An analysis of the research content and trend in Health Literacy: A semantic network analysis from 1985 to 2020

***Myung-Il Choi,** Department of Advertising & Public Relations, Namseoul University, CheonAn, 31020, Korea, *jhmi0410@nsu.ac.kr*
*Corresponding author

Abstract. This study analyzes the keywords of health literacy research, network structure and attributes, characteristics of subgroups, and association by research period. The current research conducted a semantic network analysis on 3,971 papers published in PubMed (https://pubmed.ncbi.nlm.nih.gov/) between 1985 and 2020. Seventy-one keywords were selected from 3,284 words. To visualize the network structure of health literacy research, analyze subgroups, and carry out QAP analysis, data matrices of 71*71 was prepared in Ucinet and Netdraw software. Spring-embedded algorithms and degree centrality visualized the network structure. For the sub-group analysis, we employed the modularity using NCD (Newman Community Detection) algorithm. The keywords such as patient (798 times), health (777 times), mental (360 times), adult (323 times), and low (249 times) appeared to be frequent in the health literacy research. Additionally, considering the degree centrality, such keywords as patients, assessments, education, health, low, adult, care, use, improvement, and formation played an important role around health literacy in the overall network structure. We found that the subgroup was best classified as three groups. The first group includes keywords such as develop, health, educate, mental, and formation, and was named 'Development and Evaluation of Health Literacy Measurement Scale.' In the second group, behavior, status, treatment, and adult became the keywords and they named it 'Effects and Outcomes of Health Literacy.' The third group contained keywords including, risk, women, cancer, Chinese, and quality, and was named 'Application and Use of Various Targets of Health Literacy.'

Keywords: Health Literacy, Semantic Network Analysis, Subgroup Analysis, QAP, PubMed

INTRODUCTION

Individuals recently live in a flood of health information. They can get the health information they need if they want to, and they can also ask for help from specialists such as doctors and nurses to address health concerns. Nevertheless, many people still struggle to find, understand, and use health information that is right for them [1]. For example, there are not many people who can properly understand and apply the doctor's prescription and medication manual. This is because the modern health system is generally designed for medical consumers who actively seek health information and have relevant knowledge [2]. Thus, medical consumers would have difficulty understanding warnings of prescription drugs, writing informed consent, and preparing insurance documents. It can be irresponsible to provide only health information without addressing these issues, and the resulting damage can be done not only to individuals but also to society as a whole [3].

'Health Literacy' has been highlighted as a way to solve this social issue. In 1974, efforts to create guidelines for health education for students were the first starting point of health literacy, and active research is underway as proposed by the World Health Organization (WHO) as a key strategy for health promotion in the 1990s [4]. Health Literacy encompasses the cognitive and social skills necessary for active health care and disease prevention [2, 5]. Through this ability, people can improve their communication skills to acquire health-related information, understand statistical figures, create health-related documents, and express their intentions. In addition, health literacy can enable individuals to perform health management on their own, and can be used as a tool for interaction and negotiation to receive appropriate services and management in the health system [6].

Health literacy has been effective in increasing health knowledge, reducing medical use and hospitalization [2, 7], improving health standards, and resolving health inequalities [8] in previous literature. Also, Health literacy played a significant role in affecting hospitalization, emergency room visits, immunization, drug label understanding, and health message learning. In the case of the elderly, health literacy is directly related to mortality[9]. As a result, high rates of health literacy in population promotes individual economic activities and community involvement, making the general life healthy and enriched

[5].

Apart from extensive research into the effects of health literacy on the health care of individuals and society in various fields such as medicine, health communication, and education, there have been few studies that have integrated insight into its research trends and its characteristics. To complement the limitations of the existing research, the current study employed semantic network analysis. This is an analytical technique that interprets phenomena by showing the relationships between words appearing in text as a visual network [10]. Thus, semantic network analysis is useful for discovering invisible meanings or patterns existing in text.

The current study has three main research purposes based on this preliminary investigation. First, this study seeks to identify the keywords and the attributes of the overall network structure that appear in the health literacy studies that have been accumulated so far. Second, through the subgroup analysis, we explore sub-categories of the overall network structure and figure out what attributes each subgroup has. Finally, this study aims to verify the continuity of research by analyzing the development trend of health literacy research in chronological order.

