-processing of the data was done before modeling to increase the data quality. The text in the data set was first divided into tokens using word tokenization. Then lowercase was applied to the tokens. The text was cleaned up by getting rid of the numerals, punctuation, and stop words. This was achieved using a typical English stop word list (n=153). The text is cleaned up using the stemming and lemmatization procedures.

3.3 Healthcare Bert

We develop a transformer-based deep learning model as Healthcare Bert to enhance the semantic understanding of the topics. We tuned fspanda-Medical-Bio-Bert2 available through the hugging faces tool on the healthcare corpus and generate an improved Transformers-based model to provide more accurate context vectors in contextual topic modeling. There are various pooling methods (e.g., CLS pooling, Mean pooling, Max pooling) for transformer models. (Zhao *et al* 2022). We added these three layers on top of HealthcareBert one by one and computed the coherence score through the CMT model. However, the addition of mean pooling

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

layer gives a higher coherence score so we added this layer on top of our HealthcareBert. (See algorithm -1)

Algorithm-1	
Input: Healthcare_papers_abstratcs, Pretrain fspanda-Medical-Bio-BERT2	
Output: Healthcare BERT	
1. For all Healthcare_papers_abstratcs in the dataset	
2.	Tokenize (Healthcare_papers_abstratcs) initializing it with fspanda-Medical-Bio-BERT2 instance
3.	Get the model object fspanda-Medical-Bio-BERT2
4.	Retrain/Tune this model on the Healthcare corpus
5.	Added mean pooling layer

3.4 Topic Models and CombinedTM

Latent Dirichlet allocation (LDA) is an important and widely used probabilistic topic model. It is based on a generative process denoted by the equation as follows:

$$p(\theta, z, w|\alpha,\beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta\ )\text{---(i)}$$

Since Dirichlet prior is not a location-scale family, to solve this issue in ProdLDA the decoder network is used to approximate the Dirichlet prior $p(\theta|\alpha)$ with a logistic normal distribution given by the equation where μ and ∑ represents the outputs of the encoder network as follows:

$$p(\theta|\alpha) \approx p(\theta|\mu, \textstyle\sum) = LN(\theta|\mu, \textstyle\sum)\text{-----------(ii)}$$

The encoder network has some disadvantages that it is stuck in a bad local optimum this problem is addressed using Adam optimizer, batch normalization, and dropout units in the encoder network. Another difference between LDA and ProdLDA is that β is unnormalized and w_n is defined as:

$$(\omega_n|\beta, \theta \sim Categorial(\sigma(\beta\theta))\text{---------------(iii)}$$

CombinedTM (CTM) is a contextualized topic model inspired by ProdLDA even though both of the models use the same hyperparameters. The original CTM model uses SBERT features in combination with the Bow (Bag of words). In the current study, we replaced SBERT with a novel Healthcare BERT. We fetched and processed the data set using Panda's data frame. The CombinedTM model is employed that integrates (Bianchi, Terragni & Hovy, 2021) contextualized embeddings and a bag of words model. Finally, contextual topics were generated through CombinedTM, and the sentence embedding vectors were constructed using the improved healthcare Bert we have developed in the current study as shown in (Fig.1).

3.5 Model Evaluation

For the assessment of the validity of our approach, we design the following criteria (i) quantitative assessment (Coherence score) and (ii) qualitative assessment (highly cited literature).

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

(i) Quantitative assessment (Coherence Score)

The selection of an optimal model from the list of generated models is a critical task. Human comprehension depends on the concept of semantic context, and the coherence method makes an attempt to determine the context between words in a topic. Maximising the coherence score is crucial because it provides subjects that are easier for humans to understand. This context cannot be captured by the other matrices (such as perplexity). As a result, we assess the model's performance using the coherence score (Pasquali, 2017). We generated various models with no. of topics k in the range of (10 to 100 with the step of 5) employing healthcare Bert with CombinedTM and plot the coherence score as (Coherence Cv) corresponding to each model in both cases. We choose the CombinedTM model that gives the maximum value of coherence Cv among the lists of generated models.

$$Coherence(V) = \sum_{(vi,vj)\epsilon V} score(vi, vj, \epsilon)\text{-------------------(iv)}$$

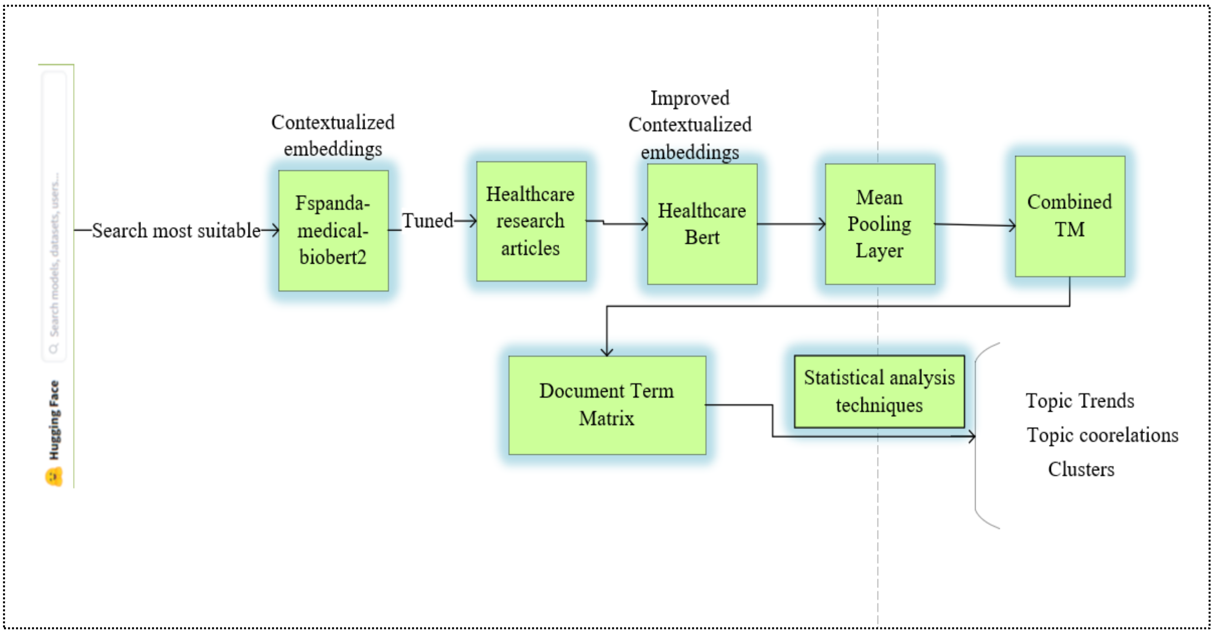

Figure.1: Contextual topic modeling framework

(ii) Qualitative assessment

For a thematic understanding of the discovered topics, we also adopted a qualitative approach to cross-validate the topic's themes. In this method, we summarize the highly cited literature in section (2.1) which facilitates the general assessment of topic themes.

