Predictors of Health Information Seeking Behavior: A Systematic Review and Network Analysis

Ardalan Mirzaei, Parisa Aslani, Edward Joseph Luca, Carl Richard Schneider

Submitted to: Journal of Medical Internet Research
on: June 21, 2020

Table of Contents

Predictors of Health Information Seeking Behavior: A Systematic Review and Network Analysis

Ardalan Mirzaei[1*] BPharm, MPhil, GCertEdStud (Higher Ed); Parisa Aslani[1*] BPharm (Hons), MSc, PhD, GCertEdStud (Higher Ed); Edward Joseph Luca[2] BA, MBA; Carl Richard Schneider[1*] BN, BPharm (Hons), PhD, PGCert (Higher Ed)

[1]The University of Sydney School of Pharmacy The University of Sydney AU
[2]The University of Sydney Library The University of Sydney AU
[*]these authors contributed equally

Corresponding Author:
Ardalan Mirzaei BPharm, MPhil, GCertEdStud (Higher Ed)
The University of Sydney School of Pharmacy
Pharmacy and Bank Building A15
The University of Sydney
AU

Abstract

Background: People engage in health information seeking behavior (HISB) to support health outcomes. Being able to predict a person's behavior can inform the development of interventions to guide effective health information seeking. Obtaining a comprehensive list of the predictors of HISB through a systematic search of the literature and exploring the inter-relationship of these predictors are important first steps in this process.

Objective: This study aimed to; identify significant predictors of health information seeking behavior (HISB) in the primary literature; develop a common taxonomy for predictors of HISB; and identify the evolution of the HISB research field.

Methods: A systematic search of PsycINFO and Scopus was conducted for all years up to and including 10/12/2019. Quantitative studies identifying significant predictors of HISB were included. Information seeking was defined broadly and not restricted to any one source of health information. Data extraction of the significant predictors was performed by two authors. A network analysis was conducted to observe relationships between predictors over time.

Results: A total of 6,402 articles were retrieved, and after screening, 300 studies were retained for analysis. A total of 1,413 significant predictors were identified. These predictors were categorized into 67 predictor clusters. The most central predictors were age, education, gender, health condition and financial income. Over time, the inter-relationship of predictors in the network became denser, with the growth of new predictor grouping reaching saturation (1 new predictor identified) in the past 7 years, despite increasing publication rates.

Conclusions: A common taxonomy was developed, classifying 67 significant predictors of HISB. A novel temporal network was developed to track the evolution of research in HISB field, showing a maturation of new predictor terms and an increase in primary studies reporting multiple significant predictors of HISB. HISB research literature has experienced evolution with decreased characterization of novel predictors of HISB over time. A parallel increase in the complexity of predicting HISB has been identified with an increase in literature describing multiple significant predictors of HISB.

(JMIR Preprints 21/06/2020:21680)
DOI: https://doi.org/10.2196/preprints.21680

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Predictors of Health Information Seeking Behavior: A Systematic Literature Review and Network Analysis

Ardalan Mirzaei [1*]
Parisa Aslani [1]
Edward J Luca [2]
Carl R Schneider [1]
[1] The University of Sydney School of Pharmacy, Faculty of Medicine and Health, The University of Sydney, NSW, Australia
[2] The University of Sydney Library, University Library, The University of Sydney, NSW, Australia

*Corresponding author:
Room N411, Pharmacy and Bank Building A15, The University of Sydney, Sydney 2006, NSW Australia. Tel.: +61 2 8627 5225; fax: +61 2 4455 5206. E-mail address: ardalan.mirzaei@sydney.edu.au (A Mirzaei)

Keywords: information seeking; network analysis; health; review; temporal analysis

Abstract

Background

People engage in health information seeking behavior (HISB) to support health outcomes. Being able to predict a person's behavior can inform the development of interventions to guide effective health information seeking. Obtaining a comprehensive list of the predictors of HISB through a systematic search of the literature and exploring the inter-relationship of these predictors are important first steps in this process.

Objective

This study aimed to; identify significant predictors of HISB in the primary literature; develop a common taxonomy for predictors of HISB; and identify the evolution of the HISB research field.

Methods

A systematic search of PsycINFO, Scopus and PubMed was conducted for all years up to and including 10/12/2019. Quantitative studies identifying significant predictors of HISB were included. Information seeking was defined broadly and not restricted to any one source of health information. Data extraction of the significant predictors was performed by two authors. A network analysis was conducted to observe relationships between predictors over time.

Results

A total of 9549 articles were retrieved, and after screening, 344 studies were retained for analysis. A total of 1595 significant predictors were identified. These predictors were categorized into 67 predictor categories. The most central predictors were age, education, gender, health condition and financial income. Over time, the inter-relationship of predictors in the network became denser, with the growth of new predictor grouping reaching saturation (1 new predictor identified) in the past 7 years, despite increasing publication rates.

Conclusion

A common taxonomy was developed, classifying 67 significant predictors of HISB. A time-aggregated network method was developed to track the evolution of research in HISB field, showing a maturation of new predictor terms and an increase in primary studies reporting multiple significant predictors of HISB. HISB research literature has experienced evolution with decreased characterization of novel predictors of HISB over time. A parallel increase in the complexity of predicting HISB has been identified with an increase in literature describing multiple significant predictors of HISB.

Introduction

Health information seeking has been defined as “the ways in which individuals obtain information, including information about their health, health promotion activities, risks to one’s health, and illness”[1]. A consumer’s health information seeking behavior has the potential to influence the process and outcomes related to coping or adjusting to an illness or condition[1].
The conceptualization of health information seeking behavior (HISB) has evolved since Lenz’s first definition in 1984, which identified two dimensions of HISB; extent and method[2]. Lambert and Loiselle’s comprehensive review[1] of the concept of HISB found definitions to have ranged, over time, from actions or behavior used to obtain knowledge[3], clarify or confirm knowledge[4], satisfy a query[5], identify information sources[6,7], or demonstrate a coping strategy[8,9]. Other known models from the information science perspective include the Comprehensive Model of Information Seeking (CMIS) which looks at the antecedents, information carrier characteristics, and information seeking actions[10]. As well as the book by Case 2002 on the research on information seeking, needs and behaviors[11]. A recent paper by Zimmerman and Shaw describes HISB as an ‘umbrella term’ for many forms of information seeking, such as ‘direct seeking, information monitoring or browsing, and the passive receipt of information’[12]. Alternatively, the passive receipt of health information by people has been defined separately as information scanning[13]. These people may not be active in their search, but are still receptive enough to receive information. Thus, while the concept of HISB has existed for well over 30 years, there remains a lack of consensus on the definition of HISB and the theories that model it.
Despite these inconsistencies, previous models and theories commonly describe HISB as involving the action of seeking out information, regardless of where it is from, how it is sought, or why it is sought[1,14]. Predictors of HISB can therefore be described as the variables affecting the action(s) of seeking out information. These predictors can be contextual, such as the environment of an individual or their social networks, as well as being personal, such as sociodemographic characteristics, health status, or internal beliefs. There may also be predictors of persistence in HISB, such as satisficing. Satisficing relates to decision-making: “through which an individual decides when an alternative approach or solution is sufficient to meet the individuals’ desired goals rather than pursue the perfect approach”[15]. In the context of information seeking, it is choosing whether it is worth the cost or effort of continuing to search, or whether already acquired information suffices.
Factors that influence why people engage in information seeking include; the content of the information, which information sources or channels are used frequently and their credibility, and the barriers they may have in seeking information[1,16]. A meta-analytic review by Chang and Huang, quantified 7 of the predictors of health information seeking behavior, however, not all of the predictors were included in their review, such as behaviors (adherence) and beliefs[17]. Predicting peoples’ behavior for health information seeking requires an understanding of the predictors, and their significance and magnitude in impacting information seeking behavior.
We define “significant predictors” as those that have been shown through empirical research to have a direct effect, rather than an association (correlation), to health information seeking. An example of a direct effect is age. Nelissen et al. showed that an increase in age led to increased cancer information seeking[18]. An example of an association is the relationship between patient-physician interaction and information seeking behavior. In this case, it is unclear whether information seeking leads to better patient-physician interaction (an outcome of HISB), or a better physician-patient interaction leads to increased HISB (a predictor of HISB). Though associations between predictors and health information seeking behavior may have statistical significance in some empirical studies, knowing the direction of the effect from predictor to variable is more informative. Furthermore, while qualitative research can provide a grounding for the identification of predictors of HISB, an