2. Materials and methods

2.1 Data

The current study analyzed the titles of 3,971 research published from 1985 to June 2020 in PudMed, which is a representative academic database of medical and healthcare fields. It is common for the title of the research to present key information and keywords, such as the topic and content of the research. The abstract of a published research can be employed as textual data, but it was excluded from the analysis data because frequently used words, such as question and approach, would be likely to interfere with identifying the association between keywords. In addition, the several keywords presented in the research paper also be analyzed. However, in the past, research in health literacy may not have provided a keyword. Figure 1 shows that research in health literacy was not largely conducted before 2000s, but the study began to be active in earnest from 2010 and increased significantly from 2015.

2.2 Procedure

It goes through 4 steps of semantic network analysis. In the first step, we collected the words presented in the previous literature titles. Such words as prepositions and conjunctions that do not fit the context of the current study were excluded. In addition, words that have the same meaning but are grammatically different were recognized as one word. For example, 'cultrural' and 'culture' are recognized as one meaningful word of culture. A total of 3,284 words collected through this process.

In step 2, the words collected in the first step were set as keywords for the analysis. As shown in the key word selection formula [11], 71 keywords (frequency of appearances more than 50 times) were the final units of analysis by considering the clarity and implications of the research results. Seventy-one keywords accounted for 41.05% of all analyzed words collected in step 1.

$$N = (-1 + \sqrt{(1 + 4d)})/2 \qquad\qquad (1)$$

n= number of final selected keywords, d= total number of keywords

In step 3, selected keywords created a matrix data (71 x 71). Briefly explaining this process, the keywords 'a,' 'b,' 'c,' and 'd' presented in a published research A can produce such six combinations as 'a-b,'
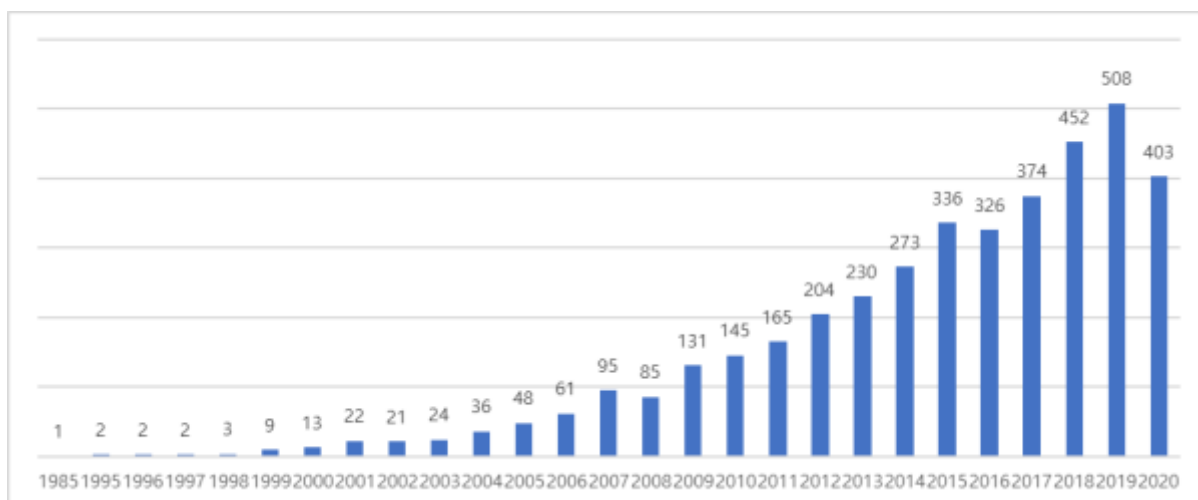
Figure 1 Number of published research in health literacy

'a-c,' 'a-d,' 'b-c,' 'b-d,' and 'c-d.' When the keywords 'a,' 'c,' and 'e' are presented in another published research B, the weighted value of the combination 'a-c' is increased, as well as the centrality of the keywords 'a' and 'c.' If the keywords 'b' and 'f' are presented in paper C, the co-occurrence combination 'b-f' is generated. In this case, 'f' is only linked to 'b', and 'b' serves as a mediator between 'f' and other keywords (a, c, d, and e) [12]. Repeating this process for the entire range of keywords in the present study produces the matrix data.

In step 4, this study completed network analysis and visualization using Ucinet and Netdraw on the matrix data generated in step 3.