3.6 Topic Trends and popularity measurement

In this study, every topic's trends are examined, and the posterior distribution is linked to the year that each document was published. Each document gives a certain topic to it that best represents its likelihood at that current time. By dividing the total number of papers each year by the number of papers assigned to this topic, the total number of papers was normalized to

https://doi.org/10.1162/dint_a_00217

determine the topic proportion for each topic each year. Following that, the cumulative proportion was used to calculate the overall popularity. Additionally, the Mann-Kendall trend test (M-K-test) is used to look for enduring upward or downward patterns in the data gathered over time. It is a non-parametric trend test method that examines discrepancies between earlier and later data points and is applicable to all distributions. It denotes that when a trend is present, sign values consistently tend to rise or fall (Wang et al 2020). Each topic's rising and declining patterns were recorded using the Mann-Kendall test.

3.7 Inter-topic distances and Topic correlations

We used LDAvis, a visualizing tool for topic models, to aid interpretations of the contextualized topics. This package plots the topics on a two-dimensional plane which gives us inter-topic distances, topic clusters, and resulting correlations.

4. Results and discussions

In this study firstly we develop a transformer-based deep learning model to enhance the semantic understanding of the topics. Google created various transformer-based machine learning methods for pre-training natural language processing which required large computational resources. So, we search first for the most suitable transformer-based model existing on the hugging face (fspanda-Medical-Bio-BERT2) that can provide word embeddings for our data set. To improve these word embeddings, we tuned this model further using Google GPU and Colab Notebook on the local corpus. Further, we added a mean pooling layer after corpus tuning. In this way, we generated a novel model as Healthcare Bert. These word embeddings are provided as context vectors to the CTM model that combine a bag of words and word embeddings to generate contextual topics of the corpus. The newly generated word embeddings also improve the topic's coherence score of the CTM models. CTM generates a document term matrix (DTM) which we further investigated using various techniques of statistics to classify the topics into hot and cold categories. Five subheadings—(i) models' evaluation/performance analysis, (ii) contextual topics, (iii) classification of topic trends, and (v) inter-topic distances topic clusters, and correlations are used to organize the study's findings.

4.1 Models' Evaluation/performance analysis

We programmed coherence matric (C_v) to compare the proposed model (CMT) with LSA, LDA, and Ber Topic. Table.1 shows the values of the coherence score (C_v) and no. of topics (K) applying our healthcare Bert with CTM and Ber Topic. We also computed the coherence score of other state-of-the-art methods (LSA and LDA) on our data set without our healthcare Bert. All values of the coherence score computed using different models are analyzed and the CTM model generated using our novel healthcare Bert with k=60 gives a clear maxima value and is chosen as an optimal model to report the contextualized topics as shown in (See Table.1).

Table.1 Performance Analysis of LSA, LDA, Ber Topic, and Our Method

Topic No	**LSA**	**LDA**	**Ber Topic**	**Our Method**
10	**0.3516**	0.4176	**0.4212**	0.5880
15	0.3363	0.4652	0.4183	0.5841

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

20	0.3210	0.4592	0.3852	0.5850
25	0.3142	0.4684	0.3781	0.5898
30	0.3119	0.4860	0.3539	0.5831
35	0.3078	0.4873	0.3554	0.5775
40	0.2976	0.4778	0.3482	0.5840
45	0.2932	**0.4979**	0.3375	0.5873
50	0.2912	0.4763	0.3373	0.5806
55	0.2849	0.4686	0.3414	0.5848
60	0.2827	0.4662	0.3272	**0.5978**
65	0.2747	0.4754	0.3370	0.5782
70	0.2739	0.4637	0.3297	0.5817
75	0.2674	0.4802	0.3251	0.5875
80	0.2676	0.4399	0.3205	0.5819
85	0.2650	0.4638	0.3294	0.5517
90	0.2601	0.4300	0.3342	0.5608
95	0.2562	0.4384	0.3309	0.5771
100	0.2563	0.4472	0.3332	0.5657

4.2 Context-based topics

This section lists the contextualized topics that our innovative healthcare Bert and the CTM model uncovered. The terms of each topic are supplied in (Appendix A) and explained in (See Table 2).

Table.2 Top 20 terms-based topic descriptions

Topic description	**Trends**	**Topic description**	**Trends**
01: 95 Ci for transplantation	↑↑	31: Asthma and costs	↓↓↓↓
02: 95 Ci testing	↑↑↑↑	32: Drug management and treatment	↑
03: Patient treatment cost analysis	↓	33: Hypertension diabetes and medication	↓
04: Cost gained strategy	↓↓	34: Global financing for public health	↑↑
05: Diabetes Intervention	↓	35: Technology adoption and its acceptance	↑
06: Health expenditure	↓	36: Nurses' job satisfaction	↓
07: Resistance in MRSA isolation and transmission	↓↓↓↓	37: Healthcare IoT	↑
08: Healthcare services	↑↑	38: Surgery and Complications	↑↑↑↑

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

09: Costs estimation	↑↑↑	39: Role of nursing as a healthcare professional	↓↓↓
10: 95 Ci for patient disparities	↑↑	40: Management of organizational processes	↓
11: Antimicrobial compliance	↓↓	41: Side effects of overwork	↓
12: 95 Ci hospital risks	↑↑↑↑	42: Smart medical monitoring system	↑↑
13: Patient treatment risks	↑↑↑↑	43: Coronavirus Investigation	↑
14: Home intervention for patient care	↓↓↓↓	44: Patient care quality and healthcare	↑↑
15: Brest cancer	↓	45: Care quality and management	↓↓↓↓
16: Hygiene compliance	↓↓↓	46: Needs of cancer patients	↓
17: Glucose control intervention	↓↓	47: Research on health policy	↑↑↑↑
18: Maternal inequalities	↑	48: Patient care quality	↓
19: Children's hospitalization rates	↓	49: Sexual substance and discrimination	↓
20: Healthcare Education	↓	50: Healthcare quality indicators	↓
21: HPV vaccination and immunization	↓	51: Healthcare devices	↑
22: Health pandemic and vaccination	↑↑	52: Women's stigma	↓
23: Reviews of healthcare research	↑↑	53: Healthcare Discrimination	↓
24: Infection-associated cost	↓	54: Telemedicine technology	↓↓↓↓
25: Pneumonia patients	↓	55: Physical and emotional effects on quality of life (QOL)	↓↓
26: Primary healthcare services	↑	56: Nursing and ethical policy	↓
27: Patient healthcare cost and services	↓	57: Consumer documentation	↓
28: Asthma	↓↓	58: Childhood disparities	↑
29: Hospital charges and costs	↓	59: Audit attendance category	↓
30: Causes and effects of anxiety and stress	↑↑↑	60: Inappropriate audit attendance	↓