ability to quantify the effect size allows for comparison of the relative importance of individual predictors.
A comprehensive list of predictors of HISB provides researchers with a focus for identifying new significant predictors or examining the relationship and effect of new interventions. Predictors of HISB can support researchers, clinicians, private institutions, and public health initiatives to optimize information-related interactions between themselves and consumers, leading to a more positive healthcare management experience[1,16].
A consistent classification of terms is the first step in formulating a comprehensive list of predictors of HISB. For instance, there is concern that the term ‘race’ and ‘ethnicity’ have been used synonymously in health research despite being separate constructs[19]. Thus, how predictors of HISB are defined in one study may not necessarily be consistent with another.

Anker and colleagues compiled a comprehensive list of predictors for HISB a decade ago through a systematic search of the literature[16]. They extracted and reported on the methods and measures used in HISB research. However, there are two critical shortcomings to the review. First, Anker and colleagues restricted their definition of HISB to an active process, in accordance with Niederdeppe’s definition of HISB[20]. Consequently, predictors for non-active acquisition of health information were not identified[20]. Second, the search strategy was restricted to a single database (PsycINFO) with the justification that HISB is a social psychological construct as opposed to a medical construct. PsychINFO has focused subject areas, and it is possible that other HISB researchers may have published in journals not indexed in PsycINFO.

Importantly, there is a need for an updated review to account for evolution in information seeking as a result of the rapid emergence and dominance of mobile digital information technology. The use of the Internet has been rising over the preceding three decades[21]. Advances in technologies such as smartphones has led to further availability and access to the Internet. Since 2011, smartphone ownership by the American population has increased from 35% to 81% in 2019, with 96% of the population owning a cellular device[22]. Similarly, the use of smartphones has led to greater access to the Internet with the exclusive use of smartphones for Internet access doubling from a reported 17.5% in 2013 to 37% of the American population in 2019[21,23]. A smartphone user is estimated to spend a daily average of 2.6 hours on their device[25]. While the Internet has become a common source of health information,[26] how the influence of the Internet has modified on the predictors of HISB over time is yet to be well characterized. Primary studies have compared sources of information used by people as part of HISB, however, most studies have only compared the findings from the early 2000s with early and mid-2010’s[27-29]. Such comparison studies report that Internet use was not a predictor of information seeking behavior, yet Huerta et al. report an increase in the use of the Internet, especially with older age groups[27]. In contrast, Li and colleagues performed a hierarchical regression analysis comparing 2002 and 2012 cohorts from the Pew Database to examine changes in online HISB[29]. They identified Internet access as a predictor for health information to have declined over time, with the authors hypothesizing this could be partly due to the increase in misinformation and rise of smartphone use resulting increased accessibility of the Internet as a source of health information in the US. The extent to which these changes have impacted predictors of HISB as a whole and across information sources and settings has yet to be reviewed.
This review aimed to identify predictors of HISB as reported in the primary literature and to explore the relationships between predictors over time.
The specific objectives were to:

- identify the significant predictors of HISB in the primary literature;
- develop a common taxonomy for predictors of HISB;
- identify the evolution of the HISB research field over time using the quantitative studies.

Methods

Selection criteria

The following section outlines the inclusion and exclusion criteria for this review.

Types of participants

Papers were included where participants were defined as health consumers or caregivers. The intent of the search for information is important: a health consumer searches for information for their own self/treatment, as opposed to a health professional who may search for information to provide therapy. Caregivers were also included as they are in a non-therapeutic relationship with the health consumer.

Papers were excluded where the participants were health students (university or college), or simulation studies where the participants sought information prospectively in hypothetical scenarios. Students studying a health-related degree were removed if in their health-related disciplines they were searching for information for their future role as health professionals.

Types of studies

Quantitative studies were eligible for inclusion. Relevant study designs included experimental and epidemiological studies, including randomized controlled trials, non-randomized controlled trials, quasi- experimental, before and after studies, prospective and retrospective cohort studies, case control studies and analytical cross-sectional studies.

An article was included if it reported the level of significance for a predictor. That is, the study showed that a certain ‘predictor’ is significant in causing ‘information seeking’, rather than associations. Significance was determined by either p values that were less than 0.05, or in the case of odds ratio, if confidence interval did not cross 1. Articles were also included if there was further work demonstrating a causation or effect size, which was often shown through logistic or linear regression, confirmatory factor analysis, or structural equation modelling.

Studies were excluded if they did not have a quantitative focus. These included qualitative studies such as focus groups, semi-structured interviews, mixed method studies (with no quantitative component) and content analysis of websites or interviews.

Search Procedures

Search limits

Papers published in English up to 10/12/2019 since database inception, were considered for inclusion. No date range was applied. A participant’s information seeking was not restricted to any one source, and all sources (e.g., online, healthcare practitioner) were included.

Databases

The databases searched were PsycINFO, Scopus and PubMed. Scopus is considered the largest abstract and citation database of peer-reviewed literature and incorporates results from Embase and MEDLINE[30].

Search terms/phrases

The keywords used were:
“Health” OR “Drug” OR “Medicine”;
AND
“Information Seeking” OR “Information Behavior” OR “Information Search” or “Satisficing”

(See Table 1 and 2 in the Supplementary File for full syntax.)

Screening

Two authors (AM and EL) independently screened a random split of articles for inclusion, by title and abstract using the selection criteria. Pilot tests were conducted to calibrate the screening process before records were split.

Data Extraction

Each included paper was counted as a data source for the purpose of extraction. Two authors (AM and EL) independently extracted the following variables: year of publication, country of the study, participant recruitment, disease states, theories used and significant predictors. Significant predictors were those variables for which direct effects were reported (not correlation) on HISB and provided significance with either *P* values less than 0.05, or confidence intervals, in the case of odds ratio, that did not cross 1. Significant relative predictors of HISB from studies reporting on comparative HISB between two or more groups were also extracted. Any uncertainty with extraction was mediated by a third author (CS). An audit of 10% of screened articles and extracted variables was performed by CS.

Analysis

Content Analysis

The first author (AM) analyzed the variables and categorized significant predictors into emerging categories. Individual predictors as identified in the individual papers were extracted and an iterative process of clustering was undertaken by two authors (AM and CS). Consensus was reached amongst the two authors for terms and categories. The categories consisted of identifying similarity between predictor terms; where predictors were the same they were categorized together, where predictors were similar with a common definition they were categorized together, and finally predictors were categorized together where there was no common definition but were described similarly in text[31].

Predictor Frequency

As part of the content analysis, a ‘word frequency’ analysis was performed. In this case, the ‘words’ chosen were the identified predictor terms. Examining the predictor term frequency assists in analyzing the strength and importance of a predictor in regards to other terms[32-35]. Each predictor that was extracted into a category was counted as 1 for that article. Multiple predictors, if they categorized the same, were categorized as 1. For example, if the article reported “Age 25-30" and “Age 65-70" as being significant factors, then they would be categorized as “Age”, however they would contribute only 1 to the “Age” category for that article, as opposed to 2. The total predictors were then reported, and the predictor frequency was then used to develop the network structure for network analysis.

