

**HEALTH CONSUMERS' KNOWLEDGE LEARNING  
IN ONLINE HEALTH INFORMATION SEEKING**

by

**Yu Chi**

B.S., Wuhan University, China, 2012

M.S., Wuhan University, China, 2015

Submitted to the Graduate Faculty of  
the School of Computing and Information in partial fulfillment  
of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH  
SCHOOL OF COMPUTING AND INFORMATION

This dissertation was presented

by

Yu Chi

It was defended on

July 26th 2021

and approved by

Dr. Daqing He, School of Computing and Information, University of Pittsburgh

Dr. Eleanor Mattern, School of Computing and Information, University of Pittsburgh

Dr. Kayla Booth, School of Computing and Information, University of Pittsburgh

Dr. Young Ji Lee, School of Nursing, University of Pittsburgh

Dissertation Director: Dr. Daqing He, School of Computing and Information, University of

Pittsburgh

Copyright © by Yu Chi  
2021

# HEALTH CONSUMERS' KNOWLEDGE LEARNING IN ONLINE HEALTH INFORMATION SEEKING

Yu Chi, PhD

University of Pittsburgh, 2021

With the increasing awareness of health consumers as active information seekers, the past decade has witnessed a shifting research interest from a physician-centered paradigm to a consumer-centered paradigm. Online health information seeking (OHIS) has become pervasive, with critical impacts on consumers' health. However, the inherent complexity and the uniqueness of health tasks pose new challenges to consumers in OHIS, such as a lack of adequate knowledge to formulate queries and evaluate the online resources with various qualities. OHIS is, by nature, a learning-oriented behavior, and knowledge learning is a critical component and outcome of consumers' OHIS. On the other hand, studies in the area of search as learning (SAL) have demonstrated that learning is a common phenomenon in the information-seeking process. However, the existing studies in OHIS mainly concentrated on viewing consumers' domain knowledge as a fixed value, even though consumers are involved in the knowledge learning in the OHIS. Therefore, this dissertation proposes a conceptual framework of health information search as learning (HearSAL) by linking the related models and prior studies from the two areas — OHIS and SAL — and conducts a systematic study to understand what, how, and how well health consumers can search and learn in online health information seeking, particularly for three increasing levels of learning objectives: Understand, Analyze and Evaluate.

Two representative health consumer groups, laypeople and cancer patients, are targeted in this dissertation study because they share the common issue of facing barriers in searching and learning in OHIS, yet they are different due to prior topic knowledge, learning duration, and learning expectation. Following the conceptual framework HearSAL, four sub-studies are conducted with emphasis on different dimensions of health consumers' search as learning in OHIS, including the following: Study 1: a user study with laypeople that examines the method dimension (e.g., search behaviors and source selections); Study 2: an analysis of an

ovarian cancer online health community that reveals the information dimension (e.g., types and amount of information); Study 3: interviews with laypeople; and Study 4: interviews with ovarian cancer patients and caregivers. The two complementary interviews highlight the outcomes of OHIS.

Major results demonstrate that, (1) health consumers' SAL behaviors and sources vary by different levels of learning objectives, and the variation is affected by the severity of health conditions; (2) Analyze is the most prevalent learning objective in the online health community, while the amount of informational support is the highest in the Evaluate level; (3) Though consumers' prior knowledge of the Understand level is the highest, compared to higher levels, consumers still tend to achieve the most knowledge increase in the Understand level of learning; and (4) Receiving more informational support drives consumers to increase the level of learning objectives.

This dissertation makes empirical, practical, theoretical and methodological contributions. The empirical studies of laypeople and ovarian cancer patients provide a deeper insight into health consumers' SAL behavior and performance in today's web environment. Based on the empirical results, practical implications are proposed for designing consumer-centered health information systems, which facilitate seeking and enhance learning. Finally, the HearSAL framework and its application in this study can serve as a theoretical and methodological basis for future explorations.

## Table of Contents

|            |   |    |
|------------|---|----|
| <b>1.0</b> | <b>Introduction</b>   | 1  |
| 1.1        | Overview  | 1  |
| 1.2        | Problem Statement   | 3  |
| 1.2.1      | Internet as a Health Information Resource: Prevalence and Challenge | 3  |
| 1.2.2      | Search as Learning as an Emerging Research Agenda: Rise and Themes  | 4  |
| 1.2.3      | Two Health Consumer Groups: Laypeople and Cancer Patients           | 6  |
| 1.3        | Research Motivation   | 9  |
| 1.4        | Significance and Contribution                                       | 10 |
| 1.5        | Overview of the Chapter Structure                                   | 12 |
| <b>2.0</b> | <b>Literature Review</b>  | 13 |
| 2.1        | Health Information Seeking Behavior                                 | 13 |
| 2.1.1      | Concepts  | 13 |
| 2.1.2      | Theories and Models   | 14 |
| 2.1.3      | Empirical Studies   | 16 |
| 2.1.3.1    | Method Dimension  | 16 |
| 2.1.3.2    | Information Dimension   | 17 |
| 2.2        | Search as Learning  | 19 |
| 2.2.1      | Concepts  | 19 |
| 2.2.2      | Theories and Models   | 19 |
| 2.2.3      | Empirical Studies   | 22 |
| 2.2.3.1    | Learning During the Information Seeking Process                     | 22 |
| 2.2.3.2    | Knowledge Change as an Outcome of HIS                               | 23 |
| 2.3        | Health Consumers' Search as Learning in OHIS                        | 25 |
| 2.3.1      | Domain Knowledge's Influence on Consumers' OHIS                     | 25 |
| 2.3.2      | Health Condition's Influence on Consumers' OHIS                     | 26 |
| 2.3.3      | Laypeople's Search as Learning in OHIS                              | 26 |

|            |   |           |
|------------|---|-----------|
| 2.3.3.1    | Laypeople’s OHIS: How They Search and Learn . . . . .   | 27        |
| 2.3.3.2    | Laypeople’s OHIS: What They Search and Learn . . . . .  | 28        |
| 2.3.4      | Ovarian Cancer Patients’ Search as Learning in OHIS . . . . .   | 29        |
| 2.3.4.1    | OvCa Patients’ OHIS: How They Search and Learn . . . . .  | 30        |
| 2.3.4.2    | OvCa Patients’ OHIS: What They Search and Learn . . . . .   | 31        |
| <b>3.0</b> | <b>Conceptual Framework and Research Questions . . . . .</b>  | <b>34</b> |
| 3.1        | Conceptual Framework . . . . .  | 34        |
| 3.1.1      | Levels of Learning . . . . .  | 35        |
| 3.1.2      | Characteristics of OHIS . . . . .   | 35        |
| 3.1.3      | Knowledge Change . . . . .  | 36        |
| 3.2        | Research Questions . . . . .  | 36        |
| <b>4.0</b> | <b>Methodology . . . . .</b>  | <b>38</b> |
| 4.1        | Research Design Overview . . . . .  | 38        |
| 4.2        | Data Collection Overview . . . . .  | 39        |
| <b>5.0</b> | <b>Study 1: Laypeople’s Health-related Search as Learning Behaviors on<br/>the Internet . . . . .</b> | <b>42</b> |
| 5.1        | Data Collection . . . . .   | 42        |
| 5.1.1      | Health Condition and Task Design . . . . .  | 42        |
| 5.1.1.1    | Scenario . . . . .  | 43        |
| 5.1.1.2    | Task Design . . . . .   | 43        |
| 5.1.2      | System and Procedure . . . . .  | 45        |
| 5.1.2.1    | Search System . . . . .   | 45        |
| 5.1.2.2    | Capturing Source Selection Behavior . . . . .   | 46        |
| 5.1.2.3    | Procedure . . . . .   | 47        |
| 5.1.3      | Participants and Dataset . . . . .  | 49        |
| 5.2        | Data Analysis . . . . .   | 50        |
| 5.2.1      | Search Behavior Measures . . . . .  | 50        |
| 5.2.2      | Online Health Source Classification . . . . .   | 51        |
| 5.2.3      | Written Summary Assessment . . . . .  | 53        |
| 5.3        | Results . . . . .   | 54        |

|            |  |           |
|------------|--|-----------|
| 5.3.1      | Search Behaviors . . . . .   | 54        |
| 5.3.1.1    | Search Queries . . . . .   | 54        |
| 5.3.1.2    | Clicks . . . . .   | 57        |
| 5.3.1.3    | Search Efficiency . . . . .  | 58        |
| 5.3.2      | Source Selections . . . . .  | 59        |
| 5.3.3      | Knowledge Change . . . . .   | 61        |
| 5.4        | Summary of Study 1 . . . . .   | 64        |
| <b>6.0</b> | <b>Study 2: OvCa Patients and Caregivers' Search as Learning Needs<br/>and Supports in Online Health Community . . . . .</b> | <b>66</b> |
| 6.1        | Data Collection . . . . .  | 66        |
| 6.1.1      | Research Setting: National Ovarian Cancer Coalition CancerCon-<br>nect Community . . . . .                                   | 66        |
| 6.1.2      | OvCa Survivors' Posts and Comments . . . . .   | 67        |
| 6.2        | Data Analysis . . . . .  | 69        |
| 6.2.1      | Identifying Initial Posts with Information Needs and Levels of Learning  | 69        |
| 6.2.2      | Identifying Topics of Information Needs in the Initial Post . . . . .  | 71        |
| 6.2.3      | Identifying Other Characteristics of Information Needs in the Initial<br>Post . . . . .                                      | 72        |
| 6.2.4      | Identifying Support Types in the Comments . . . . .  | 72        |
| 6.3        | Results . . . . .  | 74        |
| 6.3.1      | Levels of Learning in the Initial Posts . . . . .  | 75        |
| 6.3.1.1    | Distribution of Different Levels of Learning . . . . .   | 75        |
| 6.3.1.2    | Number of Topics . . . . .   | 76        |
| 6.3.1.3    | Category of Topics . . . . .   | 77        |
| 6.3.1.4    | Self-disclosure . . . . .  | 80        |
| 6.3.1.5    | Role of Posters . . . . .  | 80        |
| 6.3.2      | Support for Different Levels of Learning in the Comments . . . . .   | 82        |
| 6.3.2.1    | Number of Comments . . . . .   | 82        |
| 6.3.2.2    | Types of Support . . . . .   | 82        |
| 6.3.3      | Knowledge Change . . . . .   | 86        |



|            |  |            |
|------------|--|------------|
| 6.3.3.1    | Current Level and Knowledge Change in the Subsequent Post                      | 87         |
| 6.3.3.2    | Support Received and Knowledge Change . . . . .                                | 89         |
| 6.4        | Summary of Study 2 . . . . .   | 90         |
| <b>7.0</b> | <b>Study 3: Laypeople’s Outcomes of OHIS . . . . .</b>                         | <b>91</b>  |
| 7.1        | Data Collection . . . . .  | 91         |
| 7.1.1      | Setting and Sampling . . . . .   | 91         |
| 7.1.2      | Interview Design . . . . .   | 91         |
| 7.2        | Data Analysis . . . . .  | 92         |
| 7.3        | Results: Characterize Laypeople’s Own OHIS Story . . . . .                     | 93         |
| 7.3.1      | Information Needs . . . . .  | 94         |
| 7.3.2      | Search Starting-points . . . . .   | 96         |
| 7.3.3      | Emotion Changes . . . . .  | 97         |
| 7.3.4      | Decision-makings . . . . .   | 98         |
| 7.3.5      | Transitions between the Stages . . . . .                                       | 99         |
| 7.4        | Summary of Study 3 . . . . .   | 100        |
| <b>8.0</b> | <b>Study 4: OvCa Patients and Caregivers’ Outcomes of OHIS . . . . .</b>       | <b>101</b> |
| 8.1        | Data Collection . . . . .  | 101        |
| 8.2        | Data Analysis . . . . .  | 102        |
| 8.3        | Results: OvCa Patients and Caregivers’ Outcomes of OHIS . . . . .              | 103        |
| 8.3.1      | What Leads to Positive Affective and Cognitive Outcomes? . . . . .             | 103        |
| 8.3.2      | What Causes Negative Affective and Cognitive Outcomes? . . . . .               | 104        |
| 8.3.3      | Behavioral Outcomes Associated with Affective and Cognitive Outcomes . . . . . | 106        |
| 8.3.4      | Online Health Community: A Double-edged Sword . . . . .                        | 108        |
| 8.4        | Summary of Study 4 . . . . .   | 109        |
| <b>9.0</b> | <b>Discussions and Implications . . . . .</b>                                  | <b>111</b> |
| 9.1        | Discussion of Results . . . . .  | 111        |
| 9.1.1      | Method Dimension . . . . .   | 111        |
| 9.1.2      | Information Dimension . . . . .  | 114        |
| 9.1.3      | Knowledge Change . . . . .   | 118        |

|  |  |            |
|--|--|------------|
| 9.1.3.1  | Knowledge Change Varies by Levels of Learning Objectives       | 119        |
| 9.1.3.2  | Knowledge Change Varies by Health Conditions . . . . .         | 119        |
| 9.1.3.3  | Knowledge Change Varies by Amount of Support Received .        | 120        |
| 9.1.3.4  | Knowledge Change as a Type of Cognitive Outcome . . . . .      | 121        |
| 9.1.3.5  | Knowledge Change and Emotion Change . . . . .                  | 122        |
| 9.2  | Implications for Design . . . . .                              | 122        |
| 9.2.1  | Adapting to Consumers' Information Needs and Search Contexts . | 123        |
| 9.2.1.1  | Adapting to Learning Objectives . . . . .                      | 123        |
| 9.2.1.2  | Adapting to Health Conditions . . . . .                        | 124        |
| 9.2.2  | Advancing Search Tools and Platforms . . . . .                 | 125        |
| 9.2.2.1  | Advancing Search Engines . . . . .                             | 125        |
| 9.2.2.2  | Advancing Online Health Communities . . . . .                  | 126        |
| 9.2.3  | Promoting Desirable Outcomes . . . . .                         | 126        |
| 9.3  | Theoretical and Methodological Implications . . . . .          | 128        |
| 9.3.1  | HearSAL Framework . . . . .                                    | 128        |
| 9.3.2  | Coding Frameworks . . . . .                                    | 129        |
| 9.3.2.1  | Classification of Online Health Sources Based on Genre . .     | 129        |
| 9.3.2.2  | Approaches and Frameworks of Knowledge State Assessment        | 129        |
| 9.3.2.3  | Framework of Task Design . . . . .                             | 130        |
| 9.3.3  | Research Design . . . . .                                      | 130        |
| 9.3.3.1  | Mixing Results from Two User Groups . . . . .                  | 130        |
| 9.3.3.2  | Mixing Results from Different Types of Data . . . . .          | 131        |
| <b>10.0</b>  | <b>Conclusion . . . . .</b>                                    | <b>132</b> |
| 10.1   | Limitations . . . . .  | 132        |
| 10.2   | Future Work . . . . .  | 133        |
| 10.3   | Conclusion . . . . .   | 134        |
| <b>Appendix A. User Study Introductory Script . . . . .</b>                  |  | <b>136</b> |
| <b>Appendix B. User Study Entry Questionnaire . . . . .</b>                  |  | <b>138</b> |
| <b>Appendix C. Laypeople Interview Protocol . . . . .</b>                    |  | <b>140</b> |
| <b>Appendix D. OvCa Patients and Caregivers Interview Protocol . . . . .</b> |  | <b>142</b> |

References . . . . . 144

## List of Tables

|    |  |    |
|----|--|----|
| 1  | Overview of Data Collection Methods and Datasets . . . . .   | 40 |
| 2  | The Mapping Between OHIS Intentions and Learning Objectives . . . . .  | 43 |
| 3  | Complete Description of Task Scenarios and Subtasks . . . . .  | 44 |
| 4  | Coding Framework of Webpage Types . . . . .  | 52 |
| 5  | Written Summary Assessment Criteria . . . . .  | 54 |
| 6  | Summative Results of Search Behaviors . . . . .  | 64 |
| 7  | Coding Framework of Levels of Learning Objectives in the Initial Post . . . . .  | 70 |
| 8  | Coding Framework of Topics of Information Needs in the Initial Post . . . . .  | 71 |
| 9  | Inter-rater Agreement between Human Annotators and Classification Score for<br>Support Types in the Comments . . . . . | 74 |
| 10 | Number of Topics of Each Learning Level . . . . .  | 77 |

## List of Figures

|    |   |    |
|----|---|----|
| 1  | Anderson and Krathwohl’s Taxonomy of Learning . . . . .   | 21 |
| 2  | Conceptual Framework of Health Information Search As Learning (HearSAL)   | 34 |
| 3  | An illustration of the Research Design Overview . . . . .   | 38 |
| 4  | A Screenshot of the Homepage of HIS . . . . .   | 45 |
| 5  | An Example of the Search Result Page . . . . .  | 46 |
| 6  | An Example of the ”Save to Workspace” Interface(a): the user is highlighting<br>a piece of information and then clicking ”save to workspace.” . . . . .       | 47 |
| 7  | An Example of the ”Save to Workspace” Interface(b): A pop-up window of<br>the snippet that is going to be saved. . . . .                                      | 47 |
| 8  | User Study Procedure . . . . .  | 48 |
| 9  | a) Number of Queries (NQ) and b) Length of Queries (LQ) . . . . .   | 55 |
| 10 | a) Number of Unique Queries (NUQ) and b) Number of Unique Terms (NUT)   | 56 |
| 11 | Clicks of the Three Levels of Learning in the Two Conditions NCP: Number of<br>Clicked Pages; NSP: Number of Saved Pages; NUP: Number of Used Pages . . . . . | 57 |
| 12 | Search Efficiency of the Three Levels of Learning . . . . .   | 59 |
| 13 | Types of Selected Sources in the Severe Condition (SC) . . . . .  | 60 |
| 14 | Types of Selected Sources in the Mild Condition (MC) . . . . .  | 60 |
| 15 | Prior-Search Knowledge (PK) and After-Search Knowledge (AK) . . . . .   | 61 |
| 16 | Knowledge Change . . . . .  | 62 |
| 17 | An Example of Initial Post . . . . .  | 67 |
| 18 | Examples of Comments . . . . .  | 68 |
| 19 | Distribution of Initial Posts with Different Levels of Learning Objective U -<br>Understand, A - Analyze, E - Evaluate . . . . .                              | 75 |
| 20 | Distribution of Number of Topics per Post . . . . .   | 76 |

|    |   |    |
|----|---|----|
| 21 | Distribution of Topics of Interest in Each Learning Level. DM-Disease Management; SM-Symptom Management; TM-Treatment; TD-Treatment Decision; EM-Emotion Management; SF-Self-management; PN-Practical Needs; CG-Caregiving; OT-Others . . . . . | 78 |
| 22 | Co-occurrence of Topic Pairs in One Post. (N.B., darker color indicates larger proportions) . . . . .   | 80 |
| 23 | Distribution of Posts with or without Self-disclosure in Each Learning Level (NB. w/SD - post with self-disclosure, w/o SD - post without self-disclosure)  | 81 |
| 24 | Distributions of Roles of the Posters in Each Learning Level. (NB. P - patient, CG - caregiver) . . . . .   | 82 |
| 25 | Average Number of Comments per Post in Each Learning Level . . . . .  | 83 |
| 26 | Correlation Matrix of Different Types of Support in One Comment (Advice, Referral, Fact, Personal Experience, Opinion, Emotional Support) . . . . .   | 84 |
| 27 | Amount of Different Types of Support Received in Each Learning Level . . .  | 85 |
| 28 | Distribution of Number of Posts with Information Needs per User . . . . .   | 87 |
| 29 | Change of Learning Objectives U- Understand, A- Analyze, E- Evaluate . . .  | 88 |
| 30 | Support Received and the Change of Learning Objectives . . . . .  | 89 |
| 31 | Four Important Aspects of OHIS Process and the Transitions Between the Adjacent Aspects . . . . .   | 95 |

## 1.0 Introduction

This chapter presents an overview of the research challenges, elaborates on the problem statements, and summarizes the research motivations and the significance of this dissertation. Finally, an overview of the chapter structure is displayed.