Our contextual topic modeling approach generated 60 topics. Among these topics, 15 showed significant increasing trends. These topics mainly deal with smart medical monitoring systems (Topic-42), different health pandemics and the importance of vaccination to control them (Topic-22), causes and effects of stress and anxiety (Topic-30), and health care services generally (Topic-08). Hot topics also include healthcare cost estimations (Topic-09), patient treatment risks (Topic-13), research on designing healthcare policy (Topic-47), 95 ci hospital risks (Topic-12), patient care quality (Topic-44), global financing for public health (Topic-34), 95 ci testing (Topic-02), surgery complications (Topic-38), health care research reviews (Topic-23),95 ci for patient disparities and transplantation (Topic-10 and Topic-01). Twelve topics showed significant decreasing trends. These topics covering care quality and management (Topic-45), Asthma and costs (Topic-28 &Topic-31), Hygiene compliance

(Topic-16), Telemedicine technology (Topic-54), glucose control intervention (Topic-17), Antimicrobial compliance (Topic-11), Home intervention for patient care (Topic-14), Cost gained strategy (Topic-04), Physical and emotional effects on quality of life (QOL) (Topic-55), the role of nursing as a healthcare professional (Topic-39) and Resistance in MRSA isolation and transmission (Topic-07).

Beyond these topics, contextual topic modeling also discovered thirty-three more topics which were not shown significant increasing or decreasing trends but revealed various important subjects of the healthcare landscape. For example, infection-associated cost (Topic-24), patient treatment cost analysis (Topic-03), hospital charges and cost (Topic-29), Patient healthcare cost and services (Topic-27), children hospitalization rates (Topic-19), childhood disparities (Topic-58), Healthcare discrimination (Topic-53), Sexual substance and discrimination (Topic-49), HPV vaccination and immunization (Topic-21), investigation for coronavirus (Topic-43), Hypertension diabetes and medication (Topic-33), diabetes intervention (Topic-05), drug management and treatment (Topic-32), needs of cancer patients (Topic-46), pneumonia patients (Topic-25), nursing and ethical policy (Topic-56), nurses' job satisfaction (Topic-36), patient care quality (Topic-44&Topic-48), primary healthcare services (Topic-26), health expenditure (Topic-06), consumer documentation (Topic-57), management of organizational processes (Topic-40), Healthcare quality indicators (Topic-50) inappropriate audit attendance and audit category (Topic-60 &Topic-59). Regarding education and technology, CTM discovered various topics like healthcare education (Topic-20), technology adoption and its acceptance (Topic-35), and Healthcare IoT and devices (Topic-51 & Topic-37). Beyond these topics, some deal deals with women like women's stigma (Topic-52), Brest cancer (Topic-15), and maternal inequalities (Topic-18).

For a profound understanding of the discovered topics, we also elaborated on the top-cited articles of the corpus in section (2.1) that can cross-validate the themes of the mostly contextual topics. Topic 42, Topic 51, Topic 37, Topic 54, and Topic 35, for instance, can be supported in light of "Paragraph 2." (par.2). Both sections disused the applications of technology for healthcare. Another example would be the topics (Topic-24, Topic-04, Topic-38, Topic-29, Topic-27, Topic-31, Topic-03, Topic-9) are positioned with (par.3). These units mainly focused on cost and other relevant subjects like surgery and infections. Similarly, we can also cross-validate the rest of the topics. Another perspective of this study is that since we already encoded the context of topics in the modeling, the discovered topics are more coherent and self-explanatory (See Table.2).

4.3 Classification of topic trends

Based on the p-value, different groups are created for the trends of the contextualized topics. These classes were described in terms of arrow symbols. For example, the arrow symbol ↑ (↓) showed an increasing (decreasing) trend, but not significant (if $p > 0.05$), and other arrow symbols like ↑↑ (↓↓), ↑↑↑ (↓↓↓), ↑↑↑↑ (↓↓↓↓) showed a significantly increasing (decreasing) trend (if $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively as shown in (See Table.1). The annual distributions of these topics are depicted from the graphs (See Figure 3 and 4). The fifteen topics showed significantly increasing trends, twelve topics showed significantly decreasing trends thirty-three topics do not show significantly increasing or decreasing trends.

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

4.4 Inter-topic distances topic-clusters and correlations

We analyzed the inter-topic distance map of the most fitted contextualized topic model with (K=60). Figure 2 gives us three important clusters of the topics. Cluster-1 shows thirteen topics, and cluster-2 is dense and consisted of twenty-three topics. Cluster-3 consists of ten topics. We analyzed the different topics residing in these clusters which give important patterns, and correlations among the topics and various inter-topic research directions as follows:

The most significant topics covered under cluster-1 mainly focused on statistical analysis (95-CI) for different aspects of medical care, risk factors, associated cost, and re-hospitalization in severe medical conditions. Cluster-2 is dense and is further composed of various sub-groups. In the first group generalized topics (23, 32, 36, 46, 50) narrate the reviews on different areas of healthcare like job satisfaction, drug management, the need for cancer patients, and healthcare quality indicators.