Network Analysis

Network analysis has been used in previous systematic reviews, for example, to identify relationships between authors of the included papers,[36] or in a health context for comparison of drug treatments[37]. Traditionally, quantitative data in a systematic review are pooled via meta-analysis. This requires homogenous data. A network analysis allows for the examination of relationships among heterogeneous entities[38]. A network analysis was conducted to observe relationships between predictor terms. This method can help to identify nodes (or predictor terms) that are connected to other nodes and show the relationships between terms in the literature[39,40].
The weights of the nodes were based on the frequency of the predictor, and the size based on the number of articles mentioning that predictor term. An analysis of changes over time was undertaken to compare networks of articles pre- and post-2008 (articles dated up to 31/12/2008). While Pew Research started reporting smartphone ownership from 2011,[22] the iPhone was the beginning of a

new era of phones[41]. The year 2008 was chosen to distinguish between the availability of smartphones following the introduction of the iPhone in 2007, allowing for the uptake of the device to have begun[42]. Accordingly, in 2008, global mobile broadband subscriptions overtook fixed broadband subscriptions[24]. Time based comparisons of temporal and atemporal network features was observed using time-varying networks. Such an approach has been used in ecology, transport, and social media[43-45].

Co-occurrence of individual predictors within an article was calculated according to the predictor frequency. Each individual predictor term was connected bi-directionally to another predictor from the same article. Each connection added a weight of 1 to the edge. Edges formed where a pair of predictor terms were mentioned together in one article. The visualization was created using the software R, with the code available on GitHub[46]. Co-occurrence of predictors and visualization of the network was created using the igraph package, a software package used for network visualizations between different objects on a network map[47].

To compare the different networks, the number of nodes, number of edges, and modularity were captured. Modularity measures the clustered communities of nodes, which is how the nodes cluster together forming a community group of nodes distant to another community group of nodes. The full setup and parameters are available on GitHub[46].

Results

Search Results

The literature search process is illustrated in Figure 1. From the two databases, a total of 9549 articles were retrieved, of which 2866 were duplicates. Following de-duplication, title and abstract screening was performed, followed by full-text screening. A total of 344 papers were included for synthesis in the final paper. The results of the categorized predictors are reported in Table 4 of the Supplementary file. The included articles contained papers published between 1993 and 2019, with a peak publication year in 2019 (Figure 2).

[INSERT FIGRUE 1]

Figure 1: Flowchart for article inclusion. Reasons for exclusion do not sum to excluded articles as some articles overlapped in reasons for exclusion.

Most of the research was conducted in the United States of America (n=202). Twenty-six articles reported studies from China, 12 from Australia and South Korea each, and 9 studies from the United Kingdom and Germany.

In 203 articles, the participants were recruited specifically for the study. However, another source of participants was to use existing databases of respondents such as the Health Information Trends Survey (HINTS) (n=65 papers), The Pew Research Center's Internet & American Life Project (n=9), and the Pennsylvania Cancer Registry (n=8).

Participants were predominately seeking information for chronic diseases, with cancer being the most studied condition (n=76).

Content Analysis

Fewer than half of the papers (n=167) were underpinned by a theory or model. Observing Figure 2, there had been an increase in the number of papers with a theoretical underpinning until 2014, at which point a plateau developed from 2015 to 2018. In 2019 there was the second peak of articles with a theoretical underpinning. However, still less than half of the papers published in 2019 were not supported by a theory or model (Figure 2).

A total of 1,595 non-unique significant predictors were identified. Table 4 in the Supplementary File lists the predictor categories. Predictors were classified into 68 categories (1 category was labelled as “unclassifiable/other”, unclassifiable predictors were not carried forward for network analysis.). The categories were further grouped into sociodemographic, health, information source, information content, and affective predictor groups. Sociodemographic variables of education (n=160), age (n=156), gender (n=120), were the most commonly reported significant predictor categories, followed by the health-related predictor categories of health condition (n=87). A noted increase in the number of predictors reported in the literature began in 2005 and peaked in 2019.

[INSERT FIGURE 2]

Figure 2: Total articles publications each year and the number of articles that used a theoretically grounded approach.

Network Analysis

[INSERT FIGURE 3]

Figure 3: Fruchterman-Reingold layout algorithm of the network analysis comparing the complete model with all years until 2019. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.

[INSERT FIGURE 4]

Figure 4: Fruchterman-Reingold layout algorithm of the network analysis comparing pre and including 2008 network model. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.

[INSERT FIGURE 5]

Figure 5: Fruchterman-Reingold layout algorithm of the network analysis comparing post-2008 network model. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.

The complete network with all terms and years combined resulted in 67 nodes (‘other’ node not included) and 4128 edge connections. The modularity of the groups revealed 3 clusters. The largest group of variables were predominately psychosocial predictors (n = 41). The second largest group were sociodemographic predictors (n = 22) followed by a third group that did not have any strong focus on any particular grouping (n = 4). However, most sociodemographic-related variables (age, education, gender) had the greatest eigenvector centrality before other variables (health condition and financial income), thus appearing in the center of the network (see Table 5 in the Supplementary File).

The network statistics reported in Table 5 of the Supplementary File show a difference of structure characteristics before and after 2008. Post-2008, only 15 extra nodes were identified, with no new nodes identified after 2014 (see Table 4 in Supplementary File). There was a 7.8 times greater number of edges in the post-2008 network than the pre-2008 network. The combination of increased number of edges and the limited increase in nodes, results in a more connected network post-2008, with the average number of adjacent edges to each node (mean degree of the nodes) increasing by 6.1 times compared to pre-2008. Age, Education, Gender and Health Condition are the nodes with the greatest degree of centrality indicating the greatest influence on adjacent nodes (Figure 3). Modularity was greatest before 2008 reducing to 2 in the post-2008, indicating tighter communities of nodes clustering together, with all years compared being 3. This tighter clustering is due to a greater co-occurrence of predictor terms being researched. That is, individual articles are reporting more significant predictors than previous articles. A sensitivity analysis to compare networks pre and post 2014 was conducted (most recent new node), which confirmed the dynamic shifts in network statistics post 2008 (see Table 5 in Supplementary File)

Discussion

This paper reports on a systematic search of the literature to identify and characterize predictors of HISB. The 321 included papers report 1,484 significant predictor terms of HISB that can be classified into 68 categories. A comprehensive list of HISB predictor terms was developed as a result. A novel temporal network analysis through the comparison of two sequential time-aggregated networks was conducted to characterize relationships between HISB predictors and identify changes over time. This approach has never been used previously in characterizing such relationships. Key findings were an increase in papers reporting on multiple significant predictors of HISB within a paper and a reduced rate in the identification of new predictors. The use of network analysis to map the relationships within a research field over time demonstrates the evolving nature of research, and provides insight into how the understanding of predictors of HISB has developed. A network approach was conducted by van de Wijngaert and colleagues to examine the current state in a research field pertaining to the adoption of eGovernment services[48]. However, they used structural equation modelling and a cross-sectional analysis, as opposed to our network comparison. The advantage of network comparison over time allows for characterization of the evolution of research fields. Future work would be to explore the use of temporal dynamics of networks, ideally through analysis of longitudinal datasets[43-45].

A recent meta-analysis conducted by Chang and Huang on the antecedents predicting health information seeking aggregated the antecedents into 7 categories[17]. Through their review they were able to quantify the effect sizes of the 7 categories, however, in their methodological design, not all the predictors of HISB were captured. While a valuable review, we note the difference of design and the papers that could be retrieved. Since Anker and colleagues' initial review in 2010, increased publication of papers exploring predictors of HISB have allowed for greater granularity in the identification of HISB predictors. Specifically, we have been able to develop the sociodemographic group of predictors from 5 predictors (age, gender, education, race, health literacy) to 16 predictors (caregiver, employment, household, language, sexual orientation, financial factors, societal engagement, and location of residence). A potential benefit of such granularity is the improved targeting for interventions to optimize HISB.