### 1.1 Overview

With a dramatic growth of the available health information on the Internet, an increasing number of health consumers actively use the Internet as a resource to obtain health-related information (Fox, 2013). **Online health information seeking (OHIS)**, in which consumers use the Internet to seek and receive information to address health concerns, has brought unique benefits and made it different from traditional health information seeking. For example, given its confidentiality and anonymity, OHIS could protect patients when they seek help with certain sensitive or private health problems (Kim, Park, & Bozeman, 2011; Cline & Haynes, 2001). Beyond informational support, the Internet also facilitates the process of seeking emotional support and social support for patients on their health issues from a much wider range of user populations (Chung, 2014).

Despite the above benefits, the inherent complexity and the uniqueness of health tasks also pose new challenges to consumers in their OHIS, such as a lack of adequate knowledge to formulate queries and evaluate online resources of varying quality (Cline & Haynes, 2001; Rice, 2006; White, Dumais, & Teevan, 2009). The fact that online health information is regularly full of medical terminology only makes this problem worse. Worse yet, such information does not guarantee effective knowledge learning.

**Search as learning (SAL)** is an emerging research agenda grounded in multiple research fields, including library and information science, educational psychology, learning analytics, and information retrieval (Hoppe et al., 2018). SAL researchers believe that knowledge learning is an implicit parallel process with the explicit information seeking process. As a

result, the seeker's knowledge change is an outcome of the SAL process.

This dissertation investigates health consumers' SAL behaviors in OHIS. This is a promising research topic for the following reasons:

Firstly, health-related information seeking is a typical type of domain-specific search in which people who possess higher domain knowledge are more likely to succeed. Domain knowledge has been viewed as a crucial implicit factor influencing search behaviors, strategies, and performance in health information seeking research (Hölscher & Strube, 2000; Jenkins, Corritore, & Wiedenbeck, 2003). The existing studies in HIS, however, mainly concentrated on viewing one's domain knowledge as a fixed value; regardless of this, individuals' domain knowledge can be increased even during the short-term OHIS process. Through OHIS, consumers are involved in implicit and explicit knowledge learning. Therefore, the researcher argue that OHIS is, by nature, a learning-oriented activity, and knowledge learning is a critical component and outcome of consumers' OHIS.

Secondly, even in the substantial SAL literature, there is very limited work investigating the interplay between health consumers' behaviors and their knowledge learning outcomes. Previous work suggests that domain experts and novices behave differently in searching for health-related information (White et al., 2009; Wildemuth, 2004), while, on the other hand, there exist studies that demonstrate one's knowledge level is predictable from her searching behaviors (Yu et al., 2018; J. Liu, Liu, & Belkin, 2016). Therefore, this motivated this work to investigate the factors that affect the learning outcomes in the dynamic OHIS process. The goal is to alleviate the difficulties and provide the demanding facilities that would result in a better SAL experience in the design of the domain-specific search system.

Thirdly, this dissertation work aims to examine two representative groups of health consumers: laypeople and cancer patients. The uniqueness of health consumer groups inspires the selection of these two groups (Goldner, 2006; Baggott, Allsop, & Jones, 2014). Health consumers are a vast population of people, including laypeople who are interested in general health information but have limited prior knowledge about health issues, as well as cancer survivors with the presence of a medical condition and some experience living with the condition. Different health consumer groups could behave very differently in online health information seeking in many aspects, such as information needs, source selection, and search



behaviors (Y. Zhang, Sun, & Kim, 2017; De Choudhury, Morris, & White, 2014). However, it remains unclear how one's health status would influence SAL in online health information seeking.

Specifically, I believe it is beneficial to include laypeople and cancer behaviors because of their unique characteristics regarding learning in OHIS.

On the one hand, both laypeople and cancer patients are likely to be involved in knowledge learning in OHIS and facing similar barriers such as lack of topic knowledge. On the other hand, they are different concerning prior topic knowledge, learning duration, and learning expectation.

Overall, with more and more exploration in the SAL research area, studying learning in the OHIS is a promising research agenda. Therefore, to bridge the gap in the literature, this dissertation attempts to link the related framework and prior studies in the fields of OHIS and SAL, and conducts a systematic study to understand *what*, *how* and *how well* health consumers (e.g., laypeople and cancer patients) search and learn in online health information seeking.

## 1.2 Problem Statement

### 1.2.1 Internet as a Health Information Resource: Prevalence and Challenge

Though the healthcare provider is a preferred information source, there are some limitations and constraints restricting health consumers from addressing their information needs through communication with healthcare providers. Firstly, healthcare providers are sometimes inaccessible or unavailable to consumers. Secondly, for some health conditions, like gynecological cancers, patients are not always willing to share all the information needs to the providers, such as needs regarding sexuality and intimacy. Therefore, the Internet, with its anonymity and pervasiveness, increasingly becomes an information source for many health consumers.

Nowadays, due to an explosion in the amount of online health information, more and

more people use online resources to seek this type of information. According to the Health Fact Sheet by the Pew Research Centre’s national wide project (Fox, 2013), 72% of the US Internet users have looked online for health information. They also reported that the online information affected their decisions on how to treat an illness, maintain their health, see a doctor, cope with a chronic condition, and even whether to see a doctor.

Despite the prevalence of OHIS, the Internet as a health information resource lays additional challenges upon health consumers. Above all, not all health consumers are equal in the context of OHIS. In general, health information seeking is a typical type of domain-specific search (Hanbury, 2012), in which a high level of domain knowledge is required to conduct searches successfully. Therefore, professionals who possess higher knowledge tend to have better strategies and higher success rates in solving health information seeking tasks (Bhavnani, 2002; White et al., 2009). However, for consumers who do not possess enough knowledge about their health issues, the current general-purpose search engine is not sufficient to support their health-related searches (Bhavnani, 2002).

While an increasing population relies on OHIS to address their health information needs—and while it is known that in this kind of domain-specific information seeking, learning plays an important role as both a by-product of the search process and an influential factor determining search performance (Bhavnani, 2002; White et al., 2009)—there is little existing work that combines learning and health information seeking, to my best knowledge.

### **1.2.2 Search as Learning as an Emerging Research Agenda: Rise and Themes**

SAL is recognized as a separate research agenda and emerged from several research fields, including information science, educational psychology, learning analytics, and information retrieval (Hoppe et al., 2018). Though examining the SAL process is not a new topic in the literature, research interests in SAL have arisen and gained plenty of attention, mainly through a series of information retrieval workshops, in recent years.

In 2012, participants in the SWIRL (The second strategic workshop on information retrieval in Lorne) proposed that information retrieval systems should support learning experiences by introducing a variety of tools for exploring information (Allan, Croft, Moffat,

& Sanderson, 2012). Agosti, Fuhr, Toms, and Vakkari (2014) suggested a research agenda called *"from searching to learning"* in the 2013 Dagstuhl seminar. It emphasized the importance of conceptualizing learning as a searching outcome. Sequentially, in the following years, two Searching as Learning (SAL) Workshop took place, with one co-located with IIX 2014 (Information Interaction in Context conference) and one with ACM SIGIR 2016 (Special Interest Group on Information Retrieval) (Gwizdka, Hansen, Hauff, He, & Kando, 2016). The next year, a seminar titled "Search as Learning" was held in Dagstuhl, Germany, in February 2017. Researchers from different backgrounds presented and discussed four big research themes of "(i) understanding search as a human learning process; (ii) the measurement of learning performance and learning outcomes during search; (iii) the relationship between the learning process and the search context; and (iv) the design of functionalities and search system interventions to promote learning" (Collins-Thompson, Hansen, & Hauff, 2017). Most recently, in October 2019, researchers in the first SALMM (Search as Learning with Multimedia Information) started to explore and discuss search as learning with multimedia information beyond textual documents (Ewerth, Dietze, Hoppe, & Yu, 2019).

In information science, it has long been known that knowledge acquisition is an essential component in the information-seeking process. As stated in the Anomalous State of Knowledge (ASK) model, Belkin (1980) argued that information seeking is a process to resolve the anomaly between users' current states of knowledge the problem they faced. Marchionini (2006) claimed that beyond simple lookup search, people often engage in exploratory search tasks in which learning and investigation could play essential roles.

Additionally, empirical studies have proved that the domain-specific knowledge of a person is not fixed. Moreover, the information-seeking process, though short, allows the users to gain domain-specific knowledge in their interactions with online information sources. Numerous empirical studies have provided substantial evidence suggesting learning to be a widespread phenomenon in people's search processes (Eickhoff, Teevan, White, & Dumais, 2014; Vakkari, Penmanen, & Serola, 2003; Chi, Han, He, & Meng, 2016; Collins-Thompson, Rieh, Haynes, & Syed, 2016).

Despite the research interests and great research efforts devoted to understanding or modeling SAL behaviors, there is limited research conducted in the context of health-related

information seeking. Little is known about the factors that might enhance or impede such learning during OHIS. Therefore, this dissertation aims to examine health consumers' SAL behaviors and performances.

### 1.2.3 Two Health Consumer Groups: Laypeople and Cancer Patients

With the increasing awareness of health consumers as active information seekers rather than passive information recipients, the past decade has witnessed a shifting research interest from a physician-centered paradigm to a consumer-centered paradigm (Zeng et al., 2004; Y. Zhang, 2012). However, it is also notable that falling under the umbrella term of "health consumer," individuals can behave very differently due to a diverse range of factors, such as health conditions, health tasks, and health status (Y. Zhang et al., 2017; De Choudhury et al., 2014).

In this dissertation, two particular health consumer groups—laypeople and cancer patients—are chosen because they share common barriers in learning but are unique at the same time.

On the one hand, both laypeople and cancer patients are likely to be involved in knowledge learning in OHIS and face similar barriers regarding a lack of topic knowledge. Due to the expansion of online health information, many consumers have to navigate through webpages of varying quality. However, for those who are not adequately equipped with health-related knowledge, it can be challenging to precisely articulate information needs (Pang, Chang, Pearce, & Verspoor, 2014) and accurately judge the relevance and quality of the online content (Pian, Khoo, & Chang, 2016).

On the other hand, laypeople and cancer patients are different concerning *prior topic knowledge, learning duration, and learning expectation*.

**Unique Differences between Laypeople and Cancer Patients** Above all, the prior topic knowledge of the two groups of health consumers is different. The most common reason for laypeople to conduct OHIS is for self-diagnosing their uncertain symptoms or health concerns (Fox, 2013), which means many laypeople have very limited or even no prior knowledge about their conditions. As a result, laypeople mainly face two important challenges when

seeking health information online. Firstly, they often have inadequate knowledge about their health issues, so finding the right query terms to describe their information needs is difficult (D. L. Hansen et al., 2003; Pang et al., 2014). Additionally, laypeople often lack knowledge about which online sources contain relevant information at a suitable level for them to understand (Pang et al., 2016).

Conversely, though cancer patients are also not experts in the health and medical domain, living with cancer for years, patients' prior knowledge about their health conditions is expected to be generally higher than laypeople and is dynamic over the disease trajectory. The dynamic knowledge state, in turn, further determines what information the patient needs, how one expresses the needs, and what information is at the right knowledge level to be understood. However, online health information currently is organized or provided, disregarding the health consumers' knowledge levels. It remains unclear how the different groups of consumers' preexisting knowledge affect the SAL process and how well they can conduct SAL through OHIS.

In addition, the SAL duration of the two groups is different. Usually motivated by a health-related information need, such as an uncertain symptoms, laypeople are more likely to conduct a short-term OHIS, while, in most cases, cancer patients' SAL is a constant and long-term behavior. As previous work suggested, many laypeople simply start their search by issuing queries to a search engine (Eysenbach & Köhler, 2002; Fox, 2013; D. L. Hansen et al., 2003). In D. L. Hansen et al. (2003), 87% of the laypeople visited sites directly from search results or search engines' recommendations. The whole SAL session for laypeople is relatively short term compared to a cancer patients who have constant health-related information needs. Following a cancer diagnosis, cancer patients have an increasing need for information. They are eager to learn assorted information and knowledge about their condition. Over the course of survivorship, cancer patients have various needs of self-care, self-monitoring, and self-management. To satisfy these needs, they constantly demand and acquire information through multiple sources, thus being involved in a long-term SAL course with different search strategies and source preferences.

Finally, because of the presence and severity of the health condition, the two groups' learning expectations are also anticipated to be different. Cancer patients are more likely to

be pursuing health information with a higher level of domain knowledge, such as treatment-related information.

**Cancer Survivorship** Globally, cancer is a critical public health problem and affects many people's health and quality of life. In the United States, according to the American Cancer Society's latest cancer statistics, there were 1.8 million new cancer cases diagnosed and 606,520 deaths due to cancer in 2020 (Siegel, Miller, & Jemal, 2020). Cancer is the second-leading cause of death in the United States after heart disease. In recent years, substantial efforts have been devoted to fighting against cancer. Accordingly, the overall cancer death rate decreased by 26% from 1991 to 2015 (*Cancer Statistics*, n.d.), per the data of the National Cancer Institute, following by the largest drop rate of 2.2% from 2016 to 2017 (Siegel et al., 2020). The decline of the cancer death rate effectively demonstrates the progress in the fight against cancer. With the help of progressive treatment and early diagnosis, cancer survivorship has been continuously extended since 1975. From 2009 to 2015, the five-year relative survival rate for all diagnosed cancer was 67% in the United States (Siegel et al., 2020), and the number of cancer survivors is speculated to reach 18 million by 2022 (Howlander et al., 2012).

As a result, living with cancer increasingly becomes a critical problem. Many cancer survivors experience physical and psychosocial consequences (Denlinger et al., 2014). To deal with numerous issues in survivorship, survivors express varied information needs, and they consult various sources to address their information needs (Nagler et al., 2010).

**Ovarian Cancer Patients** Given the variety of cancers, this study further narrows down the scope and focuses on ovarian cancer (OvCa) patients. Though there is solid work in cancer survivor's online HISB, OvCa is less studied as a particular type of cancer. Much attention has been paid to breast cancer, prostate cancer, and colorectal cancer patients, but many fewer online resources and related studies are available for OvCa patients and their caregivers. (Nagler et al., 2010; Hong et al., 2012; van Eenbergen et al., 2017).

OvCa patients are chosen for two reasons. Firstly, there is a large population of women living with OvCa. The five-year relative survival rate from 2009 to 2015 in the US is 47.6%, with an estimation of 229,875 women living with ovarian cancer in 2016 (NCI, 2018). But at the same time, managing this cancer can be very stressful because of its complexity and the

disease progression (Hagan & Donovan, 2013). Due to the limitation in early detection, OvCa can be developed as an aggressive and fatal disease without proper and timely healthcare. In the United States, it is the most common cause of death compared to other gynecological malignancies (Siegel et al., 2020). However, understanding and support for OvCa patients' particular information needs are extremely limited, though research has already found that cancer patients' information seeking in terms of information needs and source use are very different across cancer types (Nagler et al., 2010). It is important to know how this group of cancer survivors behaves as proactive information seekers and receives help through OHIS.

Secondly, evidence suggests that many OvCa survivors are elderly patients. Among US females of all ages, OvCa caused 14,193 deaths in 2017 (Siegel et al., 2020). Most cases happened in women aged 40 to 59 years (2,748) and 70 to 79 years (7,741) old. For these elderly patients, the current understanding of online HISB may not be applicable due to their unique characteristics and barriers (Oh, 2016). Additionally, there are on average 12 concurrent symptoms for OvCa patients with treatment (Hagan & Donovan, 2013), thus making their information needs complicated and persistent.