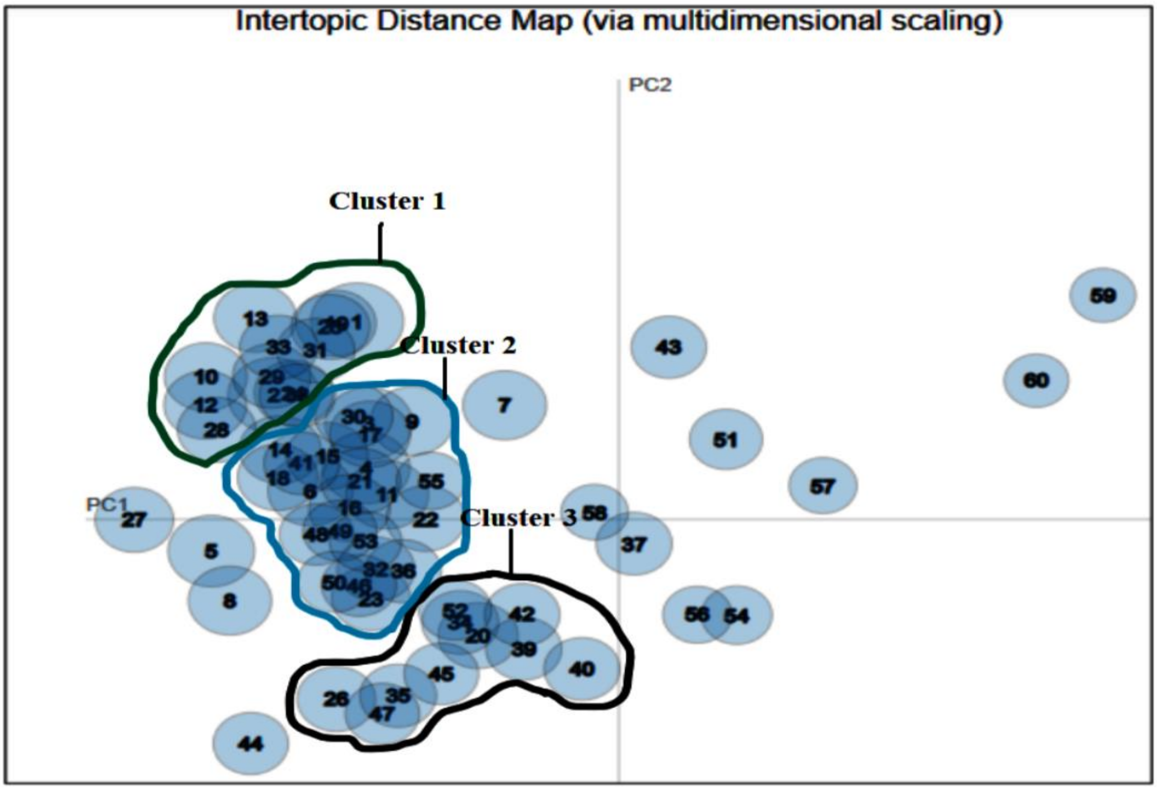

Figure: 2. Inter-topic distance and topic clusters

The topics in the second group (48,49,53) correlate due to discrimination factors in healthcare generally and sexual substance discrimination in particular. The group-3 topics (4,11,16,21,22,55) point out the pandemic (covid-19), its effects, symptoms, antimicrobial compliance, the importance of hand hygiene, estimation of the cost of vaccination, and the immunity of the people against various infections. The group-4 (topics 6,14,15,18,41) is covering health expenditure and home interventions for many health impairments like breast cancer, migraine, and maternity in different economic situations. It also clarifies the side effects of overwork, especially in women. Group 5 (topics 3,9,17,30) focuses on cost estimation and

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Figure 3. Graphical representations of topics (1-30) trends from 2000 to 2022

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

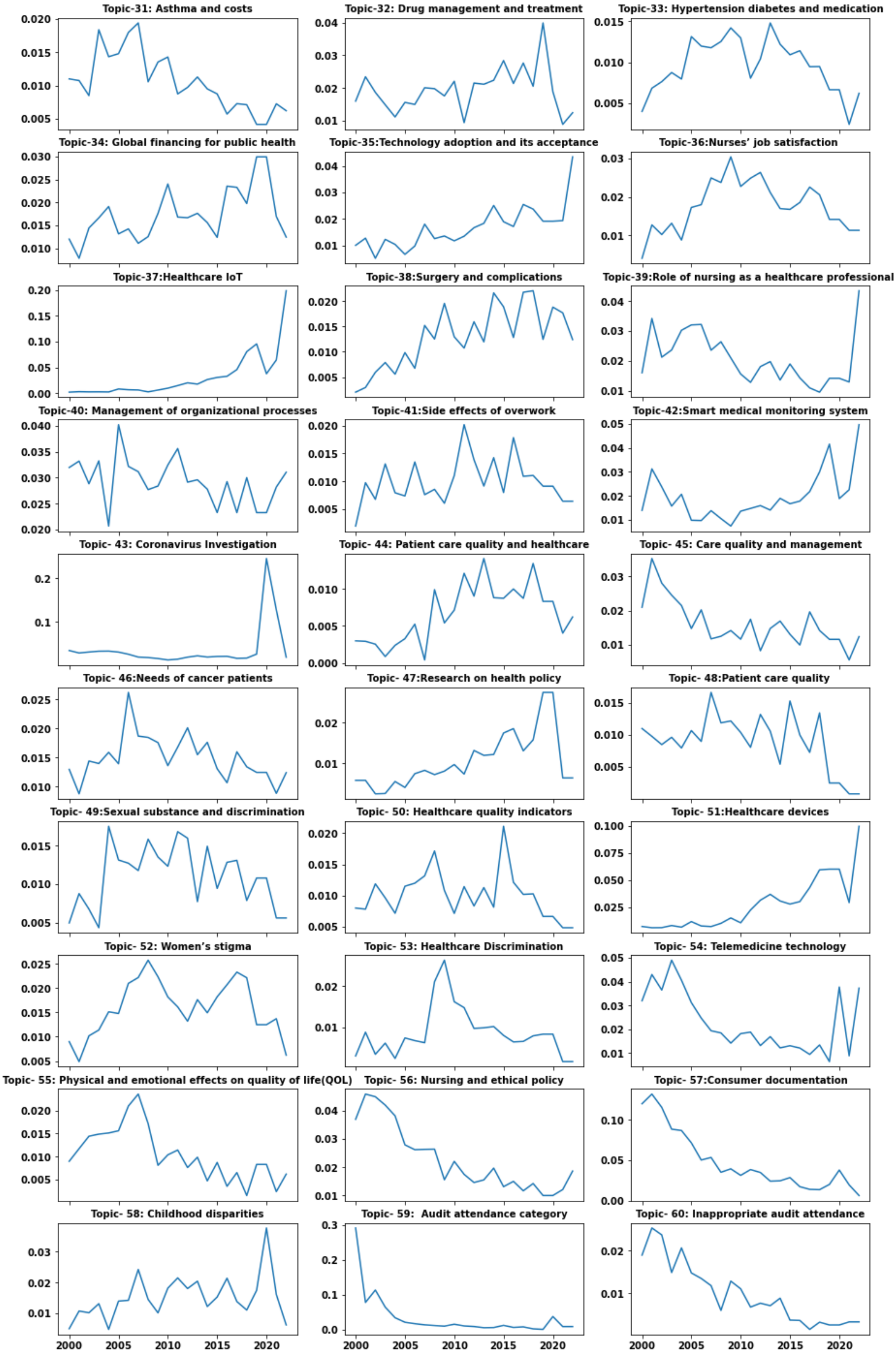

Figure 4. Graphical representations of topics (31-60) trends from 2000 to 2022

https://doi.org/10.1162/dint_a_00217

patient treatment in different diseases, especially in glucose control, drugs, and insulin. Additionally, it highlights the effects of inappropriate audit attendance cost estimation.

Cluster-3 (topics 20,26,34,35,39,40,42,45,47,52) elaborates on the significance of healthcare education, nursing, geographic organizational management of finance in the healthcare sector, uses, and application of smart medical appliances in healthcare units. It points out the discomforts of women's stigma ethically. This cluster also shows its findings on research in healthcare policies, care, quality, and clinical management, provision of primary healthcare services, and technology adoption in the treatment process.

5 Conclusion

In this study, we developed a context-based topic modeling approach that uses a transformer-based deep learning model to enhance the semantic understanding of the topics. We developed a novel Healthcare Bert to provide word embeddings as context vectors. These embeddings are combined with a bag of words to generate contextual topics of the corpus in context-based topic modeling. The newly generated word embeddings also improve the topic's coherence score. In order to categorize the themes into hot and cold categories, we further explored the document term matrix (DTM), which CTM generates using statical analysis techniques. This study also sheds light on the correlation between the topics by plotting them on a two-dimensional plane with a visualization tool. In this way, various interesting topic patterns and inter-topic research directions are pointed out. By generating rich sentence embedding vectors of the corpus under study using transformers-based models, corpus tuning, mean pooling, and the hugging face tool, this research broadens the existing topic modeling approaches which do not include contextual information in the topic extraction process. Moreover, it improves the previous topic modeling methodologies that analyze the healthcare literature from a specific disciplinary viewpoint. This process has several restrictions and things to think about. We can look at online databases like Web of Science and PubMed when choosing a data source, however, the current study solely takes into account papers that are indexed in Scopus. The current study adds context information to the topics and further gives clustering and topic correlations analysis, in future studies we may add hierarchical semantic modeling and temporal perspective in this direction.