Anker et al. reported that medication adherence was an outcome of engaging in health information seeking. However, from our review, adherence was identified as a predictor. This suggests that HISB is affected by a feedback loop, where outcomes from HISB can be a predictor for further HISB. This relationship should be examined through longitudinal studies. Few longitudinal studies have examined HISB, but our findings suggest this could lead to unidentified HISB predictors. One longitudinal study described the reciprocal relationship between health anxiety and online HISB, and how people who are 'cyberchondriac' have health anxiety exacerbated[49]. Another study looked at clinician information engagement and information seeking[50] while others have shown that information needs and preferences change with time[51,52]. These initial findings demonstrate the utility of further longitudinal studies to measure additional predictors and outcomes of HISB.

The rate of article publication on HISB increased after 2005, with a doubling of articles published in the past 10 years compared to the 30 years prior. This finding mirrors an overall increase in the rate of academic publishing[53]. Another explanation might be the establishment of data gathering institutions like the Health Information National Trends Survey (HINTS) from the National Cancer Institute, which was established in 2003. Such cohort studies provide researchers with important opportunities to examine HISB across large nationally representative sampling frames. HINTS is the most used dataset across papers, and therefore the main source for identified HISB predictors. HINTS comprises of 12 cross-sectional surveys to date over the past 15 years [54-63]. The dataset has the advantage of being a representative sample frame of the United States.

The number of articles using theory to underpin their research has also increased over time. The use of theory has become a consistent theme for describing significant predictors. Interestingly, over the past seven years there has been a plateau in the frequency of publications reporting on HISB predictors. A possible reason for this is a maturation in the literature, with an apparent saturation of identified predictors. Li and colleagues also identified an increase in publications up until 2014[64]. Our findings have extended this trend to demonstrate a plateauing of publication rates since 2015.

Participants' interactions with social media which includes social networking sites and health blogs, which were categorized as 'Environment/Network/Internet', were identified as new predictor since 2008. Hamid and colleagues reviewed the role of social media in information seeking behavior among international students, highlighting that specific information needs were satisfied by using social media[65]. Though social media can be a medium for public health intervention,[66] it can also present a challenge as a source of misinformation[67,68]. Competing misinformation has implications for providers of information using online media to target their audience. Providers or creators of information could address the rise of misinformation by ensuring that content delivered through social media is verified for quality and that continued monitoring is implemented.

The terminology used to describe predictors varied significantly between papers, and, at times, lacked precision. Articles may have mentioned race as a predictor, but upon closer inspection of the survey items used for race, overlapped with ethnicity and culture. A potential contributor to the lack of clarity in terminology is the low number of papers identified that used a theoretically grounded approach. Ambiguous terminology poses a challenge when comparing findings between papers of HISB. In response to this issue, this review has developed a common classification structure of predictor terms. This structure has the potential to be developed into a future consensus taxonomy for predictors of HISB using domain ontologies[69].

Conducting a network analysis for predictors of HISB was a novel approach to analyzing the HISB research field. The overall network analysis shows the inter-relationship of the predictor variables, however the interaction between these variables in predicting HISB is still unclear. One concern is the issue of terms being correlated with each other, resulting in collinearity. Collinearity of predictor terms may affect how an individual predictor term affects HISB. The temporal network analysis finding of a 6.1 greater mean degree demonstrates the increase of publications that report on multiple significant predictors. Increased reporting was also supported by the increase in co-reporting, represented by reducing network modularity over time. The high centrality of Age, Education, Gender and Health Condition indicate that these are the most commonly reported predictors when multiple predictors were reported in the included studies. Such studies have a greater ability to identify collinearity of predictors. The network analysis approach allowed for examination of how our understanding of predictors of HISB and their inter-relationship changed over time. Such changes have occurred in the presence of a shift towards mobile technology becoming commonplace.

Strengths and limitations

This review has several strengths. A rigorous methodology was followed with multiple authors to extract data and reach agreements on definitions. However, due to the sheer number of articles returned in the initial database searches, there is the potential risk that articles meeting the inclusion criteria could have been potentially omitted. Yet, the likelihood of omitted articles affecting the findings of this review is low due to the number of included articles. The developed taxonomy of the predictors were directly informed by the included articles via a theoretically agnostic approach and consensus between two authors (AM and CS). The content validity of the developed taxonomy would benefit from validation via conceptual synthesis and a consensus approach from experts in the field of HISB. Reliability of data extraction could be considered a limitation of our review, which we tried to mitigate through a 10% audit by a third author (CS). Employing a quantitative measure of inter-coder reliability would have increased confidence in reliability.

Limitations to the search strategy are, first, the inclusion of only articles published in English. This is a potential issue in this field, as geographic and cultural differences were identified. The US was the most represented country. However, bibliometric analysis of Internet HISB literature has been previously performed by Li and colleagues[64]. The authors similarly identified a majority of articles being from the US. A skew towards a single country may introduce a geographic bias in the literature and subsequent identification of significant HISB predictors. There is evidence that context can directly influence the HISB of individuals, such as being in a resource-poor setting[70]. Therefore, it is important to be mindful of the number of studies from high-resource settings when considering the implications of our findings in low-resource settings. For instance, the presence of the predictor variable 'healthcare source accessibility' may be more pertinent for countries without universal healthcare coverage such as the US, where access to physicians is variable[71,72], [See Supplementary File]. A second limitation is the restriction of the definition of HISB as an active behavior. This limits the ability for the review findings to represent the predictors of passive HISB or 'scanning'. Third, the review findings do not represent predictors of HISB for university students due to the exclusion of this sub-population.

Finally, systematic reviews typically include an assessment of the risk of bias. The heterogeneity of the studies and the observational nature of most study designs meant that an assessment of bias was not appropriate. This led to this systematic review not adhering to the PRISMA guidelines and protocols to its entirety. However, the systematic approach adds to the strength of this paper.

Conclusions

A systematic search of the literature identified 344 papers reporting on predictors of HISB. A common taxonomy was developed to classifying predictors of HISB into 67 categories. Only 24% (n=16) of the predictor groupings emerged since the invention of smartphones. A novel network analysis allowed identification that the growth of new predictor groupings has approached saturation with only a single new predictor identified in the past seven years, despite increasing publication rates. Publication network analysis is a promising methodology for measuring trends across scientific fields.

Conflict of Interest

None

Multimedia Appendix 1

[Supplementary File containing search terms, predictors list and their definitions, network statistics, network analysis methods]