Consequently, laypeople and OvCa patients are targeted as the research population in this dissertation.

### 1.3 Research Motivation

As mentioned in the above problem statements, while an increasing population of health consumers relies on online health information seeking to resolve their health-related concerns, and it is known that in domain-specific information seeking, knowledge learning plays an important role as both a by-product of the search process and an influential factor in determining the search performance, there is little existing work that combines SAL and health information seeking, to the best of my knowledge.

Therefore, this dissertation is *first motivated* by the gap in the literature and the lack of connection between the two search fields, OHIS and SAL. Untangling health consumers' OHIS behavior from a cognitive learning perspective will connect and contribute to the two

research communities.

Additionally, I am interested in how well the current HIS web environment can support health consumers in terms of learning. My *second motivation* is to investigate the health consumers' learning outcomes in OHIS.

In the Anderson and Krathwohl's taxonomy (A&K's taxonomy) (L. Anderson et al., 2001), the authors identified six levels of learning objectives from lowest to highest order of thinking skills, based on Bloom's taxonomy (Bloom et al., 1956). It is used in SAL studies for assessing learning outcomes (M. J. Wilson & Wilson, 2013; Collins-Thompson et al., 2016; Ghosh, Rath, & Shah, 2018). The details of the A&K's taxonomy and its applications in the previous information science studies will be presented in the related work (see Section 2.2). In this dissertation, how well consumers can learn (i.e., learning outcomes) against the cognitive levels in A&K's is assessed.

Last but not least, the uniqueness of health-consumer groups inspires the *third motivation*. There is a desire to explore how and how well different groups of consumers would search and learn.

The ultimate goal of this dissertation is to alleviate the difficulties and provide the demanding facilities that would result in a better SAL experience in the design of the domain-specific search system.

In most HIS tasks, health consumers' domain knowledge is insufficient, so they are more likely to involve themselves in the SAL process. A search system specifically supporting learning in the health-related domain is needed. Given the insufficient support for learning in the current search system, my ultimate goal in this dissertation is to provide design implications that enhance search behaviors and learning outcomes.

#### 1.4 Significance and Contribution

As one of the first to investigate health consumers' SAL behavior and performance in today's web environment, this dissertation study is expected to make contributions of significance to HIS theory, HIS and SAL research work, and healthcare practice.



**Significance to HIS theory:** Though learning outcomes are proven to be cognitive outcomes of consumers' health information seeking in some previous empirical studies (P. Zhang & Soergel, 2016; Y. Chen, Zhao, & Wang, 2020) and learning happens during the HIS process, there is no existing HIS theory or model that recognizes the learning component. In this dissertation, through connecting the theories and models in HIS and SAL, a conceptual framework that unpacks the relationship between HIS and SAL is proposed. Built upon the conceptual model, the investigation of laypeople and OvCa patients will develop a theoretical understanding of HIS from a cognitive learning perspective, thus contributing to the HIS theory.

**Significance to HIS and SAL research areas:** The work is also expected to link the two research areas and fill the gap in the literature. While an increasing population relies on online information seeking to solve their health-related questions, it is known that in this type of domain-specific information seeking, there is little existing work that combines learning and health information seeking. While simply investigating HIS as a learning process might limit the research scope and targeted population, it provides opportunities to understand and model the knowledge change in HIS at a fine-grained level and satisfy exploratory health-related information needs. Additionally, the idea of enhancing learning has implications for designing better systems to support the increasingly popular online health information seeking.

**Significance to healthcare practice:** The Internet has introduced significant changes in health consumers' information seeking, thus carrying substantial influence on today's healthcare practices. Free access to online health information has the potential to address health disparities (Gibbons et al., 2011; Lee, 2009). Having knowledge is also a critical attribute of patients' empowerment (Groen et al., 2015). Therefore, many healthcare services and providers are also taking advantage of the consumer-centered OHIS and other new technologies to empower consumers with a wealth of information, knowledge, and even self-care abilities. The findings of this study are expected to provide implications, guidance, and insights to healthcare providers, practitioners, researchers, and developers to design personalized health information systems that will enhance learning and satisfy the unmet needs of health consumers. Additionally, it is expected that some of the findings and design

implications from laypeople and OvCa patients can be generalized to other consumer groups.

## **1.5 Overview of the Chapter Structure**

The dissertation is structured as follows. First, the related theories and models and empirical studies in the two areas, OHIS and SAL, are reviewed in Chapter 2. Based on that, the conceptual framework and research questions are proposed in Chapter 3. Subsequently, Chapter 4 presents the overview of the research design and data collection methods consisting of four sub-studies, which are elaborated separately in Chapter 5, Chapter 6. Chapter 7, and Chapter 8. Then, the implications based on the major results and findings are discussed in Chapter 9. Finally, Chapter 10 concludes the dissertation.

## 2.0 Literature Review

In this chapter, literature from two research areas is first reviewed in two sections: 2.1 health information seeking behavior (HISB) and 2.2 search as learning (SAL). Then the intersection between HISB and SAL is summarized in section 2.3 health consumers' search as learning in OHIS.

In the review of the two research areas (HISB & SAL), the key concepts are first defined, such as health information seeking behavior and search as learning. Secondly, since one of the goals of this dissertation is to connect the literature of the two areas, the representative theories and models in each area are reviewed, respectively. The review in this section lays the groundwork for the conceptual framework of this dissertation. Thirdly, prior empirical studies in HISB and SAL are summarized to demonstrate how they relate to and inspire this dissertation.

### 2.1 Health Information Seeking Behavior

#### 2.1.1 Concepts

Human Information Behavior is defined as "the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use." (T. D. Wilson, 2000, p. 49). Information seeking behavior, one component of the concept of information behavior, is a crucial behavior of interest among the researchers. Åström (2007) conducted a co-citation analysis of the articles published in 1990-2004 in 21 Library and Information Science (LIS). Results suggest that information seeking and retrieval is a dominating research field. Though no agreed-upon definition for the concept of information seeking behavior (ISB), this dissertation is grounded on two widely recognized definitions: ISB is "a process in which humans purposefully engage in order to change their state of knowledge" (Marchionini, 1997, p. 5-6), and it is "the purposive

seeking for information as a consequence of a need to satisfy some goals” (T. D. Wilson, 2000, p. 49). In T. D. Wilson (1999)’s nested model of human information behavior, information search behavior is further defined as a subset of information seeking behavior, ”particularly concerned with the interactions between information user (with or without an intermediary) and computer-based information systems, of which information retrieval systems for textual data may be seen as one type.”

Health Information Seeking Behavior (HISB, henceforth) is a domain-specific ISB with health-related information as the domain of interest. HISB is defined as the search for and receipt of messages that help “to reduce uncertainty regarding health status” and “construct a social and personal (cognitive) sense of health” (Tardy & Hale, 1998, p. 338). It consists of a number of information seeking “actions used” by health consumers “to obtain knowledge of a specific event or situation” (Barsevick & Johnson, 1990).

Nowadays, due to an explosion in the amount of online health information, more and more people are involved in online health information seeking behavior(OHISB) (Fox, 2013). Unlike general-purpose online information seeking, OHIS has two unique characteristics. Firstly, understanding and interpreting health-related information require certain domain-specific knowledge, particularly when encountering resources that are full of unfamiliar, domain-specific terminologies (White et al., 2009). Secondly, when gathering health information, people prioritize information quality, since they might eventually act or make decisions based on such information.

As a result, health information seekers quite commonly navigate through a variety of online health information resources and gradually learn and gain health-related knowledge before locating preferred ones.

### **2.1.2 Theories and Models**

In this section, four previous theories and models in HISB are reviewed: 1) Lenz (1984)’s information seeking model; 2) Freimuth, Stein, and Kean (1989)’s health information acquisition model; 3) Johnson (1997, 2003)’s comprehensive model of information seeking, and 4) Longo (2005)’s integrated model of health information-seeking behaviors. One commonal-

ity shared in these four models is that they all regard HISB as a process and simplify the complex HISB into a series of stages and discuss the factors that affect each stage (Lambert & Loiselle, 2007). In general, these models characterize HISB by antecedents, actions, and outcomes.

Lenz (1984) and Freimuth et al. (1989) both depict health consumers' HISB as a decision-making flowchart, which is triggered by information seeking stimulus and ended with the decision on whether the information collected is adequate to stop or continue the HISB. In between, each model describes four steps the consumer needs to take and make decisions before they can proceed to the next step. Though these two descriptive models are very comprehensive, they fail to reveal any patient-related or disease-related factors associated with the process.

On the contrary, Johnson (1997, 2003) 's Comprehensive Model of Information Seeking (CMIS) does not elaborate much on the information needs (i.e., stimulus). Instead, CMIS highlights the importance of sources and defines health information seeking as the "purposeful acquisition of information from selected information carriers" (p. 4). Specifically, CMIS depicts the process of patients' health information seeking by defining three primary schemes: antecedents, information carrier factors, and information seeking actions. The antecedents refer to the underlying determinants of HIS, including demographics, direct experience with the problem, salience of information, and personal beliefs. The second schema is the information carrier factors consisting of characteristics of the information source and its utilities, which further shape the third schema, the patients' information seeking actions. Longo (2005) 's integrated model of health information-seeking behaviors underlines possible outcomes of either active or passive HISB.

Through a comprehensive concept analysis and review including but not limited to the aforementioned models, Lambert and Loiselle (2007) summarized and defined three key components of the HISB, namely, antecedents, characteristics of HISB, and consequences. Firstly, like general-purpose ISB, HISB is also triggered by information needs to fill a certain knowledge gap. Besides, many other antecedents further determine the follow-up actual HISB, including personal factors (e.g., sociodemographic characteristics, psychosocial variables), situational or contextual factors (e.g., individuals' environment, source of information, and

information seeking context). Secondly, characteristics of HISB can mainly be grouped into two dimensions: method dimension (i.e., search behaviors and information source) and information dimension (i.e., type and amount of information sought). The authors criticized that Lenz (1984) and Freimuth et al. (1989) covered both dimensions while Johnson (1997, 2003) and Longo (2005) only elaborated on the method dimension. Thirdly, Lambert and Loiselle (2007) expanded Longo (2005)'s HISB outcomes and categorized four overarching types of outcomes or consequences of HISB, which are cognitive outcomes (e.g., increased knowledge), behavioral outcomes (e.g., increased self-care abilities), physical outcomes (e.g., increased physical quality of life), and affective outcomes (increase hope).

### **2.1.3 Empirical Studies**

Guided by Lambert and Loiselle (2007), in this section, HISB empirical studies are grouped into two dimensions: section 2.1.3.1 method dimension, i.e., how consumers search and behave in OHIS and what online sources they select; and section 2.1.3.2 information dimension, i.e., what type and how much online information health consumers seek.

**2.1.3.1 Method Dimension** The method dimension of HISB mainly refers to the consumers' information sources and search behaviors.

Existing work on online health information sources has largely been dedicated to controlling or assessing the quality of information from different online sources, such as establishing rating systems for evaluating websites or webpages (Eysenbach, Powell, Kuss, & Sa, 2002; Silberg, Lundberg, & Musacchio, 1997; Stvilia, Mon, & Yi, 2009; Tao, LeRouge, Smith, & De Leo, 2017; Y. Zhang, Sun, & Xie, 2015). These studies usually aim to educate content providers as well as health consumers by defining the best practices of high-quality health content. Best practices usually consist of criteria related to the content (e.g., accuracy, readability, completeness) or the presentation style (e.g., layout, aesthetics, appearance) of health information. With the rise of the research interest in patient-centered OHIS, gradually, the trend of examining the credibility and quality of online health information sources from the consumers' perspective has naturally emerged (e.g., White et al., 2009; Y. Zhang, 2014).

Credibility assessment is not only studied in the field of information science. Psychology researchers devote a good deal of effort to decompose credibility assessments. Through empirical studies, Flanagin, Metzger, and Hartsell (2010) found that information type is a critical factor affecting credibility assessment. In the search for health and medical information, individuals tend to put more effort into checking credibility compared with the search for entertainment information. Information seekers utilize a two-level heuristic to judge the credibility of online sources, in which they assess the structure or website level credibility and also the content or webpage level credibility (Metzger & Flanagin, 2015; Sundar, 2008). The influences of the source characteristics on these two levels can interact with each other. Perceived credibility is also found to vary by genre of an online source. For example, participants in Flanagin and Metzger (2007)'s study believed that news organization websites are the most credible, followed by e-commerce sites, special interest sites, and personal websites.

In the context of OHIS, aligning with the definition of credibility from psychology researchers (Metzger & Flanagin, 2015), studies view credibility assessment and source selection as manifests of consumers' information evaluation experience. It has been found that consumers' credibility judgment is influenced by many source characteristics, such as the social context of online communication, the technical languages used, and authors' credentials (Thon & Jucks, 2017; Zimmermann & Jucks, 2018).

In terms of research method and design, most existing studies about source selection mainly adopt survey-based methods and suffer from similar limitations (Anker, Reinhart, & Feeley, 2011; Y. Zhang, 2014). Firstly, cross-sectional surveys collecting data about participants' source selection at one particular time usually neglect the search contexts, associated search processes, and search outcomes (Anker et al., 2011). Secondly, most prior studies constrain participants to select from one or several predefined online sources (e.g., Zimmermann & Jucks, 2018), or regard the Internet as one general health information source compared to other sources such as doctors or libraries (Case, Johnson, Andrews, Allard, & Kelly, 2004; Koch-Weser, Bradshaw, Gualtieri, & Gallagher, 2010).

**2.1.3.2 Information Dimension** The information dimension of HISB refers to the characteristic of the information sought through online HISB, especially the type and amount of

the information (Lambert & Loisel, 2007). The research found that consumers may behave variously regarding what type of online information and how much information they prefer in online HISB.

Comparing with the traditional paternalistic model in which "patients expected to be passive recipients of healthcare information from doctors" (Xie, Wang, Feldman, & Zhou, 2010), the Internet enables individuals searching for health information online to expand understandings and explanations regarding health (Ayers & Kronenfeld, 2007), as well as in making health-related decisions (Broom, 2005; Seçkin, 2010). Through log analysis, Cartright, White, and Horvitz (2011) suggested that there are two phases in exploratory online HIS with the pursuit of two different types of information: evidence-based and hypothesis-directed. The evidence-based phase mainly focuses on locating information regarding signs and symptoms, while the hypothesis-directed phase involves discrimination among different diseases under uncertainty.

Regarding the preference of the adequacy of the information acquired online, not all people with health concerns are active information seekers. Despite many people actively go online and other sources to seek health-related information on many aspects of their conditions, there also exist individuals who only want to collect a certain type of health-related information (Rutten, Squiers, & Hesse, 2006).

A better understanding of health consumers' preferences in the information dimension contributes to characterizing different types of online health information seekers, which further can imply and guide the design of better tools to serve and support seekers. Pang et al. (2014); Pang, Verspoor, Pearce, and Chang (2015)'s prior work is one of the research that synthesized health consumers' characteristics in alignment with OHIS system design. They classified health consumers into four different characteristic groups: knowledge digger, focused reader, all-around skimmer, and quick fact seeker, and proposed six design principles in support of their OHIS. Consumers are grouped based on Reading Engagement (i.e., duration of reading) and Research Tactics (i.e., motivation for digging out in-depth information). According to the definition, knowledge diggers are the consumers who demonstrate high reading engagement and research tactics. They are eager to dig much as well as in-depth health-related information. Differently, quick fact seekers can be quickly satisfied by locating



the information needed. The work confirmed that health consumers are involved in OHIS with different levels of learning objectives, and they should be supported with different design features. Furthermore, this dissertation expects to reveal, more specifically, what types of and how much information is needed to achieve different levels of learning objectives in OHIS.

## 2.2 Search as Learning

### 2.2.1 Concepts

Based on the literature exploring the interaction between search and learning, Rieh, Collins-Thompson, Hansen, and Lee (2016) identified two roles of learning in a searching process – *learning to search* and *searching to learn*, where the former refers to how people learn search experience and expertise while the latter regards learning as a parallel process with the searching process and ends with relevant knowledge increased.

The focus of this dissertation is laid upon *searching to learn* paradigm. Particularly, it is believed that during the health information seeking process, besides the explicit searching behaviors, consumers are also involved in an implicit knowledge learning process. Beyond a simple look-up search, the information needs in HIS are usually exploratory in nature, which triggers learning, and involves complex and diverse interactions with the systems. The searching and learning processes are intertwined with each other at different stages of the information seeking process. Therefore, this current literature review of search as learning domain concentrates on searching to learn and excludes papers on learning to search.

### 2.2.2 Theories and Models

Many of the studies in search as learning are founded on combining theories and models originated from two most representative fields that contribute to theorizing searching and learning: information science (IS) and educational psychology (EP). Though search as learning is recognized as a separate research agenda in recent years (Rieh et al., 2016), it

has long been aware, in the domain of information science (IS), that knowledge acquisition is an important component in the information seeking process. Several information seeking theories conceptually assume that an individual's knowledge state changes during the search process. As stated in the ASK (Anomalous State of Knowledge) model, Belkin (1980) argued that information seeking is a process to resolve the anomaly between users' current states of knowledge the problem they faced. Marchionini (1997) also claimed that beyond simple lookup search, people often engage in exploratory search tasks where learning and investigation could play essential roles.

Generally, highly adopted IS theories in the search as learning studies demonstrate two notions. Firstly, researchers employ the theories viewing information seeking as a process with different stages, such as Marchionini (2006)'s information seeking process model and Kuhlthau (1991)'s information search process (ISP) model. These theories help the researchers to conceptualize and investigate when and how learning occurs during the information seeking process (Vakkari, 2016). Secondly, theories that support the idea that searching is to acquire knowledge to satisfy the information needs are also highly cited. These theories mainly include Information Foraging Theory (Pirolli & Card, 1999), Belkin's Anomalous State of Knowledge (ASK) Model (Belkin, 1980), and Dervin's Sensemaking Theory (Dervin, 1983).