3. Results and discussion
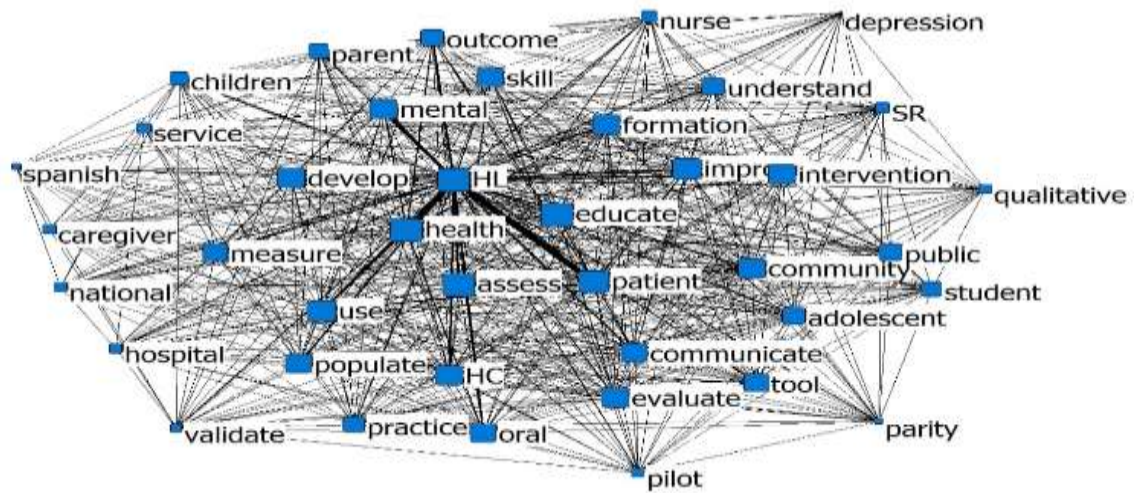### 3.1 Analysis of keywords and network structure

The following keywords; health literacy (4,052 times), patient (798 times), health (777 times), mental (360 times), adult (323 times), low (249 times), care (245 times), improve(242 times), use (219 times) and cancer (217 times) were found to be used frequently (see Table 1).

Table 1 Frequently used keywords

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
Health literacy(HL)	4,052	Diabetes	211	Communicate	162
Patient	798	Assess	200	Health care(HC)	151
Health	777	Old	197	Knowledge	148
Mental	360	Develop	189	Outcome	144
Adult	323	Oral	186	Community	138
Low	249	Populate	185	Women	134
Care	245	Measure	181	Behavior	128
Improve	242	Educate	179	Adherence	128
Use	219	Intervention	168	Student	127
Cancer	217	Formation	168	Manage	124

There is a limit to determining the role and relationship played by each keyword in the overall network structure only with the keyword frequency shown in Table 1. To compensate for this limitation, we visualized the overall network structure (see Figure 2). Keywords that tend to appear simultaneously with other keywords by using the Spring Embedded algorithm are presented in the center of the entire network structure, and the higher the relationship between the keywords, the closer to each other [13]. Also, the size of the key word appears larger as the degree centrality value increases. Degree centrality has an importance score based on the number of links on each node. In other words, it is a value that directly reflects how many relationships one keyword has with other keywords and how important it is in the entire network. As shown in Figure 2, with focus on health literacy, patient, assessment, education, health, low, adult, care, use, improve, formation have a lot of connection with other keywords and play a significant role in the entire network.
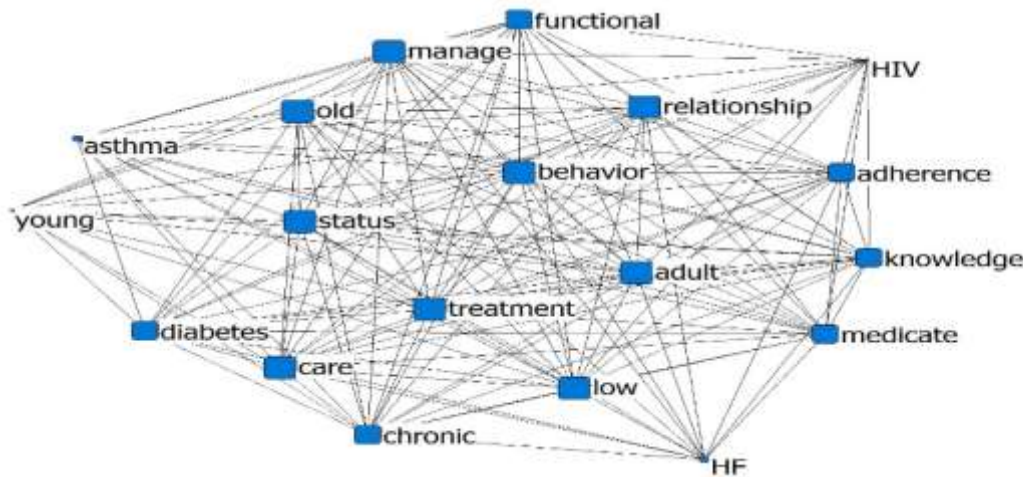
Figure 2 Overall network structure of health literacy research

3.2 Subgroup analysis

To figure out detailed characteristics of health literacy studies, this study employed a subgroup analysis. We identified the optimal subgroups based on the modularity to which the Newman Community Detection (NCD) algorithm was applied. Modularity refers to the degree to which a group is clearly distinguished because the keywords within the subgroup do not overlap with the keywords of other groups, and the connections within groups are greater than those between groups [14]. This study found that it is most appropriate to divide it into three subgroups when the modularity value was .049.