References

X

1. Lambert SD, Loiselle CG. Health information—seeking behavior. *Qual Health Res.* 2007;17(8):1006-1019.
2. Lenz ER. Information seeking: A component of client decisions and health behavior. *ANS Adv Nurs Sci.* 1984;6(3):59-72.
3. Barsevick AM, Johnson JE. Preference for information and involvement, information seeking and emotional responses of women undergoing colposcopy. *Res Nurs Health.* 1990;13(1):1-7.
4. Conley VM. Beyond Knowledge Deficit to a Proposal for Information-Seeking Behaviors. *International Journal of Nursing Terminologies and Classifications.* 1998;9:129-135.
5. Baker LM, Connor JJ. Physician-patient communication from the perspective of library and information science. *Bull Med Libr Assoc.* 1994;82(1):36-42.
6. Johnson JD. On contexts of information seeking. *Inf Process Manag.* 2003;39(5):735-760.
7. Czaja R, Manfredi C, Price J. The determinants and consequences of information seeking among cancer patients. *J Health Commun.* 2003;8(6):529-562.
8. Der Molem V. Relating information needs to the cancer experience: 1. Information as a key coping strategy. *Eur J Cancer Care (Engl).* 1999;8(4):238-244.
9. Rees CE, Bath PA. The psychometric properties of the Miller Behavioural Style Scale with adult daughters of women with early breast cancer: a literature review and empirical study. *J Adv Nurs.* 2000;32(2):366-374.
10. Johnson JD, Donohue WA, Atkin CK, Johnson S. A comprehensive model of information seeking: Tests focusing on a technical organization. *Science Communication.* 1995;16(3):274-303.
11. Case DO. *Looking for information : a survey of research on information seeking, needs, and behavior.* Fourth edition / by Donald O. Case, Lisa M. Given. ed. Bingley, UK: Emerald; 2016.
12. Zimmerman MS, Shaw Jr G. Health information seeking behaviour: a concept analysis. *Health Info Libr J.* 2020;n/a(n/a).
13. Kelly S, Eldredge SA, Dalton ED, Miller LE. Health-Information Behavior: An Initial Validity Portfolio for Active and Passive Measures. *Commun Res Rep.* 2014;31(2):171-182.
14. Lalazaryan A, Zare-Farashbandi F. A Review of models and theories of health information seeking behavior. *International Journal of Health System and Disaster Management.* 2014;2(4):193-203.
15. Simon HA. Designing organizations for an information-rich world. *International Library of Critical Writings in Economics.* 1996;70:187-202.
16. Anker AE, Reinhart AM, Feeley TH. Health information seeking: A review of measures and methods. *Patient Educ Couns.* 2011;82(3):346-354.
17. Chang C-C, Huang M-H. Antecedents predicting health information seeking: A systematic review and meta-analysis. *International Journal of Information Management.* 2020;54:102115.
18. Nelissen S, Beullens K, Lemal M, Van den Bulck J. Fear of cancer is associated with cancer information seeking, scanning and avoiding: A cross-sectional study among cancer diagnosed and non-diagnosed individuals. *Health Info Libr J.* 2015;32(2):107-119.
19. Sheldon TA, Parker H. Race and ethnicity in health research. *J Public Health (Oxf).* 1992;14(2):104-110.
20. Niederdeppe J, Hornik RC, Kelly BJ, et al. Examining the dimensions of cancer-related

information seeking and scanning behavior. *Health Commun.* 2007;22(2):153-167.
21. Pew Research Center. Demographics of Internet and home broadband usage in the United States. *Pew Research Center: Internet, Science & Tech* 2018; https://www.pewresearch.org/internet/fact-sheet/internet-broadband/, 2020.
22. Pew Research Center. Demographics of mobile device ownership and adoption in the United States. *Pew Research Center: Internet, Science & Tech* 2018; https://www.pewresearch.org/internet/fact-sheet/mobile/. Accessed February, 2020.
23. Anderson M. Mobile technology and home broadband 2019. *Pew Research Center.* 2019;2.
24. ITU-D. The World in 2009: ICT facts and figures. 2009. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2009.pdf.
25. Montag C, Błaszkiewicz K, Sariyska R, et al. Smartphone usage in the 21st century: who is active on WhatsApp? *BMC Res Notes.* 2015;8(1):331.
26. Percheski C, Hargittai E. Health information-seeking in the digital age. *J Am Coll Health.* 2011;59(5):379-386.
27. Huerta TR, Walker DM, Johnson T, Ford EW. A Time Series Analysis of Cancer-Related Information Seeking: Hints From the Health Information National Trends Survey (HINTS) 2003–2014. *J Health Commun.* 2016;21(9):1031-1038.
28. Finney Rutten LJ, Agunwamba AA, Wilson P, et al. Cancer-Related Information Seeking Among Cancer Survivors: Trends Over a Decade (2003–2013). *J Cancer Educ.* 2016;31(2):348-357.
29. Li J, Theng YL, Foo S. Predictors of online health information seeking behavior: Changes between 2002 and 2012. *Health Informatics J.* 2016;22(4):804-814.
30. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* 2008;22(2):338-342.
31. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277-1288.
32. Du H, Li N, Brown MA, Peng Y, Shuai Y. A bibliographic analysis of recent solar energy literatures: The expansion and evolution of a research field. *Renew Energy.* 2014;66:696-706.
33. Li J, Zhang Y, Wang X, Ho Y-S. Bibliometric analysis of atmospheric simulation trends in meteorology and atmospheric science journals. *Croatica Chemica Acta.* 2009;82(3):695-705.
34. Mao N, Wang M-H, Ho Y-S. A bibliometric study of the trend in articles related to risk assessment published in Science Citation Index. *Hum Ecol Risk Assess.* 2010;16(4):801-824.
35. Wang L, Zhao L, Mao G, Zuo J, Du H. Way to accomplish low carbon development transformation: A bibliometric analysis during 1995–2014. *Renewable and Sustainable Energy Reviews.* 2017;68:57-69.
36. Leppink J, Pérez-Fuster P. Social Networks as an Approach to Systematic review. *Health Professions Education* 2019; 218-224. Available at: http://www.sciencedirect.com/science/article/pii/S2452301118300695. Accessed 3, 5.
37. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. *Systematic reviews.* 2014;3:109-109.
38. Marsden PV. Network Centrality, Measures of. In: Wright JD, ed. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Oxford: Elsevier; 2015:532-539.
39. Cherven K. *Mastering Gephi network visualization.* Packt Publishing Ltd; 2015.