On the other hand, among many theories and models developed in educational psychology, the most highly adopted ones in search as learning studied can be grouped into two overarching branches: Anderson and Krathwohl's taxonomy of learning (A&K's taxonomy) (L. Anderson et al., 2001) and Piaget's cognitive development theory (Piaget, 1976). As shown in Figure 1, A&K's taxonomy of learning defined six levels of depth of learning with different cognitive complexity in a hierarchical structure. It was first introduced in 1956 by Bloom (Bloom et al., 1956) and later revised by Anderson and Krathwohl. It is widely adopted to instruct the creation of learning objectives. The theory assumes that in order to achieve a higher level of learning objective, one must master the lower levels. For example, if an individual wants to *apply* the knowledge learned to a new situation, it is required for her to *remember* and *understand* the knowledge beforehand. M. J. Wilson and Wilson (2013) were the first to use A&K's taxonomy to measure the depth of learning in the IS field. In this

dissertation, this hierarchical structure is borrowed to propose the conceptual framework. Detailed elaboration is in the section 3.1.

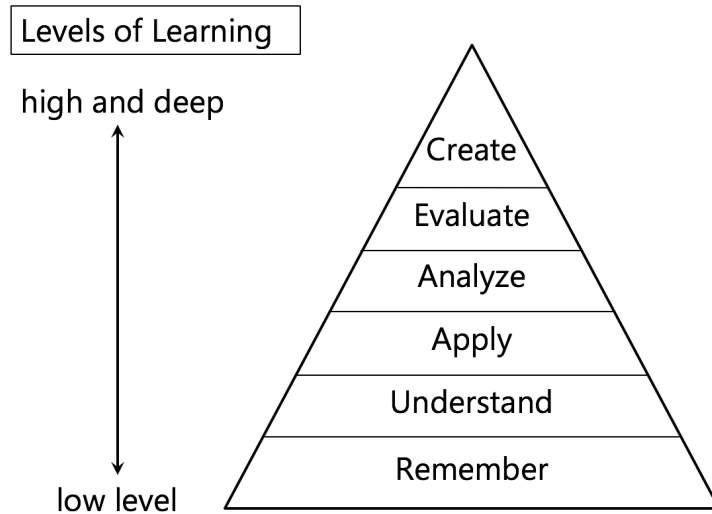


Figure 1: Anderson and Krathwohl's Taxonomy of Learning

Piaget's cognitive development theory is another popular theory addressing the cognitive learning process. Piaget (1976) identified two types of cognitive/conceptual changes in knowledge acquisition: (1) assimilation, the addition of information to existing knowledge structures; and (2) accommodation, the modification or change of existing knowledge structures. Construction-Integration Model (Kintsch & Walter Kintsch, 1998) and Schema Theory (Rumelhart & Norman, 1976) were proposed following Piaget. Schema theory extended and proposed three types of conceptual changes in ones' knowledge structure: 1) accretion, adding factual information without changing one's knowledge structure; 2) tuning, adding new knowledge knots and weakly change one's knowledge structure; and 3) Restructuring, radically changing one's knowledge structure.

The combination of the theories from different fields provides possibilities to understand and model the HISB process from a learning and knowledge acquisition perspective. Drawing upon theories in IS and educational psychology, the HISB research could better understand the health information seeking tasks that are commonly exploratory in nature. Besides,

though learning outcome is used as an indicator of the cognitive outcome of an individual's health information seeking in some prior studies (Lambert & Loiselle, 2007), it is mainly assessed through self-reporting (Suri, Majid, Chang, & Foo, 2016). Therefore, there lacks an implicit measurement of the gradually changed knowledge states. Viewing HISB as a learning process and adopting the measures developed in the search as learning studies could allow researchers to model the HIS outcomes more dynamically and accurately, thus providing the needed information at the appropriate level of knowledge.

### **2.2.3 Empirical Studies**

The review of related theories and models has shown that searching and learning are intertwined with each other at different stages of the information seeking process. In this section, the empirical studies on two stages are reviewed: Section 2.2.3.1 learning during the information seeking process; and Section 2.2.3.2 knowledge as an outcome of HIS. The review will focus on the linkages between the two fields: search as learning and health information seeking and will further demonstrate the rationale for investigating “search as learning” in the context of online HIS.

**2.2.3.1 Learning During the Information Seeking Process** Beyond addressing health information needs by the end of a HIS task, the HIS process also involves learning or knowledge acquisition. During the process of interacting with the search engine, examining search results, and extracting useful information, users acquire novel knowledge at any time, as evidence shown in the prior studies.

In information seeking behavior literature, learning has been primarily situated around the concept of knowledge acquisition. Learning in searching can occur in a relatively long-duration search, e.g., a seminar for preparing a proposal (Vakkari et al., 2003) or a course lasting for several months (Wildemuth, 2004). Furthermore, recent studies prove that domain knowledge learning takes place even in a short-term search session. Eickhoff et al. (2014) defined two types of search sessions in which users explicitly search for new knowledge, i.e., procedural and declarative sessions. Based on this, they discovered that users learn

through the search process, not only from the final results. Similarly, J. Liu, Belkin, Zhang, and Yuan (2013) reported that users' perceived domain knowledge grows significantly, especially at the beginning of the search, and later on, it may reach a plateau, which called a "ceiling effect." Chi et al. (2016) demonstrated that knowledge learning happened, and users' knowledge kept growing either when they searched individually or collaboratively, but they issued significantly more diverse queries in the collaborative condition. In a lab-based user study, Collins-Thompson et al. (2016) found that users achieved combined factual and conceptual knowledge gains through the process of searching the tasks assigned to them. Y. Chen et al. (2020) conducted one of the few previous studies that have probed into the fine-grained knowledge change within online health information seeking context. Through a user study with pre-designed health-related search tasks, the authors recruited 30 undergraduate students and explored how they learn during the OHIS process. The learning activity is quantified by three types of conceptual changes in the search queries: accretion, tuning, and restructuring (P. Zhang & Soergel, 2016).

Though a few previous studies have probed into understanding and modeling the fine-grained knowledge change within a search process, or even a single session, there are only a few studies, if any, investigating how learning occurs in the health information seeking context to the best of my knowledge. Besides, the answers to some critical research questions, e.g., what individual factors would affect learning in the HIS, what sources would be adopted, remain unexplored.

**2.2.3.2 Knowledge Change as an Outcome of HIS** Learning outcome (i.e., knowledge change) is perceived as a cognitive outcome of an individual's health information seeking (Lambert & Loiselle, 2007). Suri et al. (2016) surveyed 1062 college-going adults to assess how health literacy affects HIS behavior. They defined three types of outcomes: learning outcomes, communication outcomes, and instrumental outcomes. Learning outcomes were measured as self-reported gains in knowledge about a particular health-related topic. Statistical analysis showed learning outcomes were associated with different domain-specific health literacy skills: Find Health Information (FHI), Appraise Health Information (AHI), Actively Manage One's Health (AMH), and E-health literacy (e-HEALS).

On the contrary, users' knowledge states can be implicitly and unobtrusively reflected by the changes of users' search behaviors, e.g., query complexity, click complexity, display time, query reformulation, etc (e.g., Y. Zhang et al., 2015; Eickhoff et al., 2014; Chi et al., 2016; Bhattacharya & Gwizdka, 2019; Y. Chen et al., 2020). P. Zhang and Soergel (2016) proposed a model which could predict a user's knowledge level from implicit search behaviors. The search context employed in their study is health-related: knowledge of MeSH (Medical Subject Headings) terms. Similarly, Yu et al. (2018) proposed a supervised model to predict users' knowledge gain and state based on search behaviors in informational search sessions. The model suggests that among the 70 features of query, SERP, browsing, and mouse activities, the content of queries and browsed documents are the features most important for predicting knowledge gain and knowledge state. However, this model is not used specifically in the HIS context.

Beyond using learning outcomes as a measure of search performance, it is also believed that there is great potential to design better search systems as well as algorithms to enhance the learning outcomes by the end of the HIS process (P. Hansen & Rieh, 2016).

With the purpose of investigating how students learn about programming during online searching, Y. Lu and Hsiao (2017) conducted two classroom studies. The first study revealed differences between novice and advanced learners in terms of information seeking patterns. Novice learners tend to spend more time browsing the results and reading documents, while advanced students could achieve more in a shorter time. Applying what they found in the first study to designing an assistant system PiSA, they conducted a second classroom study to test if novices could get help from the system, which embedded term suggestions, results summarization, and other facilitates. Results showed that the system indeed could foster novices' learning. Pang et al. (2016) designed a health-specific website to support different HIS behaviors. They designed and assessed the system through a laboratory-based observational study. Participants reported that they obtained more knowledge with the help of design considerations, such as supporting serendipity and enhancing interactivity.

## 2.3 Health Consumers' Search as Learning in OHIS

### 2.3.1 Domain Knowledge's Influence on Consumers' OHIS

Domain knowledge, one's knowledge about the search topic, is recognized and examined as a critical implicit factor that affects the search behavior as well as the search performance since the early 2000s (Hölscher & Strube, 2000; Jenkins et al., 2003). Efforts have been devoted to revealing the differences between the individuals with more domain-specific knowledge (i.e., domain expertise) and others who are less familiar with the domain (i.e., domain novice) in terms of their search strategies, search behaviors, and search performances.

HIS is a typical type of domain-specific information seeking (Hanbury, 2012) because of the nature of the health-related information tasks. As a result, health-related topics, including healthcare, medicine, are commonly selected as a domain to distinguish the expertise from the novice. For example, Bhavnani (2002) identified healthcare and e-commerce as two domains to study the search strategies of users with different prior knowledge. Through a controlled lab study, he found that experts in a domain can easily solve the task because they already knew which website was better adapted to fill their needs, like MedlinePlus, while laypeople usually started with Google or other general-purpose search engines. Bhavnani (2002) also reported that better domain-specific search engines should be designed to make the search knowledge in these new domains explicit and available. Similarly, White et al. (2009) selected medicine as a domain to reveal how domain experts behave differently based on large-scale log analysis. The other three domains they selected were finance, law, and computer science. In the paper, they clarified their reason for identifying these four domains was because that the information seekers with topical interest were obviously different in knowledge, including both the general population of less or no knowledge and professional groups. The findings showed that experts have a higher success rate in their domain of expertise.

The research examining the influence of domain knowledge, however, usually views an individual's domain knowledge as a relatively stable state rather than a changing value with the process of online information seeking.

### **2.3.2 Health Condition's Influence on Consumers' OHIS**

It has long been accepted that the nature of a search task is an important factor that may impact information seekers' search behaviors, e.g., source selection behavior. Previous work has explored the influence of many different task characteristics, such as task complexity, the urgency of the task, and the domain of the task (Kelly, Arguello, Edwards, & Wu, 2015; Chi et al., 2016; Wildemuth, Kelly, Boettcher, Moore, & Dimitrova, 2018). Focusing on the domain of OHIS, Y. Zhang et al. (2017)'s study indicates that consumers' source selection varies by their individual characteristics and the type of task (i.e., factual, exploratory, and personal experience task). Health consumers with higher health literacy tend to select a search engine in all three tasks but are more likely to consult social Q&A sites for personal experience tasks and make more use of online health communities in the exploratory and personal experience tasks. While in the factual tasks, consumers with better health status prefer SNS.

Previous work has also explored whether the severity of the health condition affects users' OHIS behaviors. Drawing upon a large-scale analysis and a survey, De Choudhury et al. (2014) examined users' seeking and sharing behaviors through search engines and social media, the two most widely-used online sources, respectively. They found that users' source preferences are associated with the severity of the health conditions: when searching for severe medical conditions (e.g., cancers), users prefer search engines. On the contrary, the participants tend to use SNSs when sharing benign health conditions and symptoms (e.g., headache).

These insightful studies inspire the current study to explore online health consumers' search as learning behaviors under health conditions with different severities.

### **2.3.3 Laypeople's Search as Learning in OHIS**

Different from the cancer survivors who have been living with the condition for a period of time or even for years, laypeople in this study are the consumers who are new to their health concerns.

In this section, the related work grounded on the intersection of the aforementioned two



broader research areas: health information seeking behavior (HISB) and search as learning (SAL), is reviewed, focusing on laypeople. Specifically, guided by the two strands of HISB, the review is divided into two dimensions: method dimension: Section 2.3.3.1 how laypeople search and learn, and information dimension: Section 2.3.3.2 what laypeople search and learn.

**2.3.3.1 Laypeople’s OHIS: How They Search and Learn** Most laypeople rely on search engines to initiate OHIS, though they are also aware that the lack of knowledge makes it difficult to accurately describe the health issue and locate the pertinent information (Pang et al., 2014; Yilma, Inthiran, Reidpath, & Orimaye, 2017; Eysenbach & Köhler, 2002; Fox, 2013; D. L. Hansen et al., 2003). For example, in D. L. Hansen et al. (2003), 87% of the participants’ visited sites directly come from search results or search engine’s recommendations.

There are generally two concerns regarding this issue. Firstly, it is questionable whether or not the laypeople are capable of issuing correct or accurate queries to describe their health information needs because such action often requires a certain level of health-related background knowledge. It is revealed that laypeople’s queries suffer problems such as misspellings (D. L. Hansen et al., 2003) and too general (Eysenbach & Köhler, 2002), resulting in search failure. Secondly, some laypeople may be unaware that the ranking positions can be biased by marketing, especially when searching for the health information that may lead to purchase (Kim et al., 2011). For example, if one is searching for “weight loss”, she may be directed to advertisements or e-commerce websites via the top-ranked results. It is important to highlight the search engines’ ranking and scoring mechanisms in the information literacy programs designed for health consumers.

In a user study that compares laypeople’s and health professionals’ search behavior and performance when searching for medical-related informational tasks, the results revealed that laypeople issued significantly fewer queries, fewer search operator usages, fewer terms in a query (Kharazmi, Karimi, Scholer, & Clark, 2014). However, laypeople and health professionals were all report as a whole group, without an examination of how individuals’ search and learn evolve during the process.

Built upon the exploratory search theory, Pang et al. (2015) outlined three stages in laypeople's OHIS process. In stage 1, a layperson with health concerns starts the search by submitting a search query to a search engine. With pretty limited prior knowledge, she examines the results yielded by the search engine in an exploratory fashion. In stage 2, she gradually makes sense and learns about the related knowledge through reading and digesting online information. Then, in stage 3, as exposed to other related but unknown topics, the layperson engages in an exploratory search again. Obviously, the boundaries between the stages could be fuzzy, and the order of the stages is not rigid.

To support laypeople's exploration of online health information, Pang et al. (2015) designed a web app that incorporates a set of features that facilitate laypeople's search process. Pang et al. (2015, 2014, 2016)'s work is inspiring because laypeople are one of the populations of interest in this study, and providing design implications is also one objective of this dissertation. However, the current dissertation aims to take apart laypeople's OHIS from a cognitive learning perspective, with a focus on their behaviors, barriers, and outcomes associated with learning.

**2.3.3.2 Laypeople's OHIS: What They Search and Learn** Two dominating types of information that laypeople search online are illness-related information and wellness-related information (Yilma et al., 2017; Weaver III et al., 2010; Ogan, Ozakca, & Groshek, 2008).

Illness-related information is usually sought when laypeople encounter uncertain health issues that motivate them to look for symptoms, disease, or treatment. Pang et al. (2014) interviewed laypeople and revealed that self-diagnosis is a common information need of laypeople's OHIS. The respondents in the interview reported that, in most cases, encountering health issues drove them to conduct OHIS. This result is consistent with a US nationwide survey in 2013, which revealed that one in three adults had used the Internet to diagnose themselves or someone else (Fox, 2013). Respondents in Pang et al. (2014) further confirmed that due to a lack of health-related knowledge about their health issues, they normally try different queries, go through an exploratory search process, thus gradually learn about new keywords and new knowledge concurrently with exploring and digging around the issue.

However, it has long been concerned that laypeople's self-diagnosis using online resources could potentially delay the necessary medical intervention if the consumers do not follow up with a doctor visit. In addition, if misleading or inaccurate online information were found and believed during the customers' own OHIS, it would later be harmful to the patient-health professional relationship (Ahmad, Hudak, Bercovitz, Hollenberg, & Levinson, 2006; McMullan, 2006; Robinson, Patrick, Eng, Gustafson, et al., 1998).

On the other hand, aside from people with health concerns, there is a substantial body of evidence suggesting that healthy people, who are in pursuit of promoting a healthy lifestyle, are taking a growing proportion of online health consumers. Wellness-related information mainly includes diet, exercise, or fitness information (Weaver III et al., 2010; Jong & Drummond, 2016). Respondents in Yilma et al. (2017)'s survey reported that fitness, diet, and skin health were their most frequently searched wellness information on the Internet. Note that all the respondents in this survey are identified as laypeople in the university of a developing country.

Obviously, previous work has recognized that laypeople are involved in learning for different health-related topics in OHIS. However, most existing studies adopt cross-sectional and self-report data to present an overall picture of laypeople's information needs, and limited research has endeavored to recognize that even laypeople can engage in OHIS tasks from the simple to complex. However, there is no clarity about what type of information is needed to satisfy OHIS learning goals with different complexity. Yet it remains unclear how well laypeople can learn from different types of health information.