Declarations

Availability of data and material.

The data underlying this article is available in [Google Drive], at
https://drive.google.com/drive/folders/1TOmO7VQYnhq-__CIU6Dqk5weFqkYAbH2?usp=sharing

References

Ali, I., Kannan, D. (2022). Mapping research on healthcare operations and supply chain management: a topic modeling-based literature review. Ann Oper Res.
Alemdar, H., & Ersoy, C. (2010). Wireless sensor networks for healthcare: A survey. *Computer networks*, *54*(15), 2688-2710.
Angus, D. C., Barnato, A. E., Linde-Zwirble, W. T., Weissfeld, L. A., Watson, R. S., Rickert, T., ... & Robert Wood Johnson Foundation ICU End-Of-Life Peer Group. (2004). Use of intensive care at the end of life in the United States: an epidemiologic study. *Critical care medicine*, *32*(3), 638-643.
Abalovich, M., Amino, N., Barbour, L. A., Cobin, R. H., De Groot, L. J., Glinoer, D., Mandel, S. J., Stagnaro-

Arora, S., Thornton, K., Murata, G., Deming, P., Kalishman, S., Dion, D., ... & Qualls, C. (2011). Outcomes of treatment for hepatitis C virus infection by primary care providers. New England Journal of Medicine, 364(23), 2199-2207.

Basmatkar, P., & Maurya, M. (2022). An Overview of Contextual Topic Modeling Using Bidirectional Encoder Representations from Transformers. In *Proceedings of Third International Conference on Communication, Computing and Electronics Systems* (pp. 489-504). Springer, Singapore.

Bianchi, F., Terragni, S., & Hovy, D. (2021). Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 759–766.

Barlow, S. E., & Expert Committee. (2007). Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics*, *120*, 164-192.

Boyce, J. M., & Pittet, D. (2002). Guideline for hand hygiene in health-care settings: recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Infection Control & Hospital Epidemiology, 23(S12), S3-S40.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., ... & Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Annals of internal medicine, 138(1), W1-12.

Cassini, A., Högberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., ... & Hopkins, S. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. The Lancet infectious diseases, 19(1), 56-66.

Chinn, R. Y., & Sehulster, L. (2003). Guidelines for environmental infection control in health-care facilities: recommendations of CDC and Healthcare Infection Control Practices Advisory Committee (HICPAC).

Chau, P. Y., & Hu, P. J. H. (2002). Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Information & management, 39(4), 297-311.

Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., & Tarricone, L. (2015). An IoT-aware architecture for smart healthcare systems. IEEE internet of things journal, 2(6), 515-526.

Cho, H. W. (2019). Topic modeling. Osong public health and research perspectives, 10(3), 115.

Dantu, R., Dissanayake, I., & Nerur, S. (2021). Exploratory Analysis of Internet of Things (IoT) in Healthcare: A Topic Modelling & Co-citation Approaches. *Information Systems Management*, *38*(1), 62–78.

Devaraj, S., & Kohli, R. (2003). Performance impacts of information technology: Is actual usage the missing link?. Management science, 49(3), 273-289.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Erasmus, V., Daha, T. J., Brug, H., Richardus, J. H., Behrendt, M. D., Vos, M. C., & van Beeck, E. F. (2010). A systematic review of studies on compliance with hand hygiene guidelines in hospital care. Infection Control & Hospital Epidemiology, 31(3), 283-294.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., ... & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic resonance imaging, 30(9), 1323-1341.

Guy Jr, G. P., Machlin, S. R., Ekwueme, D. U., & Yabroff, K. R. (2015). Prevalence and costs of skin cancer treatment in the US, 2002− 2006 and 2007− 2011. American journal of preventive medicine, 48(2), 183-187.

Gilson, L. (2003). Trust and the development of health care as a social institution. Social Science & Medicine, 56, pp.1453–68

Gurcan, F., Cagiltay, N. E., & Cagiltay, K. (2020). Mapping Human-Computer Interaction Research Themes and Trends from Its Existence to Today: A Topic Modeling-Based Review of past 60 Years. International Journal of Human-Computer Interaction, 00(00), pp.1–14, 10.1080/10447318.2020.1819668

Heidenreich, P. A., Albert, N. M., Allen, L. A., Bluemke, D. A., Butler, J., Fonarow, G. C., ... & Trogdon, J. G. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation: Heart Failure, 6(3), 606-619.

Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A., & Fridkin, S. K. (2008). Antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infection Control & Hospital Epidemiology, 29(11), 996-1011.

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Harris, R. P., Helfand, M., Woolf, S. H., Lohr, K. N., Mulrow, C. D., Teutsch, S. M., ... & Force, S. T. (2001). Current methods of the US Preventive Services Task Force: a review of the process. American journal of preventive medicine, 20(3), 21-35.

Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. American journal of public health, 103(5), 813-821.

Ho, P. M., Bryson, C. L., & Rumsfeld, J. S. (2009). Medication adherence: its importance in cardiovascular outcomes. Circulation, 119(23), 3028-3035.

Iddianozie, C., &Palmes, P. (2020). Towards smart sustainable cities: addressing semantic heterogeneity in Building Management Systems using discriminative models, Sustain. Cities Soc. 62.102367.

Principe, V. A., de Souza Vale, R. G., de Castro, J. B. P., Carvano, L. M., Henriques, R. A. P., de Almeida e Sousa Lobo, V., & de Alkmim Moreira Nunes, R. (2021). A computational literature review of football performance analysis through probabilistic topic modeling. *Artificial Intelligence Review*, 1-21.

Pare, G., &Kitsiou, S. (2017). Methods for literature reviews, in: F.K.C. Lau (Ed.), Handbook of EHealth Evaluation: An Evidence-Based Approach.

Pasquali, A. R. (2017). Automatic coherence evaluation applied to topic models (Doctoral dissertation, Universidade do Porto (Portugal)).

Jalali, M.S., Razak, S., Gordon, W., Perakslis, E., Madnick, S. Health Care, and Cybersecurity: Bibliometric Analysis of the Literature. J Med Internet Res 2019;21(2).

Jack, B. W., Chetty, V. K., Anthony, D., Greenwald, J. L., Sanchez, G. M., Johnson, A. E., ... & Culpepper, L. (2009). A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Annals of internal medicine, 150(3), 178-187.