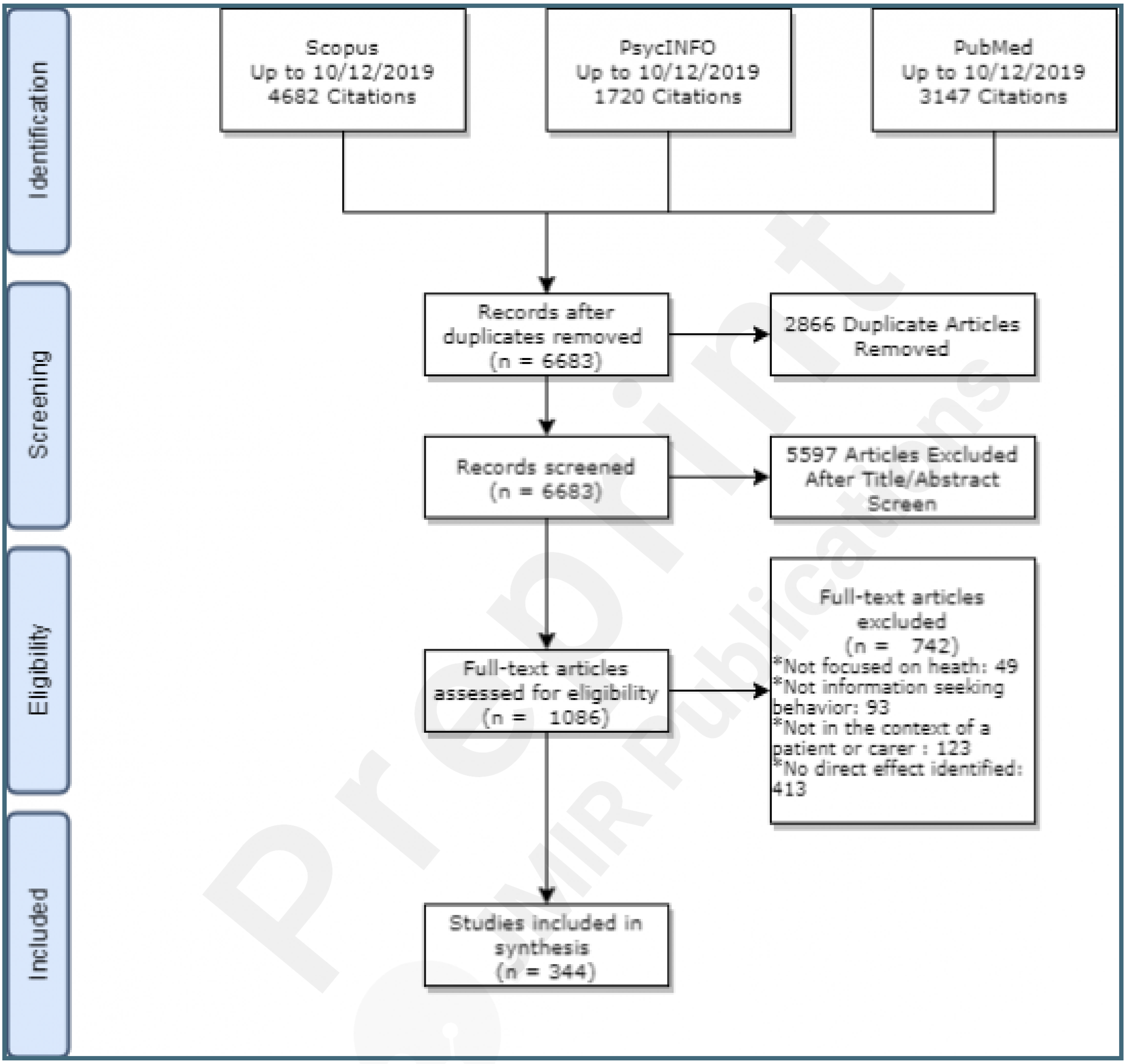
40. Khokhar D. *Gephi cookbook.* Packt Publishing Ltd; 2015.
41. Sarwar M, Soomro TR. Impact of smartphone's on society. *European journal of scientific research.* 2013;98(2):216-226.
42. Islam N, Want R. Smartphones: Past, Present, and Future. *IEEE Pervasive Comput.* 2014;13(4):89-92.
43. Blonder B, Wey TW, Dornhaus A, James R, Sih A. Temporal dynamics and network analysis. *Methods in Ecology and Evolution.* 2012;3(6):958-972.
44. Santoro N, Quattrociocchi W, Flocchini P, Casteigts A, Amblard F. Time-varying graphs and social network analysis: Temporal indicators and metrics. *arXiv preprint arXiv:11020629.* 2011.
45. Tang J, Musolesi M, Mascolo C, Latora V. Temporal distance metrics for social network analysis. Proceedings of the 2nd ACM workshop on Online social networks; 2009; Barcelona, Spain.
46. Mirzaei A. Algorithm and code used for calculating and displaying a network analysis for the predictors of health information seeking behaviour. [Code Repository]. 2020; https://github.com/ardimirzaei/HISB_Predictors_Network_Analysis.
47. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, complex systems.* 2006;1695(5):1-9.
48. van de Wijngaert L, Bouwman H, Contractor N. A network approach toward literature review. *Qual Quant.* 2014;48(2):623-643.
49. te Poel F, Baumgartner SE, Hartmann T, Tanis M. The curious case of cyberchondria: A longitudinal study on the reciprocal relationship between health anxiety and online health information seeking. *J Anxiety Disord.* 2016;43:32-40.
50. Moldovan-Johnson M, Tan AS, Hornik RC. Navigating the cancer information environment: The reciprocal relationship between patient-clinician information engagement and information seeking from nonmedical sources. *Health Commun.* 2014;29(10):974-983.
51. Harrison DE, Galloway S, Graydon JE, Palmer-Wickham S, Rich-van der Bij L. Information needs and preference for information of women with breast cancer over a first course of radiation therapy. *Patient Educ Couns.* 1999;38(3):217-225.
52. Matsuyama RK, Kuhn LA, Molisani A, Wilson-Genderson MC. Cancer patients' information needs the first nine months after diagnosis. *Patient Educ Couns.* 2013;90(1):96-102.
53. Bjork B-C, Roos A, Lauri M. Scientific journal publishing: yearly volume and open access availability. *Information Research: An International Electronic Journal.* 2009;14(1).
54. Health Information National Trends Survey 1 (HINTS 1): Methodology Report. National Cancer Institute Bethesda, MD; 2003.
55. Westat. Health Information National Trends Survey 2 (HINTS 2): Methodology Report. National Cancer Institute Bethesda, MD; 2005.
56. Westat. Health Information National Trends Survey 4 (HINTS 4): Cycle 1 Methodology Report. National Cancer Institute Bethesda, MD; 2012.
57. Westat. Health Information National Trends Survey 4 (HINTS 4): Cycle 2 Methodology Report. National Cancer Institute Bethesda, MD; 2013.
58. Westat. Health Information National Trends Survey 4 (HINTS 4): Cycle 3 Methodology Rreport. National Cancer Institute Bethesda, MD; 2014.
59. Westat. Health Information National Trends Survey FDA (HINTS-FDA): Methodology Report. National Cancer Institute Bethesda, MD; 2015.
60. Westat. Health Information National Trends Survey FDA2 (HINTS-FDA2): Methodology Report. National Cancer Institute Bethesda, MD; 2017.

61. Westat. Health Information National Trends Survey 5 (HINTS 5): Cycle 1 Methodology Report. National Cancer Institute Bethesda, MD; 2017.
62. Westat. Health Information National Trends Survey 5 (HINTS 5): Cycle 2 Methodology Report. National Cancer Institute Bethesda, MD; 2018.
63. Westat. Health Information National Trends Survey 4 (HINTS 4): Cycle 4 Methodology Report. National Cancer Institute Bethesda, MD; 2015.
64. Li F, Li M, Guan P, Ma S, Cui L. Mapping Publication Trends and Identifying Hot Spots of Research on Internet Health Information Seeking Behavior: A Quantitative and Co-Word Biclustering Analysis. *J Med Internet Res.* 2015;17(3):e81.
65. Hamid S, Bukhari S, Ravana Sri D, Norman Azah A, Ijab Mohamad T. Role of social media in information-seeking behaviour of international students: A systematic literature review. *Aslib Journal of Information Management.* 2016;68(5):643-666.
66. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. *NPJ Digit Med.* 2018;1:47.
67. Rolls K, Massey D. Social media is a source of health-related misinformation. *Evidence Based Nursing.* 2020:ebnurs-2019-103222.
68. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science & Medicine.* 2019;240:112552.
69. McDaniel M, Storey VC. Evaluating Domain Ontologies: Clarification, Classification, and Challenges. *ACM Comput Surv.* 2019;52(4):Article 70.
70. Dutta R. Information needs and information-seeking behavior in developing countries: A review of the research. *International Information & Library Review.* 2009;41(1):44-51.
71. Naylor KB, Tootoo J, Yakusheva O, Shipman SA, Bynum JPW, Davis MA. Geographic variation in spatial accessibility of U.S. healthcare providers. *PLOS ONE.* 2019;14(4):e0215016.
72. Aldridge ML, Daniels JL, Jukic AM. Mammograms and Healthcare Access Among US Hispanic and Non-Hispanic Women 40 Years and Older. *Fam Community Health.* 2006;29(2):80-88.