#### **2.3.4 Ovarian Cancer Patients' Search as Learning in OHIS**

An observational study conducted by Zeng et al. (2004) indicated that specific and complicated health information retrieval goals would motivate patients towards OHIS in advance. Bundorf, Wagner, Singer, and Baker (2006) surveyed more than twelve thousand individuals and reported that consumers who are "uninsured with a chronic condition" tend to seek health information online. These studies suggest OHIS is prevalent among cancer survivors, this particular group of health consumers who are very likely to encounter specific

as well as complicated health information retrieval goals, and many of whom are "uninsured with a chronic condition." As indicated by some SAL literature, a search and learning process is expected to happen with the complex information needs (Rieh et al., 2016).

The reviews of two dimensions are presented in this section: method dimension: 2.3.4.1 how OvCa patients search and learn, and information dimension 2.3.4.2 what OvCa patients search and learn.

**2.3.4.1 OvCa Patients' OHIS: How They Search and Learn** In a study characterizing OvCa patients' experience from pre-diagnosis to treatment (Pozzar & Berry, 2019), women with OvCa reported that they are eager to learn and engage in the treatment plan with the providers. Though patients value their communication with the health care providers very much, the fact is that the providers are not always accessible. Therefore, the patients would turn to the Internet and conduct online seeking for more information as well as emotional support (Pozzar & Berry, 2019; Mayer et al., 2007). Furthermore, the Internet is found to be cancer patients' most preferred information source over other traditional information sources (e.g., health care providers, libraries, friends and families, etc.) in many previous nationwide surveys (Mayer et al., 2007; Chou, Liu, Post, & Hesse, 2011; Hartoonian, Ormseth, Hanson, Bantum, & Owen, 2014). Data from the Pew research center suggests that 51% US adults with a chronic disease have looked for information on the Internet (Fox & Purcell, 2010).

Instead of regarding the Internet as one collective information source, it is crucial to aware that there are many different types of online sources that provide cancer patients with information of varying contents, formats, and qualities. Among various sources, the online health community (OHC), also known as the online support group, where the cancer community can seek, share, and exchange information with peers is a critical online source. Participating in OHC is one of the three primary approaches for consumers to seek online health information, and the other two are searching directly with the search engine and interacting with health professions (Cline & Haynes, 2001). For cancer patients who have a constant and evolving need for information, OHC is particularly important with its around-the-clock availability, anonymity, immediate and time-delayed responses (Cline & Haynes,

2001; Grandinetti, 2000). Participating in OHC facilitates cancer patients to chart the path from being unknowing to knowing (Harkin, Beaver, Dey, & Choong, 2017).

Despite the benefits, the impact of participating in OHC remains inconsistent in literature as it is hard to quantitatively and homogeneously measure the outcomes of OHC (van Eenbergen et al., 2017; Houlihan & Tariman, 2017). In addition, patients participating in OHC may even demonstrate dynamic roles due to varying needs, learning styles, interaction characteristics, expectations, and context, thus making it more complicated to deliver desired support (Huh et al., 2016; D. Yang, Kraut, Smith, Mayfield, & Jurafsky, 2019). N. Liu, Tong, and Chan (2017)'s survey results suggest that one's social identity (e.g., group self-esteem) and personal identity (e.g., perceived disease severity) together characterize one's propensity to seek health information in OHC. In addition, there are also concerns about the quality of the narratives shared by patients in OHC (Bekker et al., 2013; J. Zhang, 2019). Given the conflicting evidence, inconclusive instructions of OHC, and limited research on OvCa patients, there is a dearth of research investigating how OvCa patients are navigating through the online environment and use online sources, such as OHC, to perform the search as learning.

**2.3.4.2 OvCa Patients' OHIS: What They Search and Learn** Pozzar and Berry (2019) reviewed the literature regarding the factors that affect OvCa patients' treatment decision making, and it is suggested that lack of knowledge is a critical barrier reported by the OvCa patients in several studies. They are eager to learn more information about their disease and the treatment (Stewart et al., 2000; Ekwall, Ternstedt, Sorbe, & Graneheim, 2011; C. Anderson et al., 2011). Beyond being able to engage in the decision-making of the treatment, getting more information and knowledge about one's cancer brings in additional benefits. Women with OvCa are more likely to feel confident and empowered, thus more willing to adhere to treatment and health behavior, and eventually get better quality of life (Ekwall et al., 2011; M. Yang & Xu, 2019).

In terms of what they want to learn, there was substantial literature on what particular information needs and health-related topics motivated cancer patients to go through OHIS. The highly mentioned information needs include cancer-specific information, treatment-related

information, prognosis information, emotional support, recovery, medical system information, et al (Rutten, Arora, Bakos, Aziz, & Rowland, 2005; Nagler et al., 2010; Maddock, Lewis, Ahmad, & Sullivan, 2011; Halkett et al., 2010; Ramsey, Corsini, Peters, & Eckert, 2017). It is notable that cancer patients' information needs are dynamic and change over the illness trajectory (Rutten et al., 2005; Mistry, Wilson, Priestman, Damery, & Haque, 2010; Jacobs, Johnson, & Mynatt, 2018). Most existing studies about female cancer patients, however, concentrates on breast cancer, while limited attention has been paid to OvCa patients.

The most common information needs of OvCa patients include the status and nature of cancer, treatment information, and self-care and empowerment issues (Stewart et al., 2000; Pozzar & Berry, 2019). Through content analysis of the posts on Ovarian Cancer National Alliance (OCNA), an ovarian cancer online health community, Madathil, Greenstein, Juang, Neyens, and Gramopadhye (2013) classified OvCa patients and caregivers' information needs into three types, namely ovarian cancer-specific information, treatment-related information, and coping information. The analysis further found that the ovarian cancer-specific information needs are more likely to be expressed in a negative tone comparing with the other two types of needs. Surprisingly, the ones' stage of ovarian cancer does not predict the type of information needs. However, the individual's phase of diagnosis and treatment indeed has an influence on their information needs. Patients and caregivers were often concerned about the treatment-related information (e.g., side effects of procedures) in the after diagnosis and waiting for the treatment phase.

Regarding how much information and how frequently cancer patients search and learn, inconsistent results were found in previous studies. In Stewart et al. (2000), most OvCa patients reported a need for detailed and much information regardless of their phase of diagnosis, i.e., diagnosis, during treatment, or after treatment. Differently, some studies claimed that the time since diagnosis plays an important role (Davis, O'Malley, Bator, Ohman-Strickland, & Hudson, 2020). Patients are more likely to experience an aggressive and rapid information seeking phase after being diagnosed with cancer and especially before the treatment, which results in a deluge of information (Hayes et al., 2008). However, at this stage, due to the stress of being diagnosed, more information does not guarantee more

learning (McEwen, 1998). Conversely, Mistry et al. (2010) demonstrated that demographic factors are more critical in determining ones' information needs as well as the amount of information needed comparing with time since diagnosis. At the same time, collectively, all the patients reported feeling under-informed regardless of demographic factors and time since diagnosis. These studies and inconsistent evidence call for a comprehensive understanding of what and how much OvCa patients want to search and learn through OHIS over the course of the survivorship.

### 3.0 Conceptual Framework and Research Questions

#### 3.1 Conceptual Framework

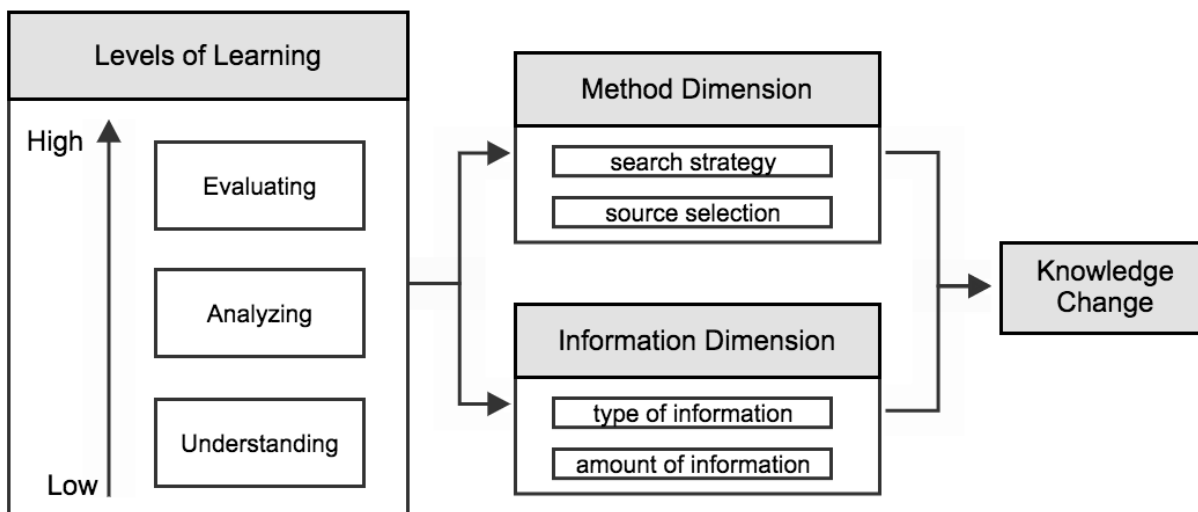


Figure 2: Conceptual Framework of Health Information Search As Learning (HearSAL)

Built upon the theories and empirical studies reviewed in Chapter 2, a conceptual framework of **Health Information Search As Learning (HearSAL)** is proposed, as shown in Figure 2. Because this dissertation work studies health consumers' information seeking as a cognitive learning process in the context of OHIS, the conceptual framework employs theories and models from both OHIS and SAL. Specifically, two theoretical approaches introduced in Chapter 2 are borrowed: L. Anderson et al. (2001)'s taxonomy of learning levels and Lambert and Loiselle (2007)'s two dimensions of HISB characteristics.

In general, as shown in the framework, it is assumed that *being motivated by different levels of learning objectives, health consumers conduct online health information seeking with two dimensions of characteristics (i.e., information and method) and finally achieve knowledge change as an outcome.*



### 3.1.1 Levels of Learning

*Understand, Analyze, and Evaluate* represent three levels of the six learning objectives in A&K's taxonomy, from lower cognitive complexity to higher cognitive complexity (see Figure 1). Understand means the learner can explain and synthesize ideas or concepts beyond simply recalling the concepts. To master the level of Analyze, one can draw connections and differentiate among ideas or concepts. Above that, the Evaluate level requires the learner to have the ability to make decisions, judgments, or critiques. Aside from these three levels, other levels in A&K's taxonomy are Remember, Apply, and Create. Instead of adopting all six levels, three levels are chosen for three reasons. First, it is more feasible and manageable to differentiate between three levels rather than including all six levels. Second, the Remember level is too basic, and Apply, which involves using the information in a new situation, is not very suitable for health-related tasks. Additionally, Create, the highest level in the hierarchy, is out of the scope of health consumers' objectives in OHIS, but more applicable to health professionals. Third, the levels of learning objectives should be close to real OHIS scenarios. As claimed in Cartright et al. (2011)'s search log analysis, there are two representative OHIS intentions: evidence-based and hypothesis-directed. With the evidence-based intention, consumers mainly focus on locating information regarding signs and symptoms, which can be mapped to the Understand level of learning. The hypothesis-directed intention, which drives consumers to draw connections and discriminate among different situations and conditions under uncertainty, aligns with the Analyze level. Besides, the Evaluate level is added, as it corresponds to decision-making intention, which involves seeking information to make a treatment decision.

### 3.1.2 Characteristics of OHIS

Another component of the HearSAL framework is HISB, with its two dimensions. As reviewed in Section 2.1.2, Lambert and Loiselle (2007) conceptualized two dimensions of HISB based on assorted characteristics: method dimension and information dimension. Additionally, it highlights HISB as a process of SAL with increased knowledge as a cognitive outcome of the process. The method dimension of HISB mainly refers to the individuals'

source selection behaviors and search behaviors. The information dimension of HISB refers to the characteristic of the information sought through online HISB, especially the types and amounts of the information.

### 3.1.3 Knowledge Change

In this conceptual framework, aligning with previous work, *knowledge change* (also called "knowledge state change") is defined as the change in one's *knowledge state* over a period of time (M. J. Wilson & Wilson, 2013; Eickhoff et al., 2014; Yu et al., 2018). Here, "knowledge state" refers to an individual's temporal knowledge regarding a specific topic (e.g., health-related knowledge). In particular, it is assumed that consumers' knowledge state evolves throughout OHIS and SAL for different levels of learning objectives, thus resulting in knowledge change. Four aspects of OHIS—including the search behaviors, source selection, type of information, and amount of the information collected—will influence and contribute to the knowledge change. As for the measurement of one's implicit knowledge state and knowledge change, it is believed that the knowledge state can be reflected by the explicit behaviors, such as the usage of language (e.g., number of professional terminologies) and search behaviors (e.g., query complexity, click complexity) (M. J. Wilson & Wilson, 2013; Eickhoff et al., 2014; Chi et al., 2016).

## 3.2 Research Questions

In this dissertation, the conceptual framework provides an orienting lens to help shape the research questions and guide the research design. Therefore, this dissertation answers four research questions. RQ1 and RQ2 are asked to disclose the linkage between levels of learning and OHIS behaviors from the *method dimension* (RQ1) and *information dimension* (RQ2). RQ3 examines *knowledge change*. Finally, RQ4 is to propose design implications based on the results and findings.

- **RQ1:** How do health consumers search and learn in OHIS to achieve different levels of

learning objectives (method dimension)?

**RQ1a:** How do health consumers' search behaviors vary by different levels of learning?

**RQ1b:** How do health consumers' selected online sources vary by different levels of learning?

- **RQ2:** What information do health consumers search and learn in OHIS to achieve different levels of learning objectives (information dimension)?

**RQ2a:** What are health consumers' information needs of different levels of learning?

**RQ2b:** What types of information can health consumers acquire for different levels of learning?

- **RQ3:** How well do health consumers learn in OHIS (knowledge change)?
- **RQ4:** What are the implications for system design?

## 4.0 Methodology

This chapter provides the high-level overview of the methodology. In general, four sub-studies are conducted to answer the research questions derived from the conceptual framework. Section 4.1 first presents an overview of the research design as well as a breakdown of the components. Section 4.2 then introduces the overview of the data collection methods and the rationale for choosing each specific data collection method. The sections in this chapter highlight the relationship among the four studies and their relationship with the research questions, while the details of each study, including data collection, analysis and results, are elaborated on separately in Chapters 5–8. Chapter 9 presents the discussions and implications of the results and findings.

### 4.1 Research Design Overview

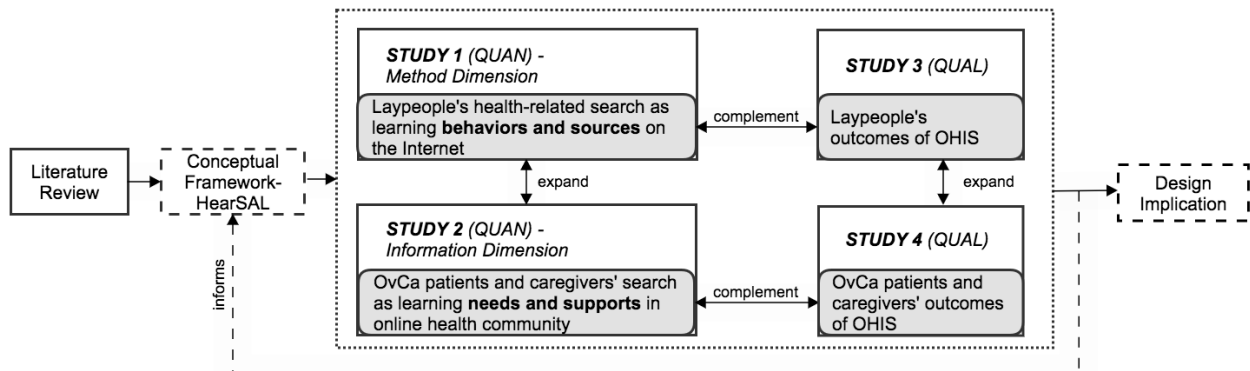


Figure 3: An illustration of the Research Design Overview

This dissertation employs a mixed-method research design (Creswel, 2009, p. 215). Specifically, the research design is guided by a conceptual framework (i.e., Figure 2), and mixed methods are used to collect both qualitative data and quantitative data. Figure 3

delineates the essential components and sequence of flow sequence in the research design. In the chart, blocks show each component in the study, and directional arrows illustrate the sequence of the components in the entire research flow. The blocks shaded gray indicate the four sub-studies of this dissertation.

Once the conceptual framework (Chapter 3) has been proposed based on the literature review (Chapter 2), the four studies are conducted (Chapter 5,6,7,8). Subsequently, the results and findings of the four studies are triangulated and inform the design implications (Chapter 9).

In accordance with the conceptual framework, my investigation of consumers' search as learning in OHIS is divided into four sub-studies: two quantitative studies (Study 1 & 2) focusing on the method dimension (RQ1) and information dimension (RQ2) of health consumers' search as learning in OHIS, and two interview studies that directly elicit health consumers' experience and perspectives towards OHIS and subsequent outcomes, with an emphasis on knowledge change as a cognitive outcome (RQ3). Specifically, laypeople (Study 3) as well as ovarian cancer patients and caregivers (Study 4) are interviewed. Finally, the research question with respect to design implications (RQ4) is answered by synthesizing all the results and findings.

The relationship between Study 1 and Study 2 is expansion, as the two studies highlight different inquiry components and adopt different methods and population groups, thus extending the inquiry's breadth and range (Creswel, 2009). By using the qualitative interview method, Study 3 and Study 4 contribute insights into health consumers' knowledge change through OHIS in consumers' natural settings, which is less revealed and discussed in Study 1 and Study 2. Therefore, Study 3 and Study 4 are complements to Study 1 and Study 2, respectively.