Subgroup 1 contains a total of 40 keywords. Along with health literacy, develop, health, education, mental, and formation were found to play important roles (see Figure 3). We named subgroup 1 Development and Evaluation of Health Literacy Measurement Scales, considering the comprehensive relationship between the keywords.

Subgroup 2 consists of a total of 19 keywords. Key words such as behavior, status, treatment, and adult were found to play an important role (see Figure 4). We named it The Effects and Outcomes of Health Literacy.

We identified that subgroup 3 consists of a total of 12 keywords, such as risk, women, cancer, Chinese, and quality, and named it Apply and Use of Various Objects of Health Literacy (see Figure 5)

Figure 3 Network structure of subgroup 1: Development and Evaluation of Health Li teracy Measurement Scales
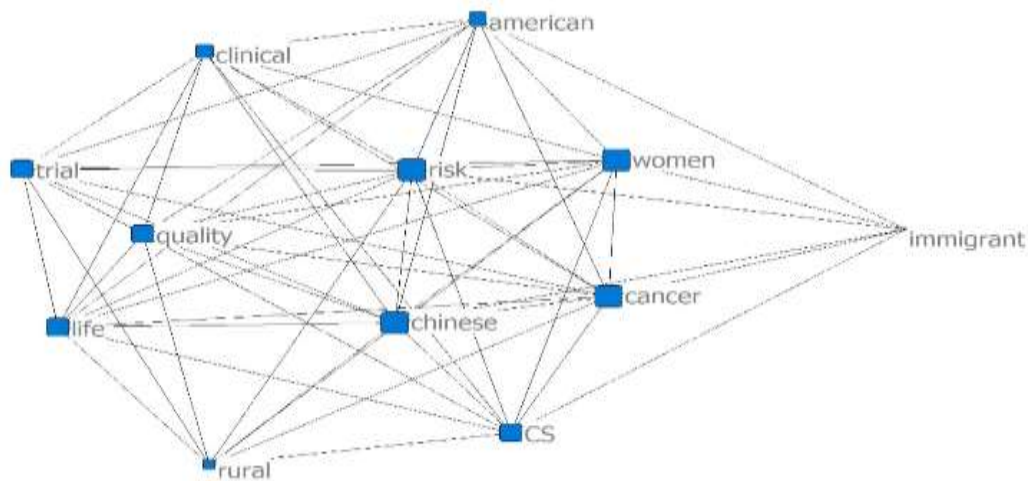
Figure 4 Network structure of subgroup 2: The Effects and Outcomes of Health Li teracy

Figure 5 Network structure of subgroup 3: Apply and Use of Various Objects of Health Literacy

3.3 Continuity of health literacy research: Based on the correlation analysis by research period

For correlation analysis by research period, we classified the period of research in health literacy into three stages. There are 700 published research (17.6%) in the first period between 1985 and 2010. In the second (2011- 2015), there are 1,208 studies (30.4%), and in the third period (2016-2020), there are 2,063 (52%) studies. Moreover, we employ quadratic assignment procedure (QAP) to test the similarity between networks of research in health literacy in each period. To analyze the correlation between networks composed of the similar keywords, QAP randomly rearranges the composition of those keywords to assess for statistical significance. The Pearson correlation coefficient tells the degree of similarity between networks, and the closer the value is to 1, the more similar the structure between networks is [15].

Table 2 shows a high correlation between the first, second, and third periods of health literacy research. The correlation coefficient between the 1st and the 2nd period is .916 ($p < .01$), the coefficient between the 1st and the 3rd period is .893 ($p < .01$), and between the 2nd and the 3rd period the correlation coefficient is .956 ($p<.01$). Thus, the study of health literacy shows high research continuity between the first, second, and third periods.

Table 2 Correlation of health literacy studies between research periods

	1st period	2nd period	3rd period
1st period	1		
2nd period	.916*	1	
3rd period	.893*	.956*	1

*p<.01

4. Conclusion

This study attempts to analyze the attributes of health literacy research, one of the important concepts in the flood of health information. And we employ a semantic network analysis of 3,971 papers for answering research questions.