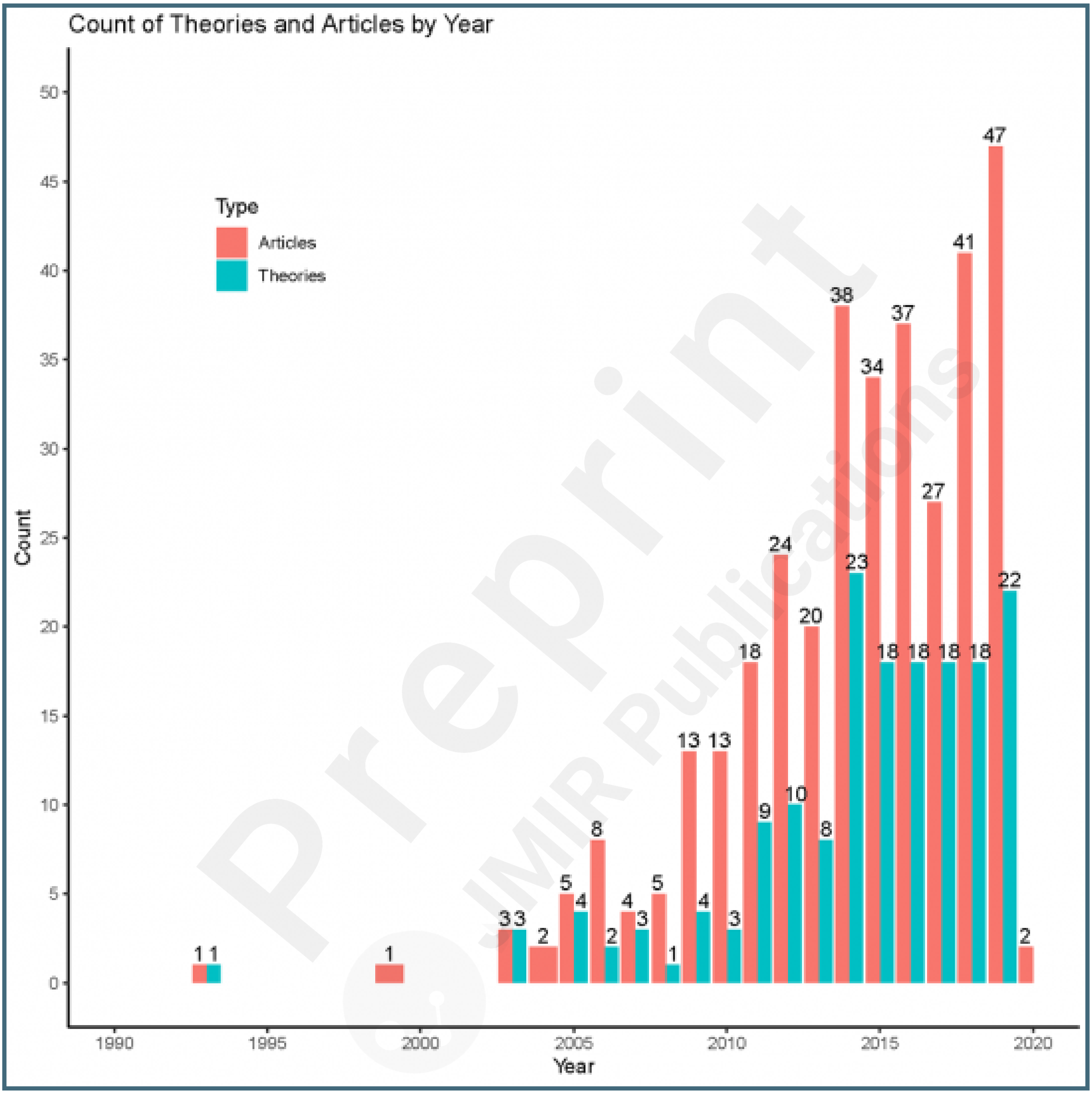
Supplementary Files

Figures

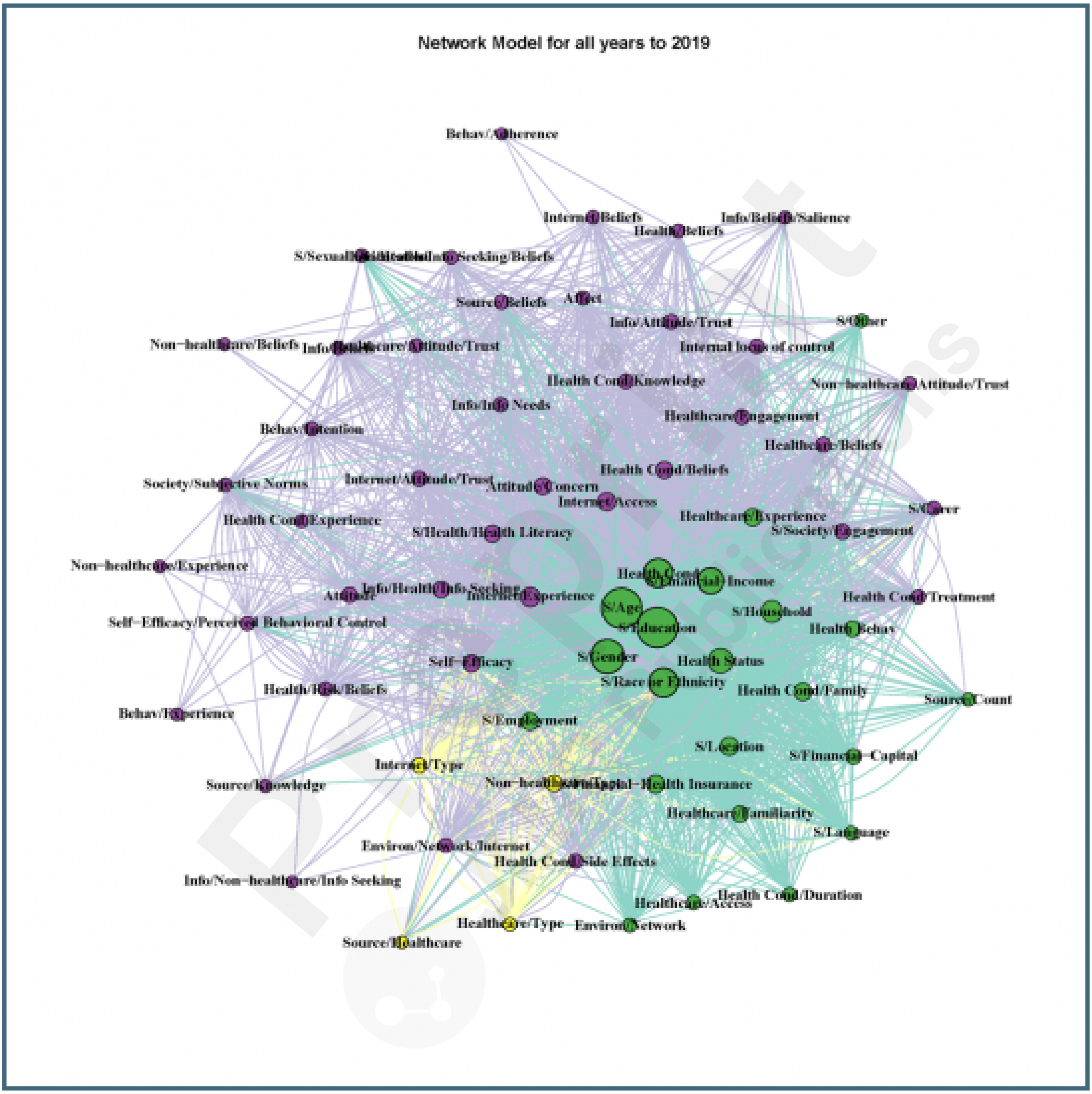
Flowchart for article inclusion. Reasons for exclusion do not sum to excluded articles as some articles overlapped in reasons for exclusion.