## **4.2 Data Collection Overview**

Generally, health consumers engage in OHIS through three major approaches: 1) directly searching and navigating the internet, 2) participating in online health communities, and 3)

Table 1: Overview of Data Collection Methods and Datasets

| <b>Study</b> | <b>Target Population</b>     | <b>Data Collection Method</b> | <b>Dataset</b>  |
|--------------|------------------------------|-------------------------------|---|
| 1            | Laypeople                    | User study                    | 5298 search log records from 24 participants who search for two scenarios and six tasks. Each user study lasted approximately 2h.         |
| 2            | OvCa patients and caregivers | Web scraping                  | 909 initial posts and 14816 comments posted from June 2010 to December 2020 in National Ovarian Cancer Coalition CancerConnect Community. |
| 3            | Laypeople                    | Semi-structured interview     | 24 interview transcripts, each last approximately 30 min.   |
| 4            | OvCa patients and caregivers | Semi-structured interview     | 20 interview transcripts, each last approximately 1h.   |

communicating with health professionals on the web (Cline & Haynes, 2001). In this study, as the focus is on consumer self-motivated active search as learning, the data collection attempts to cover the first two means of accessing online health information.

Notably, as suggested in the literature review in Section 2, laypeople mainly rely on search engines to initiate the OHIS (Pang et al., 2014; Yilma et al., 2017; Eysenbach & Köhler, 2002; Fox, 2013; D. L. Hansen et al., 2003), due to their limited prior knowledge about the health concern and information sources. In contrast, the online health community (OHC) has the benefits of peer interaction and distinguishes itself for OvCa patients and caregivers who have a constant and evolving need for information. Therefore, the data from laypeople and OvCa survivors are collected from two settings. In addition, two semi-structured interviews with the same populations are conducted to offset the weakness of the two quantitative studies.

Table 1 provides an overview of the data collection methods as well as the datasets. The laypeople’s dataset contains 5298 search log records and interview transcripts from 24 participants, each of whom searched for two scenarios with different severities and six subtasks with different levels of learning goals. The data is collected from a lab-controlled user study consisting of a pre-search survey, in-situ search recording, and a post-search semi-structured interview. The data was collected from May 2017 to August 2017.

The data collection for understanding the OvCa community is collected from the National Ovarian Cancer Coalition (NOCC) CancerConnect Community<sup>1</sup>, a peer-support online forum for OvCa community, as well as semi-structured interviews with women with OvCa and their family caregivers. Upon approval, 909 initial posts and 14816 comments posted from June 2010 to December 2020 in NOCC are collected. In addition, 20 participants who were either OvCa patients (n = 18) or family caregivers (n = 2) were interviewed from May 2020 to November 2020.

---

<sup>1</sup><http://nocccommunity.ovarian.org/>

## 5.0 Study 1: Laypeople’s Health-related Search as Learning Behaviors on the Internet

As described in the research design previously, Study 1 is conducted to unveil the method dimension of consumers’ SAL in OHIS and answer RQ1. A user study was designed and carried out to collect rich behavioral data. This chapter presents the data collection, analysis, and results of the user study.

### 5.1 Data Collection

This section elaborates on the user study design and data collection for understanding laypeople’s SAL, including the choice of health conditions and task design, system and interface, procedure, and participants. Overall, it is a controlled study around scenario-based health problems, with a live search system for the recruited laypeople to search on. The study design was approved by the Human Research Protection Office (formerly IRB) of the University of Pittsburgh.

#### 5.1.1 Health Condition and Task Design

Two health conditions with different severity levels are selected: Severe Condition (SC) and Mild Condition (MC). The SC task is a more urgent and complex health issue compared to the MC. This is because, as suggested in previous work, even laypeople may behave and perform differently in search as learning for health conditions with different severity (De Choudhury et al., 2014). As for the specific health issues, multiple sclerosis (a demyelinating disease) is chosen for the former and weight loss for the latter. It should be noted that weight loss could be a severe condition in certain situation, but in the current study, the scenario of weight loss is designed as a mild condition.



Table 2: The Mapping Between OHIS Intentions and Learning Objectives

| <b>Levels of Learning Objectives</b> | <b>HIS Intentions</b> | <b>Task Design Guidance</b>   |
|--------------------------------------|-----------------------|---|
| Understand                           | Evidence-based HIS    | pursuit of details and relevance of signs and symptoms  |
| Analyze                              | Hypothesis-based HIS  | pursuit of content on one or more illnesses, and on the discrimination among different diseases under the uncertainty |
| Evaluate                             | Decision-based HIS    | given specific situation of a condition, collect causes, treatments, tests and other information to make decision     |

**5.1.1.1 Scenario** The widely-used simulated work task situation (Borlund, 2000) is adopted as the guideline for the task design. A simulated work task situation requires a description of a realistic search scenario, which motivates the participants to search to fulfill the goals. In the scenarios, the participants are asked to imagine that they need to search on behalf of friends or family members who are experiencing a health issue. Particularly, the participants were required to search for someone else instead of for themselves because although the tasks are designed to resemble real-world needs as closely as possible, these needs are still not originating from the participants themselves. Additionally, searching for others is a very common situation in real life (Cutrona et al., 2015).

**5.1.1.2 Task Design** The design of the specific search task is guided by Table 2. The simulated tasks should be both 1) close to real OHIS information needs on the web and 2) capable of reflecting learning with different levels of objectives. Therefore, I map the evidence-based OHIS and hypothesis-based OHIS intentions established by Cartright et al. (2011) through large-scale real search logs to the two levels of A & K's taxonomy (L. Anderson et al., 2001): Understand and Analyze, and propose a third intention for the Evaluate level of learning: decision-based OHIS. Based on the mapping, the specific search goals for each

OHIS intention and level of learning are described in Table 2. This framework is used to design the specific scenario-based search tasks in the user study. The complete scenario and task description are shown in Table 3.

Table 3: Complete Description of Task Scenarios and Subtasks

| Conditions | Severe Condition (SC) - Multiple Sclerosis  | Mild Condition (MC) - Weight Loss   |
|------------|---|---|
| Scenario   | Your friend Cathy, age 32, was diagnosed with <b>Relapsing-Remitting Multiple Sclerosis (RRMS)</b> 8 years ago. To treat her form of Multiple Sclerosis (MS) she has been prescribed a medication known as Avonex. She was functioning well until about a month ago when she started experiencing vision loss and muscle spasms with greater frequency. Upon visiting her physician, she learned that her Relapsing-Remitting Multiple Sclerosis had changed into Secondary Progressive Multiple Sclerosis (SPMS). <b>You want to do some research on her condition and give her suggestions.</b> | Your 15-year-old nephew Johnson is considering <b>going on a diet and losing weight.</b> His weight is 200 lbs and he is 5 feet and 8 inches tall. Overweight teenagers are at high risk of having health issues such as diabetes and high blood pressure. But you know that it is also important to be healthy while going on a diet. So, <b>you want to collect some information and give suggestions to your nephew.</b> |
| Understand | First find out what is Secondary Progressive Multiple Sclerosis (SPMS)? <b>Describe the disease and list five symptoms of SPMS.</b>   | Our bodies need protein from the foods we eat to build and maintain bones, muscles and skin. First <b>find out how much protein your nephew needs given his current condition</b> (e.g., weight, height, age, gender).  |
| Analyze    | Doctors have identified 4 major types of MS. Your friend experienced 2 types of MS (RRMS and SPMS). Please find out how these two differ from the 2 other types of MS? <b>Describe the differences and similarities in how patients with each type of MS behave.</b>  | Protein can be categorized into two types: <b>complete protein and incomplete protein. Please find out the major difference between them.</b> Which of the following food(s) provide complete protein: yogurt, rice, beans, coconut, nuts?  |

|          |  |  |
|----------|--|--|
| Evaluate | Your friend is seeking a new treatment plan for her SPMS. You want to collect some information and create a basic treatment plan for her before she goes to see the doctor. Given the fact that she also has a pre-existing heart condition, <b>do you think she should continue with Avonex? Why? If not, what alternative medicine would you suggest? Are there any specific types of exercises and suggested dietary practices that are recommended for individuals with MS? Please also include them in your treatment plan.</b> | Your nephew is considering a diet and exercise plan. You want to help him and design a plan for him. Some people argue that the high protein diets can help reduce the desire for food. People with high protein diets can remain healthy and lose weight at the same time. But there are also papers arguing high protein diets can do harm, especially for long-term. Your nephew wonders <b>if he should go on a high protein diet. Besides, what alternative diet would you suggest? Please give your opinions and reasons in your recommended plan.</b> |
|----------|--|--|

### 5.1.2 System and Procedure

**5.1.2.1 Search System** A search system called HIS (health information-seeking system)<sup>1</sup> is designed and implemented to collect and record search behaviors. Figure 4 shows a screenshot of the homepage. The search function is wrapped around Google Search API. Like Google, there is a search box in the center of the page for entering queries. A unique user ID and password are assigned to each participant beforehand. After login, the participant with the ID S01 would be directed to the homepage.



Figure 4: A Screenshot of the Homepage of HIS

<sup>1</sup><http://crystal.exp.sis.pitt.edu/his/login.html>

Then, Figure 5 shows a screenshot of the search result page, directly provided by Google. All users' interactions with the system, such as queries, clicks, and dwell times, are automatically logged to a MySQL database. Each record in the log refers to a source clicked by the participants.

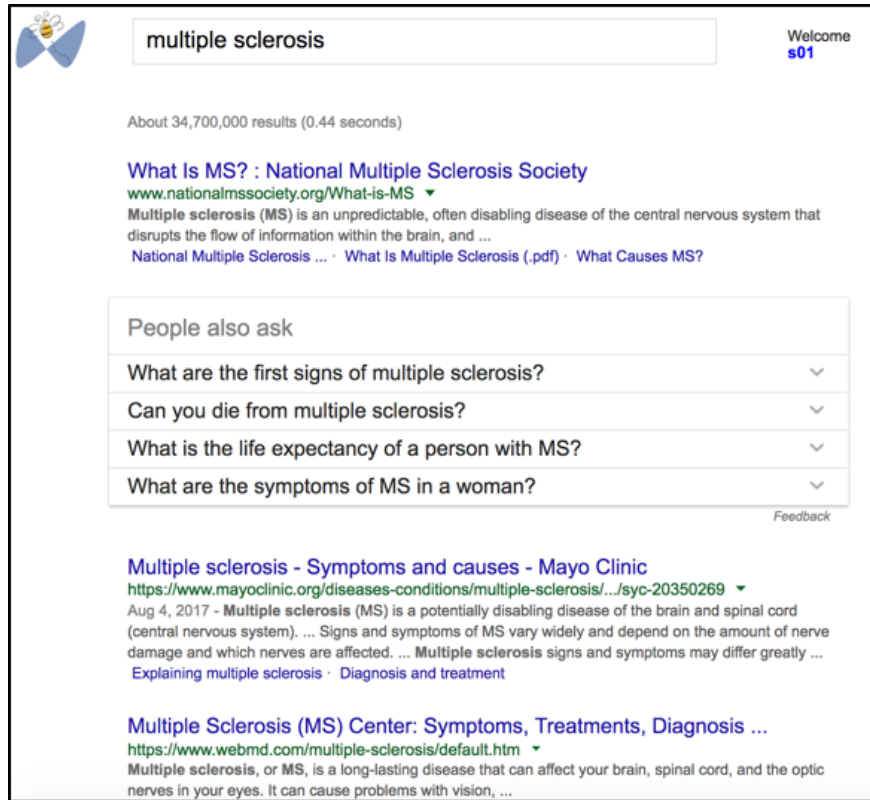


Figure 5: An Example of the Search Result Page

**5.1.2.2 Capturing Source Selection Behavior** In order to capture and collect rich data on the laypeople's source selection behavior, a special feature "save to workspace" is designed and embedded in the system. A "save to workspace" button is located in the bookmark menu of the browser. As shown in Figure 6, whenever the participant finds a useful piece of information, she can simply highlight the snippet and click "save to workspace". This action triggers a separate pop-up page (see Figure 7), which prompts the participant to rate the relevance of the snippet from 1 to 5 before she clicks the "submit" button. The "save to workspace" function is supported on any webpage including the search result page.

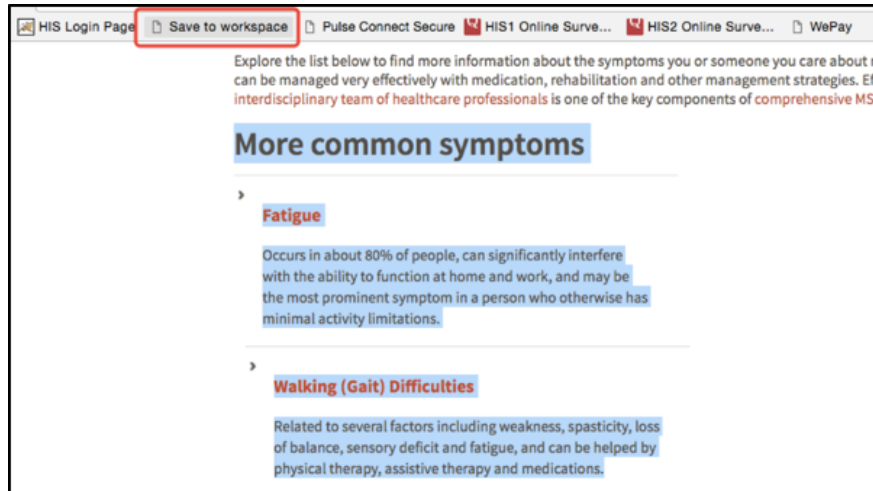


Figure 6: An Example of the "Save to Workspace" Interface(a): the user is highlighting a piece of information and then clicking "save to workspace."

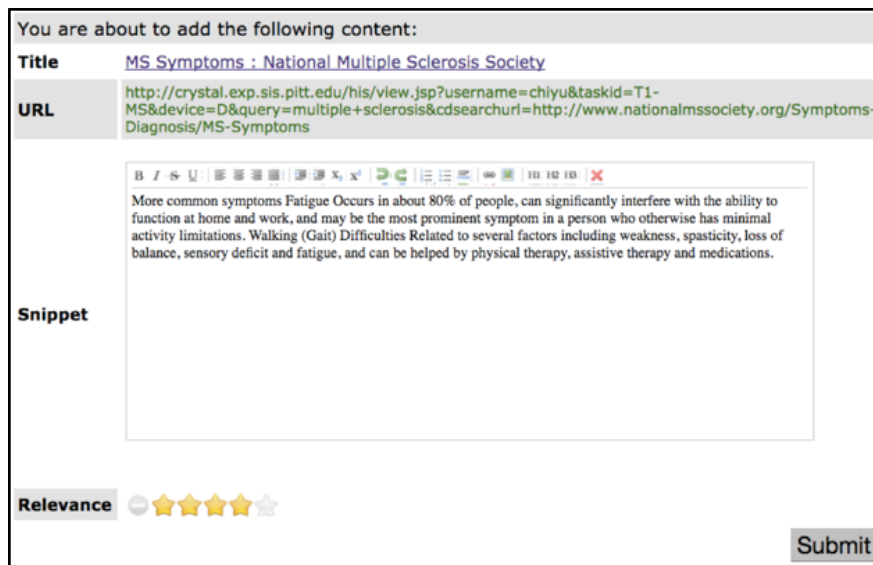


Figure 7: An Example of the "Save to Workspace" Interface(b): A pop-up window of the snippet that is going to be saved.

**5.1.2.3 Procedure** Figure 8 provides an overview of the procedure of the user study for each participant. The user study was conducted in a controlled lab setting. Upon arrival,

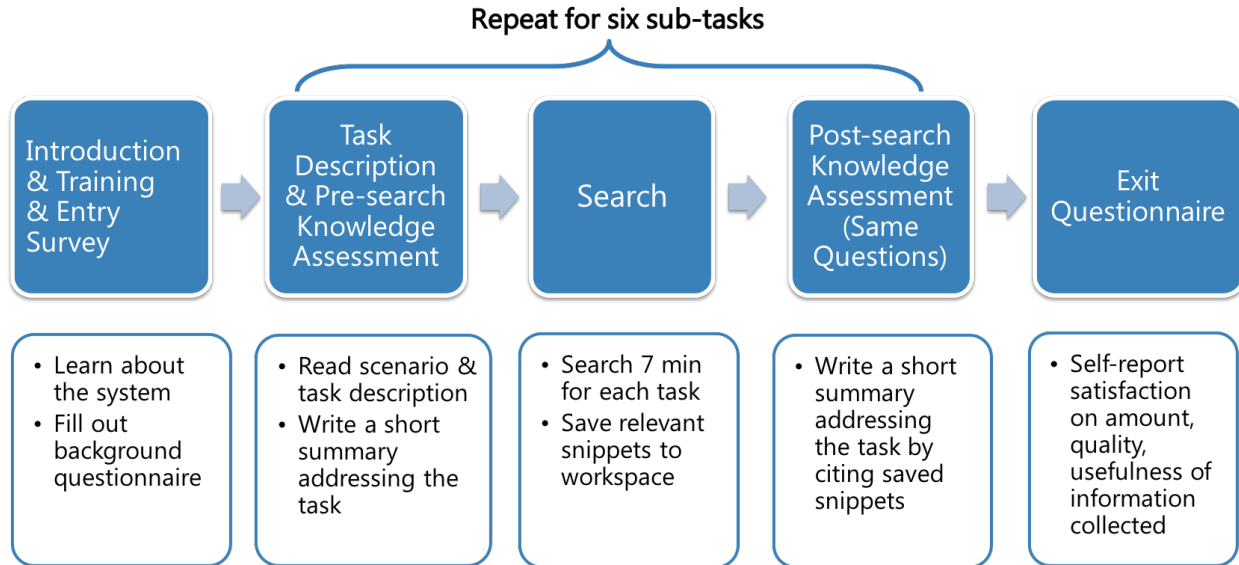


Figure 8: User Study Procedure

each participant received a brief introduction about the study (see Appendix A) and was asked to fill out an entry questionnaire (see Appendix B) that aimed to collect background information (e.g., age, education) and past experience on OHIS. Then the participant worked on the search system for 5 minutes with a prepared training task and explicitly informed that she could freely search, click and view webpages and that all the interactions with the system would be logged and analyzed afterward. The “save to workspace” function was demonstrated to the participant so that she knew how to save snippets into the workspace to generate solutions to the tasks.