From the keyword analysis, we find that such keywords as health literacy (4,052 times), patient (798 times), health (777 times), mental (360 times), adult (323 times), low (249 times), care (245 times), improve (242 times), use (219 times), and cancer (217 times) frequently appear in health literacy studies.

Additionally, considering the degree centrality value, those keywords, including health literacy, patient, assess, educate, health, low, adult, care, use, improve, and formation, play a significant role in associating with other keywords. It can be concluded that health literacy research is the mainstream in developing education programs for patients and adults and measuring their effectiveness. These results might be due to a number of functional and practical studies aimed at achieving results such as resolving health inequality [11] and reduced hospitalization and emergency room use [12].

The results of the subgroup analysis explain this situation more clearly. Specifically, we identify three distinctive research trends, such as 'development and evaluation of health literacy measurement scales,' 'effects and outcomes of health literacy,' and 'apply and use of various objects of health literacy.' This research trend reflects the characteristics of modern society in which the role of consumers in understanding and utilizing various health-related information has become important. However, recipients of health information through the media are limited in recognizing the potential impact of fragmented health knowledge on solving their health-related problems. In order to compensate for this challenge, existing health literacy studies have developed a measurement scale with high reliability and validity for analyzing practical effects. Trustful health literacy measures might be beneficial to the planning and operation of health promotion activities at the social level.

Lastly, we investigated how the research on health literacy has chronologically evolved. Health literacy research has developed over three major periods, and the trend of research in each period shows high continuity. Although new media have appeared and consumers' demands for health information have changed, the research subject and purpose of health literacy remains consistent. Thus, we found that health literacy studies have maintained a certain research trend and attributes over the past 35 years.

In sum, there are two primary discoveries through this study. First, the research on health literacy have mainly focused on developing educational programs for various users of health information (patients, adults, etc.) and verifying their effectiveness at the social and individual levels. Second, given the research trends and attributes, health literacy studies have evolved while preserving a consistent direction through the three main research periods.

With the advent of an aging and information-oriented society, the need for health literacy research is growing socially, and the possibility of contributing to the public interest is also increasing. However, since the quantity and quality of research has grown dramatically since 2010, there are relatively few studies that can comprehensively understand studies of health literacy. This study could contribute academically to narrowing this gap and could substantially increase the effectiveness of health campaigns through health literacy. In the future, further studies are necessary to improve a systematic health literacy education program that can contribute to personal and social health management, and to evaluate the effectiveness.

5. Acknowledgements

6. References

[1] Cutilli CC. Do your patients understand?: Determining your patients' health literacy skills. Orthopedic Nursing. 2005;24(5):372-7.
[2] Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Viera A, Crotty D et al. Health literacy interventions and outcomes: An updated systematic review. Evidence Report/Technology Assessment No. 199. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
[3] Schloman BF. Health literacy: A key ingredient for managing personal health. Online Journal of Issues on Nursing. 2004;9(2):14-21.
[4] Nutbeam D. Health promotion glossary. Health Promotion International. 1998;13(4):349-64.
[5] World Health Organization. Health literacy: The solid fact. Copenhagen. Denmark: WHO Regional Office for Europe; 2013
[6] Ishikawa H, Yano E. Patient health literacy and participation in the health-care process. Health Expectations. 2008;11(2):113-22.
[7] Manganello JA. Health literacy and adolescents: A framework and agenda for future research. Health Education Research. 2008;23(5):840-7.
[8] Dodson S, Good S, Osborne RH. Health literacy toolkit for low and middle-income countries: A series of information sheets to empower communities and strengthen health systems. New Delhi: WHO Regional Office for South-East Asia; 2015.

[9] Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. American Journal of Public Health. 2002;92(8):1278-83.

[10] Popping R. Computer-assisted text analysis. London, Thousand Oaks: Sage Publications; 2000.

[11] Donohue. JC. Understanding Scientific Literature: A Bibliographic Approach. Cambridge: The MIT Press; 1973.

[12] Im JH, Lee KS, Kim KY, Hong NS, Lee SW, Bae HJ. Follow-up study on mortality in Korean stroke patients. Journal of the Korean Medical Association. 2011;54(11):1199-208.

[13] Kamada T, Kawai S. An algorithm for drawing general undirected graphs. Information processing letters. 1988;31(1):7-15.

[14] Newman MEJ. Fast algorithm for detecting community structure in networks. Physical Review E. 2004;69(6):1-5.

[15] Martin L. A general permutation-based QAP analysis approach for dyadic data from multiple groups. Connections. 1999;22(2):50-60.