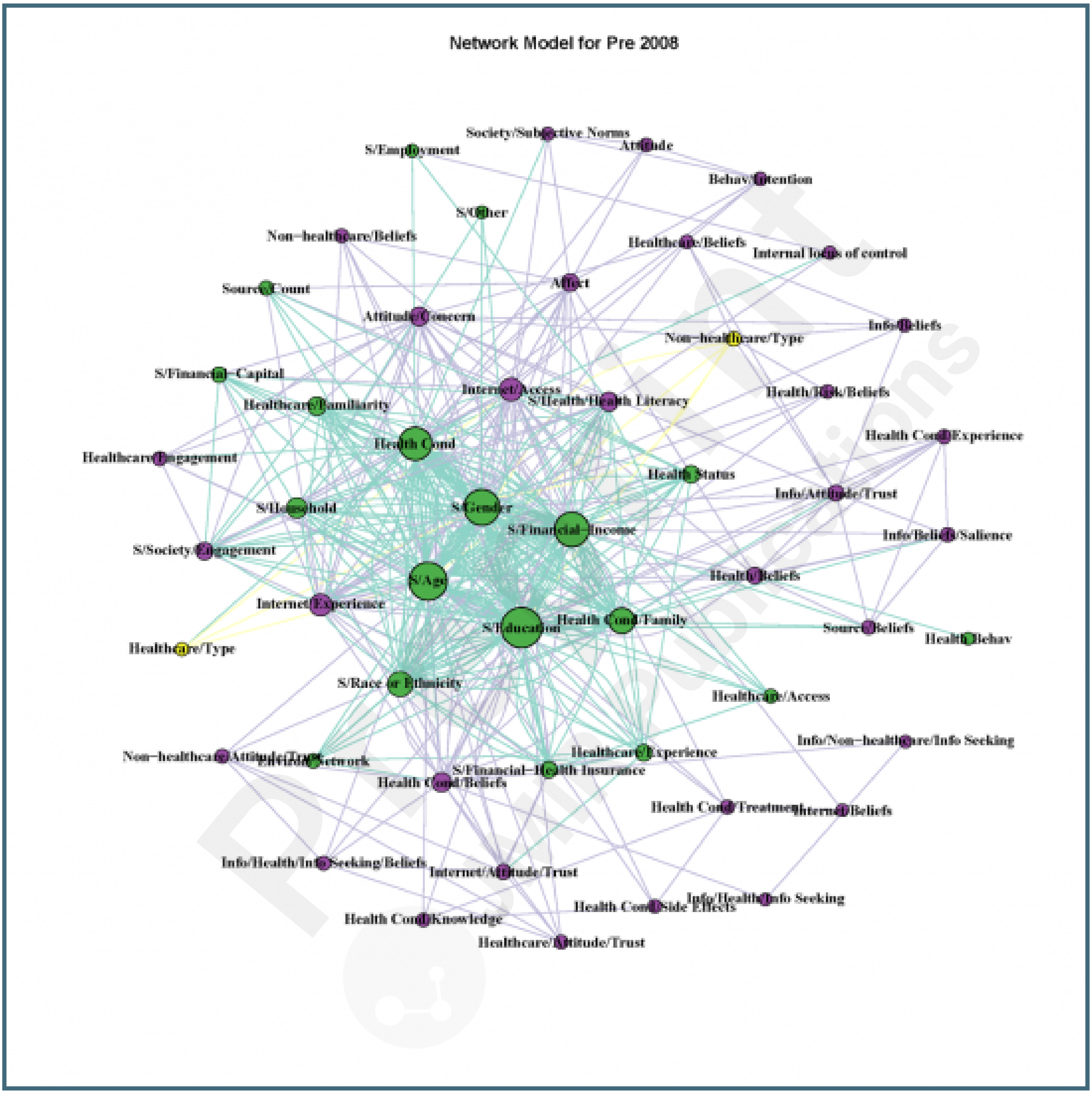
Total articles publications each year and the number of articles that used a theoretically grounded approach.



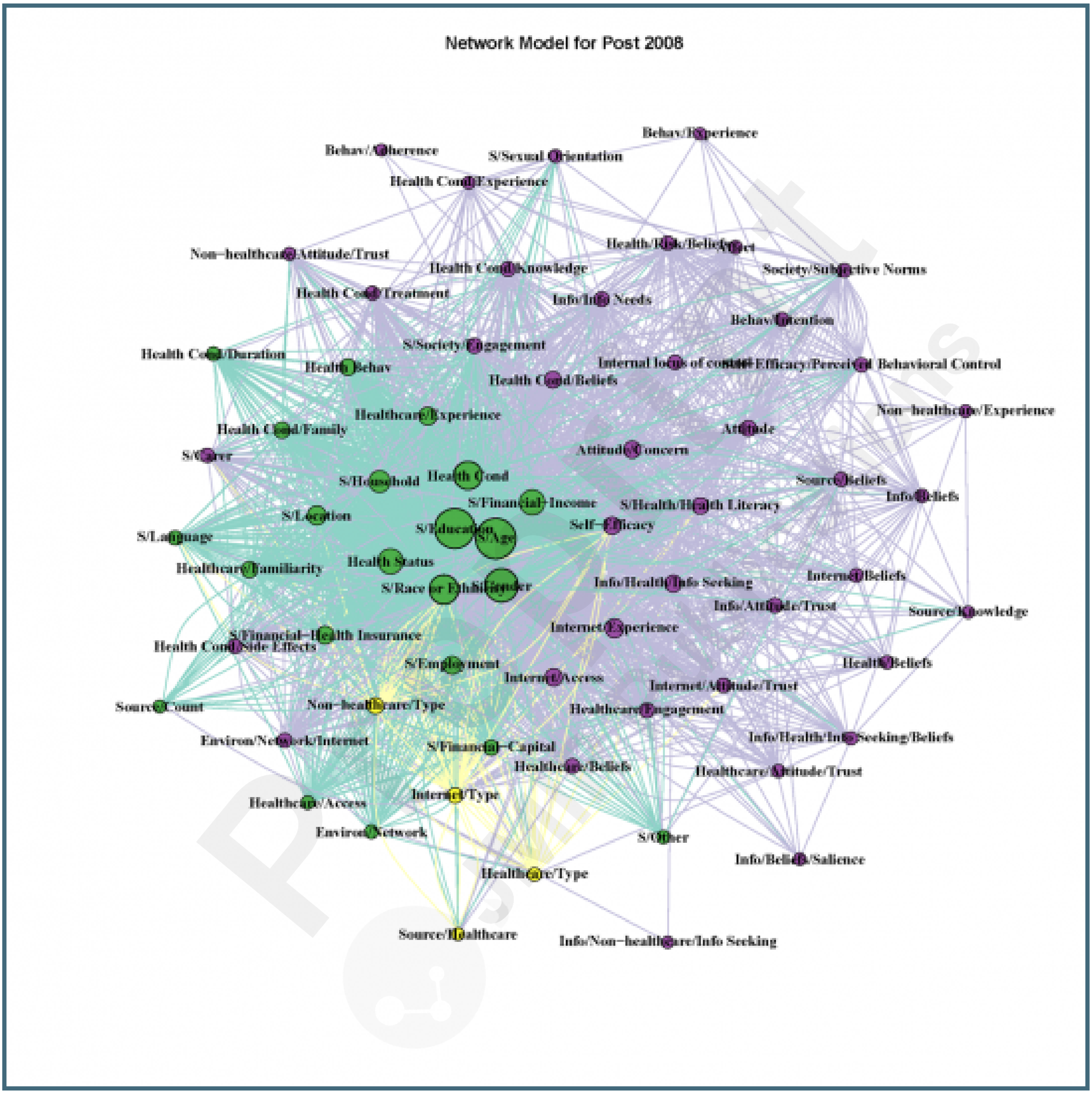
Fruchterman-Reingold layout algorithm of the network analysis comparing the complete model with all years until 2019. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.



Fruchterman-Reingold layout algorithm of the network analysis comparing pre and including 2008 network model. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.



Fruchterman-Reingold layout algorithm of the network analysis comparing post-2008 network model. Color coding according to the modularity group membership of the complete model. Repository available on GitHub.



Multimedia Appendixes

Search terms, Predictors list and their definitions, Network statistics, Network analysis methods.
URL: http://asset.jmir.pub/assets/4e93442ec788345f4fab447bba83b7b9.docx