The experiment adopted a within-subject design, so each participant worked on 3 sub-tasks associated with both health conditions. The order of the two health conditions was rotated under the Latin square principle to minimize fatigue effects. As a result, each participant completed 6 sub-tasks.

In the search for each health condition, the participant first received a printed description of the scenario. Then the participant was presented with the first sub-task and asked to write a short summary about what they had already known about the task by typing in

the system. Each sub-task requested the written summary to assess the participants' prior-search knowledge about the task. The participants were asked to perform searches for 7 minutes per subtask, considering the total length of the study. The two participants of the pilot study confirmed that 7 minutes is appropriate for the tasks, and in the formal data collection, all the participants met the time requirements. After the 7-minute search, another written summary about the sub-task was requested again from the participants to assess the after-search knowledge. In total, the participant would search for 21 minutes to complete all three subtasks and provide 3 prior-search and 3 after-search written summaries. After a five-minute break, the participant moved onto the second health condition with the same procedure. After finishing the search for both health conditions, the participant was requested to fill out an exit questionnaire about their satisfaction of the search. The whole study took approximately 1.5 hours, and each participant was compensated US \$20.00.

### 5.1.3 Participants and Dataset

Participants were recruited through flyers posted on the campuses of the University of Pittsburgh and Carnegie Mellon University. Two criteria were applied to screen eligible participants before they formally signed up. Firstly, individuals who had completed or were currently pursuing health- or medicine-related education programs were excluded because they were not considered laypeople. Secondly, potential participants must be native English speakers over the age of 18. In total, 24 college students were recruited (15 females and 9 males, 17 undergraduates, and 7 graduate students, 18-33 years old). On a scale from 1 to 5, participants described their search activity as very frequent ( $M= 4.79$ ,  $SD=.104$ ) and very experienced ( $M= 4.25$ ,  $SD= .124$ ). Regarding OHIS, most had searched for health-related information recently (in the last week:  $n= 15$ , 62.8%; last month:  $n= 8$ , 33.3%; last six months:  $n= 1$ , 4.2%). The majority of them also reported searching on a weekly ( $n=8$ , 33.3%) or monthly ( $n=8$ , 33.3%) basis.

As a result, 48 complete search logs (144 search sessions) were collected from 24 participants. The dataset contains 5,298 clicked webpages, among which 965 (18.21%) were saved into the workspace.

## 5.2 Data Analysis

The data collected from the user study was analyzed with three steps. First, the recorded search logs were analyzed to reveal the participants' different search behaviors on the query level and clicking level (Section 5.2.1). Then, a coding framework was applied to classify all the online sources visited by the participants (Section 5.2.2). This analysis provided an in-depth understanding of participants' source selection behavior. Finally, the 144 written summaries were assessed manually to measure the participants' prior-search and after-search knowledge (Section 5.2.3).

### 5.2.1 Search Behavior Measures

Issuing queries and clicking and navigating search results are two typical search behaviors (J. Jiang, He, & Allan, 2014). Therefore, the search logs were analyzed on both the query and click levels.

On the **query** level, four measures were computed:

- Number of Queries (NQ): number of queries per participant per sub-task
- Length of Queries (LQ): number of query terms per participant per sub-task
- Number of Unique Queries (NUQ): number of unique queries among all the participants per sub-task
- Number of Unique Terms (NUT): number of unique terms among all the participants per sub-task

On the **click** level, three measures were used in accordance to the different source selection behaviors:

- Number of Clicked Pages (NCP): number of clicked pages per participant per sub-task
- Number of Saved Pages (NSP): number of pages saved to the workspace per participant per sub-task
- Number of Used Pages (NUP): number of pages cited as references in the after-search written summaries per participant per sub-task



Combining the query- and click-level measures, another measure was calculated:

- Search Efficiency (SE):  $SE = (\sum_{i=1}^M RSP_i) / NQ$

A larger SE indicates that the participant collected more relevant information with fewer queries. In the formula above,  $SE$  indicates the search efficiency of a participant in a sub-task.  $RSP_i$  is the relevance score of the saved page  $i$ , and the relevance score was rated by the participant when the page  $i$  was saved to the workspace.  $M$  and  $NQ$  are the total number of the saved pages and the total number of the queries issued by the participant to complete the sub-task.

### 5.2.2 Online Health Source Classification

Due to the lack of a widely adopted classification scheme for online health information sources, it was decided to develop one in order to classify the sources selected by the participants. Initially, the websites were classified by their different domain names. Domain names are designed to represent websites distributed among various hosts and network systems, with a string of characters usually separated by dots as their structure. However, since no restrictions were imposed on the sources that the participants could select during the search, as expected, it was found that the sources they visited were scattered across different types of websites. Even the websites with the same domain names might function very differently during the search process. For example, participants might consult a social network site and an e-commercial site in the same task. Both websites have dot com domain names, but it is not appropriate to rigidly classify them into a single simple category. Given that the website type classification is too broad, a coding framework for classifying webpage types was developed. The mapping to the webpage category was completed through content analysis by two coders.

Researchers (Kanaris & Stamatatos, 2009) often distinguish webpages by topic (e.g., health vs. sport) and genre (e.g., online news vs. online forums). In this study, the topic is restricted to health. Thus, a webpage was classified based on its genre and a set of nine genre types emerged. Each webpage was mapped to a genre through manual coding by two coders. To generate the base codes, the coders separately open-coded the transaction log of

Table 4: Coding Framework of Webpage Types

| <b>Code</b> | <b>Webpage Type</b>       | <b>Description</b>   |
|-------------|---------------------------|--|
| SERP        | Search engine result page | A search engine result page for a given query.   |
| HSP         | Health-specific page      | A webpage that offers focused information about one specific health issue, whose content is often contributed by professionals. Webpages promoting or selling any product such as pharmaceutical, home remedy, diet product are excluded and classified as E-commerce pages (ECP). |
| WBP         | Wiki-based page           | A wiki page that collects voluntary creation from general users.   |
| ALP         | Academic literature       | A webpage whose main content is academic literature.   |
| PP          | Personal webpage          | A webpage sharing authors' personal ideas and experiences; self-portrayals, personal homepages with informal content.  |
| NAP         | News article page         | A webpage filled with long reviews, reports, newsfeeds or photos.  |
| SNS         | Social network service    | A webpage supporting interactions among users. Through the interactions, users may engage in questioning, answering, emotional support, human experiences, etc. Wiki-based pages are excluded as their core is to co-edit instead of interaction.                                  |
| ECP         | E-commerce page           | A webpage with the main purpose of promoting products and supporting online purchase.  |
| OEP         | Online education page     | A webpage from educational institutes with the main goal of online education with tutorials (in digital print, video, etc.).   |

the first participant (i.e., S01). Seven types of webpages emerged, and the coders achieved an acceptable agreement (Cohen’s Kappa = 0.891). During the subsequent debriefing, the coders discussed controversial records and two more types were added. Table 4 provides the final nine genre types.

It is assumed that webpages belonging to the same type of genre resemble each other in terms of presentation style, purpose, function, or even content. Some webpage genres may suggest higher-quality content than others. For example, Academic literature (ALP) is more likely to provide professional instructions. However, it is worth noting that the current coding framework was not used to measure the quality of the webpage specifically, so it is possible that the quality of a certain type is not identical.

### 5.2.3 Written Summary Assessment

The prior-search and after-search written summaries were assessed using the method proposed by Kammerer, Nairn, Pirolli, and Chi (2009). First, a coding framework was defined to assess the quality of the written summary in accordance with the task description. For example, in the Understand level in SC, the summaries were rated based on the description of the secondary progressive multiple sclerosis (0-1 point) and 5 symptoms of it (0-1 point per symptom). Then two coders who were familiar with the search tasks and the assessment criteria rated all the written summaries manually. Pearson correlation was used to measure the inter-rater reliability, and the two coders obtained an acceptable agreement on all the sub-tasks, ranging from 0.563 to 0.875,  $p < 0.001$ .

The the scores of the written summaries were used to gauge the participants’ knowledge state before and after the search session, and the knowledge change:

- Prior-search Knowledge (PK): normalized scores of the prior-search written summaries;
- After-search Knowledge (AK): normalized scores of the after-search written summaries;
- Knowledge Change (KC):  $KC = AK - PK$

Table 5: Written Summary Assessment Criteria

| Conditions            | Understand   | Analyze  | Evaluate   |
|-----------------------|--|--|--|
| Severe Condition (SC) | Disease description (0-1 point);<br>5 symptoms (0-1 point each)  | Names of the four types of multiple sclerosis (0-1 point each);<br>Characteristics of each type (0-1 point each)   | Covering the following topics, for each topic, points from 0 (not covered) to 2 (detailed focused coverage): Avonex and heart risk (0-2 points); alternative medicine (0-2 points); exercises (0-2 points); Dietary (0-2 points) |
| Mild Condition (MC)   | Amount of protein needed (0-2point);<br>Characteristics (e.g., gender, age) of the patients considered (0-2 points), and the reason (0-2 points) | Characteristics of complete and incomplete proteins (0-1 point each);<br>Difference between the two types of proteins (0-1 point);<br>Types of protein of the 5 foods (0-1 point each) | Covering the following topics, for each topic, points from 0 (not covered) to 2 (detailed focused coverage): high protein diets, pros (0-2 points) and cons (0-2 points); alternative diet (0-2 points); reasons (0-2 points)    |

### 5.3 Results

#### 5.3.1 Search Behaviors

A two-way within-subjects ANOVA was conducted to reveal the effect of the two independent variables (levels of learning objectives and severity of health conditions) and their interaction effect on the three groups of search behavior measures: search queries, clicks, and search efficiency.

**5.3.1.1 Search Queries** Figure 9 shows the averaged number of queries (NQ) and length of queries (LQ) issued by the participants for sub-tasks with different levels of learning objectives. Besides, the queries of the two health conditions were separated to unveil the effects of the conditions.

Combining the results of NQ and LQ, it was found that in the search and learning of

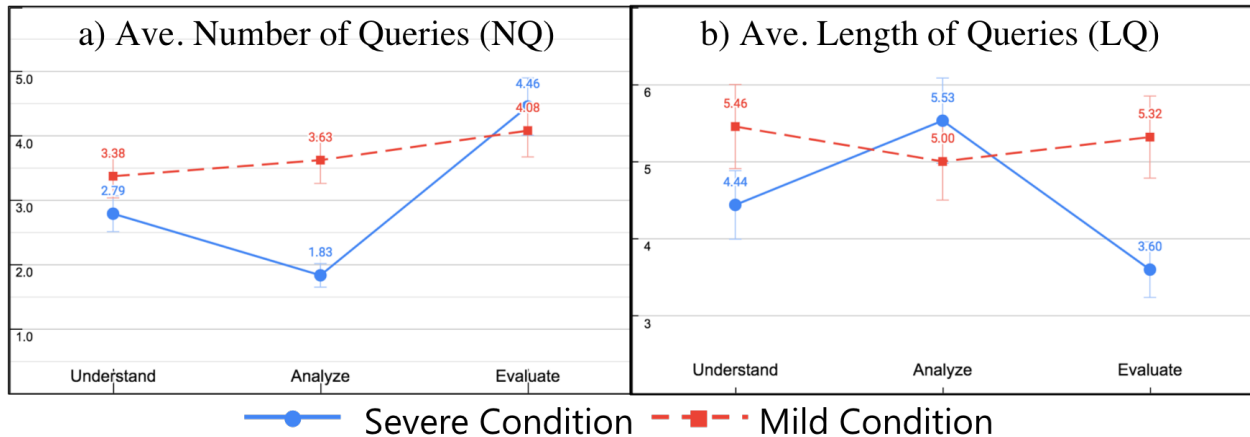


Figure 9: a) Number of Queries (NQ) and b) Length of Queries (LQ)

the severe condition (SC), to achieve the Analyze level of learning, participants used longest but fewest queries, whereas to achieve the Evaluate level, participants used the shortest but the most number of queries.

### Number of Queries

The results of the two-way repeated measures ANOVA revealed that there was a significant interaction between levels of learning and severity of health conditions on number of queries (NQ) issued by the participants ( $F(2, 46) = 8.130, p = .001, \text{partial } \eta^2 = .261$ ). Significant main effects were also found on levels of learning ( $F(2, 46) = 12.809, p = .000, \text{partial } \eta^2 = .358$ ) and severity of health conditions ( $F(1, 23) = 11.265, p = .003, \text{partial } \eta^2 = .329$ ). The significant differences in the number of queries among levels of learning for each condition were followed by simple pairwise comparisons among three levels while using Bonferroni correction for control of familywise Type I error rate.

For SC, participants issued the most queries in the Evaluate task ( $NQ = 4.46$ ), followed by the Understand task ( $NQ = 2.79, p = .001$ ), and the fewest in the Analyze task ( $NQ = 1.83, p = .000$ ). The difference between the Understand task and the Analyze task is also significant ( $p = .041$ ). There were no significant differences in the number of queries among

the three levels of learning objectives for MC.

### Length of Queries

Similar to the NQ, there was significant interaction effect between levels of learning and severity of health conditions on length of queries (LQ) ( $F(2, 46) = 9.467, p = .000$ , partial  $\eta^2 = .292$ ). Besides, significant main effects were also found on levels of learning ( $F(2, 46) = 6.724, p = .003$ , partial  $\eta^2 = .226$ ) and severity of health conditions ( $F(1, 23) = 8.188, p = .009$ , partial  $\eta^2 = .263$ ) on LQ.

Significant differences in LQ among the three levels of learning were only found in the SC. The longest queries were issued in the Analyze level ( $LQ = 5.53$ , followed by the Understand level ( $LQ = 4.44, p = .046$ ), and the shortest in the Evaluate level ( $LQ = 3.60, p = .000$ ). The LQ in the Understand level was longer than the Evaluate level  $p = .037$ ).

### Number of Unique Queries and Terms

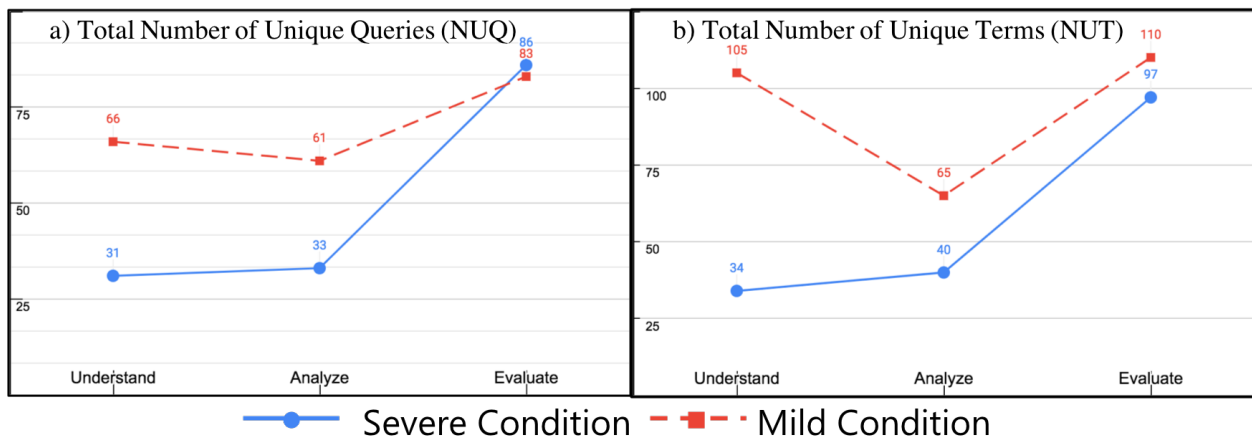


Figure 10: a) Number of Unique Queries (NUQ) and b) Number of Unique Terms (NUT)

The total number of unique queries (Figure 10 (a) NUQ) and unique terms (Figure 10 (b) NUT) among all the 24 participants were plotted across the three levels of learning and split by the two health conditions. As shown in the figure, in the SC, the NUQ and NUT were similar in the Understand and the Analyze level, which might mainly come from

the task description. However, in the MC, the participants were able to use many unique terms in the Understand level, which might arise due to their intrinsic familiarity with the tasks in the MC (i.e., protein amount). It is notable that, in both conditions, participants issued the largest number of unique queries and terms in the Evaluate tasks, suggesting that the participants used diverse queries and terms beyond the task description to achieve the Evaluate level of learning regardless of the severity of the health condition.