Jensen, P. A., Lambert, L. A., Iademarco, M. F., & Ridzon, R. (2005). Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005.

Joosten, E. A., DeFuentes-Merillas, L., De Weert, G. H., Sensky, T., Van Der Staak, C. P. F., & de Jong, C. A. (2008). Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. Psychotherapy and psychosomatics, 77(4), 219-226.

Kim, H., Choi. H., Kang, H., An, J, Yeom, S, &Hong, T. (2021). A systematic review of the smart energy conservation system: from smart homes to sustainable smart cities, Renewable Sustainable Energy Rev. (2021),

Koroteev, M. V. (2021). BERT: A review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*.

Klevens, R. M., Edwards, J. R., Richards Jr, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., & Cardo, D. M. (2007). Estimating health care-associated infections and deaths in US hospitals, 2002. Public health reports, 122(2), 160-166.

Kakkar, A. K., Brenner, B., Dahl, O. E., Eriksson, B. I., Mouret, P., Muntz, J., ... & RECORD2 Investigators. (2008). Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomised controlled trial. The Lancet, 372(9632), 31-39.

Krishna, S., Boren, S. A., & Balas, E. A. (2009). Healthcare via cell phones: a systematic review. Telemedicine and e-Health, 15(3), 231-240.

Koivikko, A., & Løwenstein, H. (Eds.). (1985). Allergy: European Journal of Allergy and Clinical Immunology: Proceedings of a Postgraduate Course, Sources of Allergens, June 5, 1984, Turku, Finland: in Association with the XV Nordic Congress in Allergology.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, *62*(10), 1-34

Lessa, F. C., Mu, Y., Bamberg, W. M., Beldavs, Z. G., Dumyati, G. K., Dunn, J. R., ... & McDonald, L. C. (2015). The burden of Clostridium difficile infection in the United States. New England Journal of Medicine, 372(9), 825-834

Ludvigsson, J. F., Otterblad-Olausson, P., Pettersson, B. U., & Ekbom, A. (2009). The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. European journal of epidemiology, 24(11), 659-667

Martins, J. M., Pathmanathan, I., Tan, D. T., Lim, S. C., & Allotey, P. (Eds.). (2021). Systems Thinking Analyses for Health Policy and Systems Development. In Systems Thinking Analyses for Health Policy and Systems Development: A Malaysian Case Study (pp. i-ii), Cambridge: Cambridge University Press.

Mustakim, M., Wardoyo, R., Mustofa, K., G. R. Rahayu, G.R. and Rosyidah, I, "Latent Dirichlet Allocation for Medical Records Topic Modeling: Systematic Literature Review," *2021 Sixth International Conference on Informatics and Computing (ICIC)*, 2021, pp. 1-7, 10.1109/ICIC54025.2021.9632993.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... & Turner, M. B. (2015). Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*, *131*(4), 29-322.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., ... & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 263-269.

Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., ... & Fridkin, S. K. (2014). Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine*, *370*(13), 1198-1208.

McAlindon, T. E., Bannuru, R., Sullivan, M. C., Arden, N. K., Berenbaum, F., Bierma-Zeinstra, S. M., ... & Underwood, M. (2014). OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis and cartilage, 22(3), 363-388.

Marelli, A. J., Mackie, A. S., Ionescu-Ittu, R., Rahme, E., & Pilote, L. (2007). Congenital heart disease in the general population: changing prevalence and age distribution. Circulation, 115(2), 163-172.

Merikangas, K. R., He, J. P., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., ... & Olfson, M. (2011). Service utilization for lifetime mental disorders in US adolescents: results of the National Comorbidity Survey–Adolescent Supplement (NCS-A). Journal of the American Academy of Child & Adolescent Psychiatry, 50(1), 32-45.

Muraro, A., Werfel, T., Hoffmann-Sommergruber, K., Roberts, G., Beyer, K., Bindslev-Jensen, C., ... & EAACI Food Allergy and Anaphylaxis Guidelines Group. (2014). EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. Allergy, 69(8), 1008-1025.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics, 19(6), 1236-1246.

Muraro, A., Roberts, G., Worm, M., Bilò, M. B., Brockow, K., Fernández Rivas, M., ... & EAACI Food Allergy and Anaphylaxis Guidelines Group. (2014). Anaphylaxis: guidelines from the E uropean A cademy of A llergy and C linical I mmunology. Allergy, 69(8), 1026-1045.

Rejeb, A., Treiblmaier, H., Rejeb, K. *et al.* Blockchain research in healthcare: a bibliometric review and current research trends. *J. of Data, Inf. and Manag.* 3, 109–124 (2021). https://doi.org/10.1007/s42488-021-00046-2

Roger, V.L., Go., A.S., Lloyd-Jones, D.M., Benjamin, E.J., Berry, J.D., *et al.* (2012). Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation.125(1):2-220. https://: 10.1161/CIR.0b013e31823ac046.

Ramakrishna, S., Fujihara, K., Teo, W. E., Yong, T., Ma, Z., & Ramaseshan, R. (2006). Electrospun nanofibers: solving global issues. Materials today, 9(3), 40-50.

Reginster, J. Y., & Burlet, N. (2006). Osteoporosis: a still increasing prevalence. Bone, 38(2), 4-9.

Shi, L., Cheng, G., Xie, S. R., & Xie, G. (2019). A word embedding topic model for topic detection and summary in social networks. Measurement and Control (United Kingdom), 52(9–10), 1289–1298. https://doi.org/10.1177/0020294019865750

Sachs, J.D. (2001). Macroeconomics and Health: Investing in Health for Economic Development. Geneva: World Health Organization

Saheb, T., Dehghani, M., & Saheb, T. (2022). Artificial intelligence for sustainable energy: A contextual topic modeling and content analysis. *Sustainable Computing: Informatics and Systems*, *35*(February), 100699. https://doi.org/10.1016/j.suscom.2022.100699

Siegel, R.L., Miller, K.D., Jemal, A. (2019) Cancer statistics, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34. doi: 10.3322/caac.21551. Epub 2019 Jan 8. PMID: 30620402.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., ... & Henry, D. A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. bmj, 358

Sokol, M. C., McGuigan, K. A., Verbrugge, R. R., & Epstein, R. S. (2005). Impact of medication adherence on hospitalization risk and healthcare cost. Medical care, 521-530.

Stevens, J. A., Corso, P. S., Finkelstein, E. A., & Miller, T. R. (2006). The costs of fatal and non-fatal falls among older adults. Injury prevention, 12(5), 290-295.

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

Tablan, O. C., Anderson, L. J., Besser, R., Bridges, C., Hajjeh, R., CDC, & Healthcare Infection Control Practices Advisory Committee (2004). Guidelines for preventing health-care-associated pneumonia, 2003: recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports*, *53*(RR-3), 1–36.
Trung, T. Q., & Lee, N. E. (2016). Flexible and stretchable physical sensor integrated platforms for wearable human-activity monitoring and personal healthcare. Advanced materials, 28(22), 4338-4372.
Taylor, C. A., Bell, J. M., Breiding, M. J., & Xu, L. (2017). Traumatic brain injury–related emergency department visits, hospitalizations, and deaths—United States, 2007 and 2013. MMWR Surveillance Summaries, 66(9), 1.
Wang, X., Gu, Y., Xiong, Z., Cui, Z., & Zhang, T. (2014). Silk-molded flexible, ultrasensitive, and highly stable electronic skin for monitoring human physiological signals. Advanced materials, 26(9), 1336-1342.
Wang F, Shao W, Yu H, Kan G, He X, Zhang D, Ren M and Wang G (2020) Re-evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series. Front. Earth Sci. 8:14.
Zimlichman, E., Henderson, D., Tamir, O., Franz, C., Song, P., Yamin, C. K., ... & Bates, D. W. (2013). Health care–associated infections: a meta-analysis of costs and financial impact on the US health care system. JAMA internal medicine, 173(22), 2039-2046.
Zhao, S., Zhang, T., Hu, M. *et al.* (2022). AP-BERT: enhanced pre-trained model through average pooling. *Appl Intell* **52**, 15929–15937 (2022).

Appendix A

P-value	Topic	Top Terms
0.0065	01:95 ci for transplantation	transplantation, transplant,95, ci, crc, hr, hazard, blacks, stage, women, cancer, hispanics, survival, years, age, whites, hispanic, ses, among, disparities
0.0004	02:95 ci testing	hiv, aor, tb, ci, 95, testing, tuberculosis, risk, hcv, hcws, among, infected, posit ive, tested, infection, art, hepatitis, test, exposure, occupational
0.355	03: Patient treatment cost analysis	cost, costs, incremental, per, qaly, utility, placebo, treatment, patient, perspecti ve, compared, gained, effectiveness, model, total, analysis, patients, trial, year, life
0.0283	04: cost gained strategy	qaly, cost, gained, strategy, incremental, model, effective, effectiveness, per, s ensitivity, lifetime, perspective, life, years, screening, costs, year, treatment, sa vings, analysis
0.653	05: diabetes intervention	care, diabetes, primary, intervention, based, quality, screening, control, outco mes, practice, usual, management, self, intervention, controlled, trial, participa nts, group, effectiveness, practices
0.978	06: Health expenditure	households, health, china, expenditures, pocket, services, expenditure, househ old, payments, rural, utilization, inequality, urban, private, outpatient, children, insurance, areas, income, spending
0.0004	07: Resistance in mrsa isolation and transmission	isolates, outbreak, mrsa, aureus, resistant, isolation, ca, isolated, transmission, strains, resistance, hcw, environmental, hcws, infection, staphylococcus, antim icrobial, spread, tb, contact
0.0283	08: Healthcare services	hiv, services, health, facilities, healthcare, study, facility, rural, access, care, co mmunity, maternal, people, workers, district, women, service, delivery, tb, pri mary
0.0018	09: Costs estimation	billion, costs, million, indirect, disease, estimates, direct, burden, attributable, cost, fractures, per, estimated, incidence, year, estimate, total, annual, expectan cy, countries
0.0047	10:95 ci for patient disparities	95, ci, patients, hispanic, black, race, racial, white, whites, disparities, use, odd s, ed, ethnicity, likely, opioid, among, blacks, receive, associated
0.0033	11: antimicrobial compliance	antimicrobial, compliance, hygiene, catheter, icus, hand, antibiotic, infection, s urveillance, control, hospitals, units, bed, waste, rates, icu, feedback, infection s, intensive, resistant
0.0009	12:95 ci hospital risks	95, ci, risk, associated, hospital, rate, study, adverse, days, icu, events, rr, meta , mortality, studies, catheter, intervention, readmission, outcomes, ratio

0.0001	13: Patient treatment risks	patients, risk, hr, hazard,95, claims, treatment, ci, years, failure, matched, thera py, vs, disease, costs, date, diagnosis, higher, survival, compared
0.0001	14: home intervention for patient care	care, patients, home, intervention, group, usual, hf, clinic, primary, palliative, control, life, hospital, gp, randomized, months, program, groups, patient, telep hone
0.1611	15: Brest cancer	cancer, breast, prostate, stage, lung, colorectal, survivors, screening, survival, crc, women, cervical, treatment, diagnosis, chemotherapy, african, black, diag nosis, racial, disparities
0.0013	16: hygiene compliance	Hand, waste, hygiene, doctors, compliance, physicians, staff, errors, workers, s afety, reported, attitude, reporting, knowledge, nurses, training, pharmacists, h ealthcare, respondents, medical
0.0162	17: glucose control intervention	exercise, group, glucose, control, pressure, intervention, weeks, blood, training , week, participants, placebo, randomised, self, activity, controlled, randomize d, difference, trial, groups
0.7931	18: Maternal inequalities	maternal, inequalities, health, mortality, birth, neonatal, inequality, factors, out comes, countries, risks, socioeconomic, antenatal, income, deaths, low, pregna ncy, 95, ci, women
0.2904	19: children hospitalization rates	rates, children, hospitalizations, trends, age, rate, hispanic, injury, uninsured, n ationwide, hospitals, states, pediatric, blacks, per, medicaid
0.6918	20: Healthcare education	students, learning, training, practice, skills, education, implementation, eviden ce, knowledge, educational, change, teaching, development, course, clinical, in terventions, research, behavior, consensus, nursing
0.0504	21: hpv vaccination and immunization	hpv, vaccination, immunization, vaccine, influenza, cervical, uptake, vaccines, recommendation, screening, women, parents, pregnant, coverage, adolescent, knowledge, among, crc, cancer, years
0.0151	22: health pandemic and vaccination	pandemic, covid,19,2020, influenza, vaccination,2021, vaccine, vaccines, willi ngness, workers, sars, coronavirus, hcws, cov,2019, personal, stress, work, pub lic
0.0047	23: Reviews on health care research	Reviews, meta, articles, studies, systematic, search, trials, medline, searched, r eview, inclusion, outcomes, included, evidence, quality, research, intervention s, literature, bias, criteria
0.615	24: Infection-associated cost	catheter, cdi, per, infection, associated, days, infections, cost, rates, surveillanc e, incidence, urinary, nosocomial, icus, 1000, hospital, rate, costs,000, cases
0.9368	25: pneumonia patients	acquired, ca, pneumonia, cdi, onset, patients, aureus, sepsis, children, hospital, associated, icu, infection, mrsa, resistant, infections, community, admission, re spiratory, isolates
0.383	26: Primary healthcare services	care, palliative, health, services, primary, dementia, people, community, needs, mental, service, integrated, home, chronic, access, system, support, manageme nt, families, disease
0.4127	27: patient healthcare cost and services	care, costs, patients, healthcare, services, cost, medical, study, data, health, tot al, patient, use, service, claims, hospital, system, related, outpatient, year
0.0283	28: Asthma	ed, asthma, veterans, va, use, care, ptsd, pain, visits, mental, children, primary, reported, visit, emergency, medical, department, disorders, medication, disord er
0.328	29: Hospital charges and costs	Hospital, charges, costs, los, hospitalizations, stay, length, readmission, admiss ions, inpatient, readmissions, total, hospitalization, sepsis, cost, nationwide, as sociated, mortality, ed, admission
0.0015	30: Causes and effects of anxiety and stress	anxiety, stress, violence, depressive, depression, ptsd, sleep, burnout, psycholo gical, work, symptoms, workers, covid, mental, physical, ci, associated, factor s, risk, disorder
0.0001	31: asthma and costs	patients, vs, asthma, mean, copd, costs, group, symptoms, severe, months, day s, 12, disease, pulmonary, per, total, year, one, median, treatment

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

0.290	32: drug management and treatment	treatment, clinical, drug, management, guidelines, drugs, guideline, therapy, recommendations, therapies, disease, adherence, patients, treatments, medication, consensus, evidence, pain, may, therapeutic,
0.4756	33: hypertension diabetes and medication	hypertension, diabetes, cam, drugs, prevalence, prescription, drug, use, years, age, women, population, medication, medications, adherence, prescribed, prescriptions, glucose, 95, among
0.0162	34: Global financing for public health	financing, countries, health, public, sector, government, policy, private, policies, china, equity, global, africa, expenditure, spending, country, european, world, funding, universal
0.249	35: Technology adoption and its acceptance	data, information, adoption, technology, ehr, healthcare, systems, model, use, privacy, system, acceptance, electronic, sharing, decision, records, research, user, users, used
0.8325	36: Nurses job satisfaction	job, burnout, satisfaction, dimensions, reliability, items, validity, nurses, safety, instrument, patient, culture, teamwork, nurse, work, nursing, leadership, factor, item, quality
0.9195	37: Healthcare iot	smart, scheme, algorithm, network, iot, networks, proposed, internet, remote, devices, energy, cloud, security, efficient, accuracy, wireless, secure, prediction, real, propose
0.0004	38: Surgery Complications	volume, surgeons, surgery, postoperative, undergoing, mortality, surgical, hospitals, readmission, patients, hospital, complications, outcomes, procedures, trauma, factors, risk, los, covid, day
0.0039	39: role of nursing as healthcare professional	nurses, nursing, nurse, professional, students, caring, role, practice, roles, ethical, working, blackwell, skills, ethics, professionals, team, gps, managers, members, themes
0.0537	40: Management of organizational processes	process, processes, organisational, methodology, organisations, waste, paper, approach, systems, organizational, simulation, theory, value, technology, chain, supply, decision, managers, complexity, conceptual
0.5609	41: Side effect of overwork	productivity, costs, migraine, hrqol, indirect, copd, direct, asthma, disease, impairment, pain, work, burden, chronic, life, healthcare, impact, reported, total, related
0.0151	42: Smart medical monitoring system	internet, mobile, remote, medical, smart, information, monitoring, privacy, devices, security, system, telemedicine, secure, communication, iot, cloud, proposed, technologies, technology, user
0.2673	43: Coronavirus Investigation	ct, viral, ml, sars, respiratory, cell, ant, syndrome, dose, cov, skin, infectious, virus, coronavirus, rapid, epidemic, detected, samples, min, water
0.024	44: patient care quality and healthcare	patient, care, quality, study, healthcare, research, satisfaction, practice, primary, nurses, health, nursing, process, using, information, professionals, service, design, support, communication
0.0005	45: Care quality and management	care, quality, management, clinical, organizations, system, standards, evidence', guidelines, systems, recommendations, development, new, medicine, based, stroke, practice, improvement, programs, delivery
0.107	46: Needs of cancer patients	Patients, cancer, caregivers, needs, palliative, patient, family, information, preferences, life, professionals, care, making, end, communication, families, decision, unmet, support, diagnosis
0.0001	47: Research on health policy	research, health, social, policy, community, implementation, services, equity, communities, methods, framework, barriers, approach, cultural, qualitative, people, evidence, interventions, factors, determinants
0.0723	48: Patient care quality	Care, quality, patient, hospital, hospitals, performance, Accreditation, satisfaction, beds, critical, measures, emergency, safety, nursing, physicians, centered, improvement, medical, physician, nurse
0.634	49: sexual substance and discrimination	sexual, substance, cam, discrimination, mental, health, use, veterans, women, services, abuse, seeking, stigma, among, alcohol, barriers, unmet, reported, disorders, men

Data Intelligence Just Accepted MS.
https://doi.org/10.1162/dint_a_00217

0.1388	50: health care quality indicators	data, ehr, performance, quality, administrative, indicators, electronic, informati on, hospitals, records, clinical, used, systems, use, patient, using, measures, res earch, model, predictive
0.4145	51: Healthcare devices	skin, device, pressure, mechanical, detection, energy, monitoring, sensor, fall, flexible, algorithm, conventional, detect, power, technique, ms, imaging, accur acy, wearable, human
0.4435	52: Women stigma	stigma, women, hiv, experiences, sexual, social, partners, pregnant, themes, su pport, qualitative, pregnancy, participants, seeking, aids, young, men, depth, b eliefs, interviews
0.123	53: Health care discrimination	care, health, discrimination, disparities, ethnic, access, racial, minority, dental, americans, preventive, cultural, quality, survey, children, insurance, provider, satisfaction, whites, providers
0.0001	54: telemedicine technology	telemedicine, technologies, digital, market, technology, internet, adoption, sec urity, requirements, industry, consumers privacy, infrastructure, business, sect or, integration, users, solutions, networks, consumer
0.0011	55: Physical and emotional effects on quality of life (QOL)	qol, pain, functioning, back, exercise, self, physical, life, cognitive, hrqol, sym ptom, parents, depression, symptoms, functional, intensity, scale, emotional, a ctivity, distress
0.793	56: Nursing and ethical policy	nursing, article, ethical, policy, competence, practice, ethics, workforce, cultur al, education, leadership, professional, concept, leaders, nurse, political, ways, must, changing, programs
0.559	57: Consumer documentation	consumer, legal, documentation, pharmaceutical, consumers, act, food, advanc e, youth, crisis, inappropriate, assistance, accreditation, continuity, market, suc cess, travel, organ, encounters, reimbursement
0.0813	58: Childhood disparities	americans, childhood, disparities, minority, oral, ethnic, cardiovascular, africa n, prevention, morbidity, disease, racial, kidney, minorities, american, disorder s, populations, include, heart, obesity
0.1039	59: audit attendance category	audit, attendance, category, cessation, episode, continuity, encounters, consult ations, employment, netherlands, completion net, class, ms, falls, inappropriate , intensity, nutrition, knee, preventable,
0.218	60: Inappropriate audit attendance	audit, attendance, inappropriate, cessation, continuity, list, category, waiting, p reventable, urgent, class, employment, completion, consultations, nutrition, ms , accreditation, spent, malaria, encounters