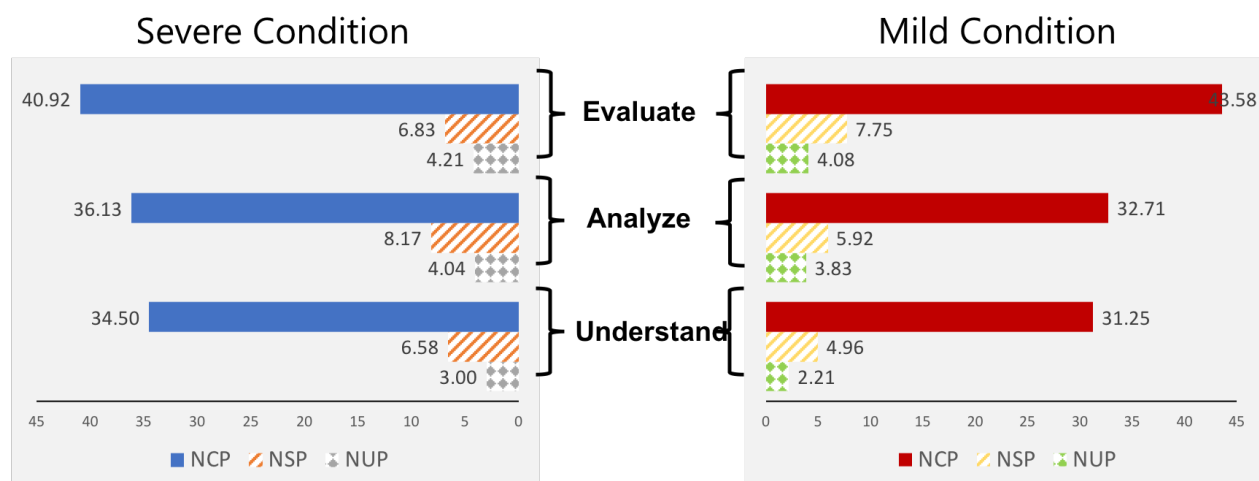


Figure 11: Clicks of the Three Levels of Learning in the Two Conditions

NCP: Number of Clicked Pages; NSP: Number of Saved Pages; NUP: Number of Used Pages

**5.3.1.2 Clicks** Figure 11 depicts the average number of clicked pages (NCP), saved pages (NSP), and used pages (NUP) of the 24 participants for the three sub-tasks with different levels of learning in the two health conditions (left: severe condition; right: mild condition). Note that since the saved pages are from the clicked pages, and the used pages, the ones cited in the written summaries, are further selected from the saved pages in the workspace, NCP is inherently larger than NSP, and NUP is the smallest among the three types of clicks.

The results of the two-way ANOVA indicated that there was a significant interaction effect between level of learning and severity of health conditions on the number of saved pages (NSP) ( $F(2, 46) = 8.668, p = .001, \text{partial } \eta^2 = .274$ ). However no significant interaction effects were found on the number of clicked pages (NCP) ( $F(2, 46) = 1.309, p = .280$ ) and

used pages (NUP) ( $F(2, 46) = 0.538, p = .0588$ ). In addition, it is notable that level of learning had a significant main effect on all the three measures of clicks: NCP ( $F(2, 46) = 15.023, p = .000$ , partial  $\eta^2 = .395$ ), NSP ( $F(2, 46) = 5.982, p = .005$ , partial  $\eta^2 = .206$ ), and NUP ( $F(2, 46) = 11.929, p = .000$ , partial  $\eta^2 = .342$ ). Severity of condition itself had no significant main effect on the clicks.

Follow-up pairwise comparison using Bonferroni correction revealed that significant different numbers of clicks affected by the level of learning mainly appeared in the MC. To complete the search tasks of the Evaluate level in the MC, participants significantly clicked ( $NCP = 43.58$ ), saved ( $NSP = 7.75$ ), and used ( $NCP = 4.08$ ) more pages than the Understand level task ( $NCP = 31.25, p = 0.001$ ;  $NSP = 4.96, p = 0.000$ ;  $NUP = 2.21, p = 0.000$ ). The clicks in the Evaluate level were also more than those in the Analyze level ( $NCP = 32.71, p = 0.002$ ;  $NSP = 5.92, p = 0.034$ ;  $NUP = 4.96, p = 1.000$ ), and the differences were significant except for the NUP. The differences of clicks between the Analyze level and the Understand level was only found significant in the number of used pages ( $p = 0.039$ ). In contrast, though the clicks also appeared to increase with the level of learning objectives in the SC, statistically significant result was only found between the NUP in the Evaluate level ( $NUP = 4.21$ ) and the NUP in the Understand level ( $NUP = 3.00, p = 0.017$ ).

**5.3.1.3 Search Efficiency** The search efficiency is visualized in Figure 12. A larger SE indicates that the participant collected more relevant information with fewer queries. Therefore, it is evident that participants' achieved the highest SE in the Analyze level in the SC tasks, while there was a dramatic drop from the Analyze level to the Evaluate level following it. Conversely, the SE tended to be similar among the three levels of learning in the MC tasks.

A significant interaction effect between levels of learning objectives and severity of health conditions on search efficiency was found by using the two way ANOVA ( $F(2, 46) = 17.833, p = .000$ , partial  $\eta^2 = .437$ ). There was a significant main effect of levels of learning objective on the search efficiency ( $F(2, 46) = 13.579, p = .000$ , partial  $\eta^2 = .371$ ). Significant effect of severity of condition was also found ( $F(1, 23) = 10.386, p = .004$ , partial  $\eta^2 = .311$ ).

With respect to the follow-up comparison results, significant differences in the search



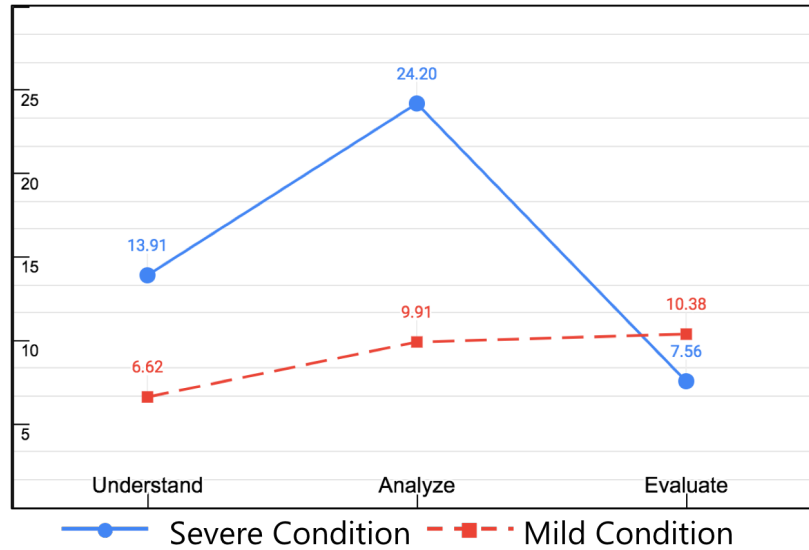


Figure 12: Search Efficiency of the Three Levels of Learning

efficiency among the three levels of learning were mainly revealed in SC. Participants' search efficiencies of the three sub-tasks in the SC were significantly different from each other. Specifically, the search efficiency was the highest in the Analyze level ( $SE = 24.20$ ), compared to the Understand level ( $SE = 13.91, p = 0.002$ ) and the Evaluate level ( $SE = 7.56, p = 0.000$ ). Search efficiency of the Understand level was higher than that of the Evaluate level ( $p = 0.012$ ). In MC, participants' search efficiency increased slightly with the increase of learning levels, though no statistically significant results were found.

### 5.3.2 Source Selections

The types of the source selected in the subtasks with different levels of learning objectives were plotted. Figure 13 and Figure 14 visualize such information with two area charts: Figure 13 shows the types of the clicked sources in the SC task, while Figure 14 displays the sources in the MC task. The distribution of source selections in each learning level was further divided into two stages evenly based on the time to demonstrate the source change on a fine-grained level.

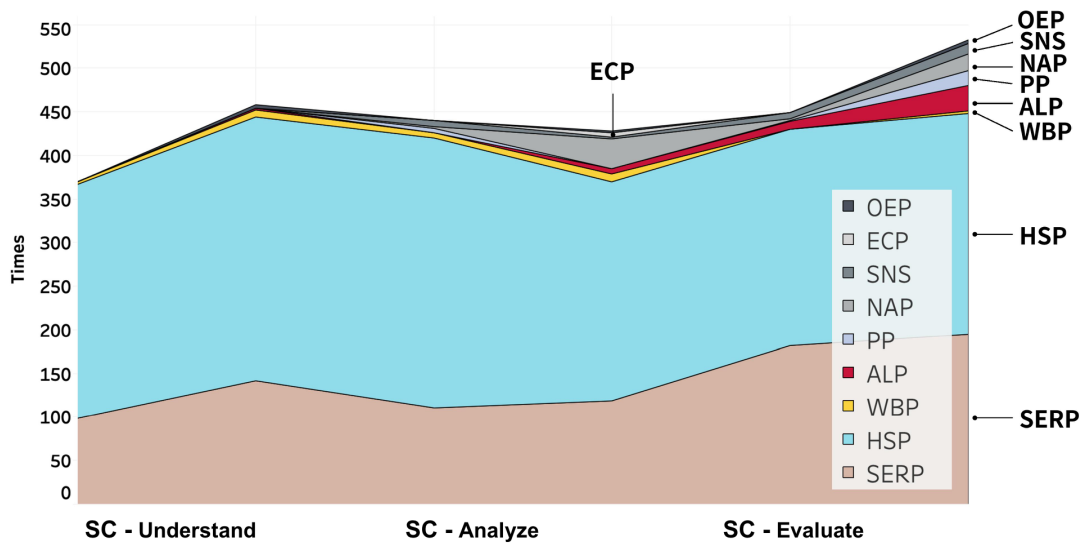


Figure 13: Types of Selected Sources in the Severe Condition (SC)

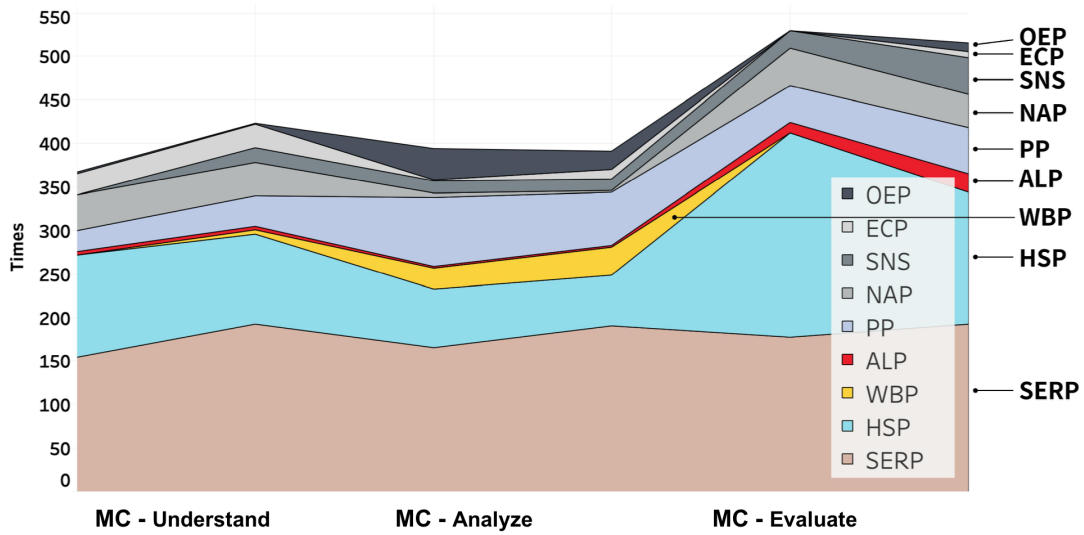


Figure 14: Types of Selected Sources in the Mild Condition (MC)

In the SC (Figure 13), Health-specific page (HSP) is the top clicked source throughout the three levels, particularly in the Understand level. However, in the MC (Figure 14), HSP does not play the same dominant role in the Understand level of learning, but its share increases dramatically in the Evaluate level; this indicates that participants' selection of HSP varies by the required knowledge level of the OHIS task they are trying to solve. Laypeople tend to select more health-specific webpages, such as WebMD.com, when they lack sufficient knowledge to solve their HIS task. Secondly, in the MC search process, participants tend to consult more diverse resources, while in the SC search process, source selection centers around HSP and search result page (SERP). Moreover, personal page (PP), which was supposed to deliver online users' personal experience and knowledge, is rarely visited when solving the SC task. However its share increases gradually throughout the MC process. Additionally, laypeople will select more diverse types of sources with the increase of the levels of the learning objectives, i.e., in the Evaluate tasks. Consequently, search engines are also employed more frequently in higher learning level tasks.

### 5.3.3 Knowledge Change

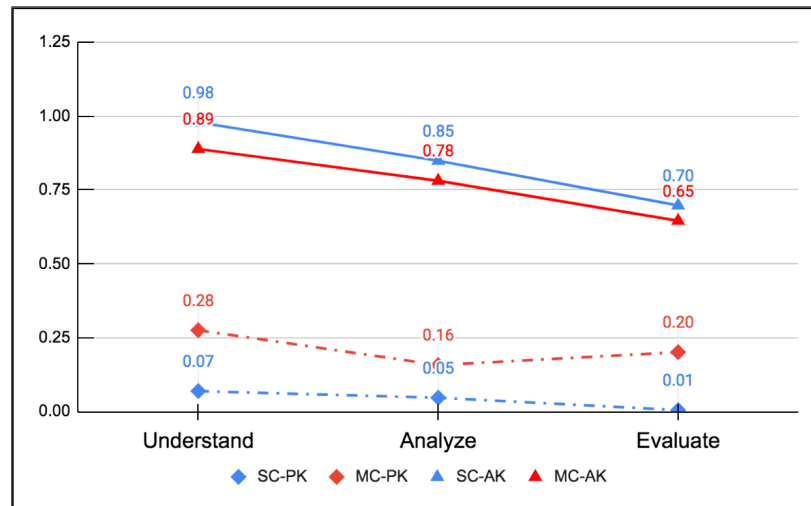


Figure 15: Prior-Search Knowledge (PK) and After-Search Knowledge (AK)

### Prior-search Knowledge and After-search Knowledge

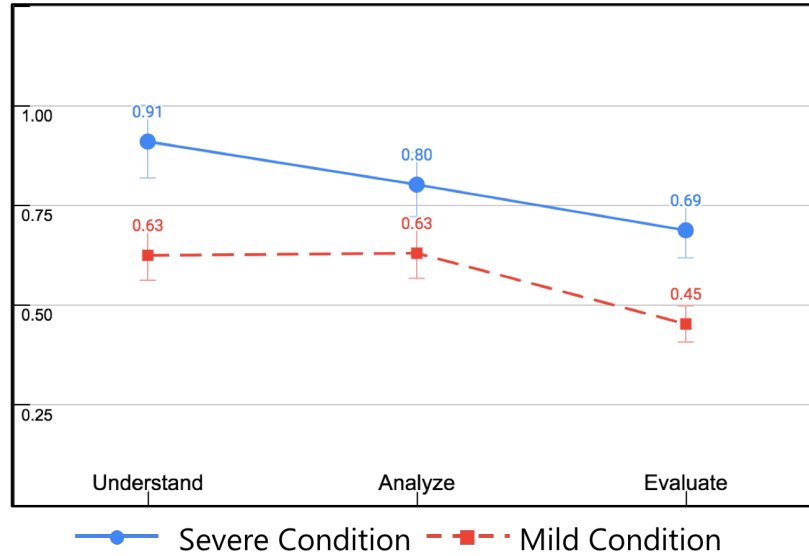


Figure 16: Knowledge Change

Knowledge change is defined as the change between the prior-search and the after-search knowledge state. In order to explore how the knowledge change differs in the three levels of learning and the two health conditions, the prior-search knowledge state and after-search knowledge state were first computed and visualized in Figure 15.

Results of two-way ANOVA indicated that levels of learning had a significant main effect on both prior-search knowledge (PK) state ( $F(2, 46) = 4.11, p = .023$ , partial  $\eta^2 = .152$ ) and after-search knowledge (AK) state ( $F(2, 46) = 33.342, p = .000$ , partial  $\eta^2 = .592$ ). Severity of condition also had a significant main effect on PK ( $F(1, 23) = 37.244, p = .000$ , partial  $\eta^2 = .618$ ) and AK ( $F(1, 23) = 5.678, p = .026$ , partial  $\eta^2 = .198$ ). It is notable that the effect size of the severity of the condition on PK was larger than that of learning levels, while learning levels had a larger effect size on AK compared to the severity of the condition. This might be because the participants' prior knowledge of the two health conditions was quite different, as one of them was a serious and complex disease (i.e., multiple sclerosis). At the same time, the other was a health-related issue (i.e., weight loss) that they were more likely to be familiar with.

In general, AK decreased with the increase of levels of learning objectives in both conditions. Pairwise comparisons among the three levels of learning suggested that participants' AK were all significantly different from each other, except for the difference between the Evaluate level and the Analyze level in the MC. Such results might indicate that laypeople were more likely to achieve a higher knowledge state through search as learning for easier tasks.

In contrast, there was no significant interaction effect between levels of learning and severity of condition on participants' PK ( $F(2, 46) = 1.253, p = .295$ ) or AK ( $F(2, 46) = 0.119, p = .888$ ).

### **Knowledge Change**

Figure 16 shows the knowledge change (KC) patterns among the three levels of learning in the two conditions. Results of two-way ANOVA demonstrated that there were significant main effects on KC from levels of learning ( $F(2, 46) = 14.224, p = .000$ , partial  $\eta^2 = .382$ ) and severity of condition ( $F(1, 23) = 45.317, p = .000$ , partial  $\eta^2 = .663$ ), though no significant interaction effect between the two factors was found ( $F(2, 46) = 0.534, p = .590$ ). The significant differences on KC among levels of learning objective for each condition were followed by simple pairwise comparisons, and the differences were found to be significant between the Evaluate level and the Understand level in both SC ( $p = .003$ ) and MC ( $p = .047$ ).

As shown in Figure 16, it is notable that participants' knowledge change generally decreased with the increasing levels of learning objectives. It should be emphasized that though the participants' prior knowledge appeared to be higher in the lower learning levels than higher levels, which means that they had already known much about the task, they still tended to gain more knowledge for the easier and less complex tasks than search tasks with a higher level of learning objectives in OHIS.

Additionally, the knowledge gain and the after-search knowledge state of SC were higher than that of MC regardless of learning levels, suggesting that OHIS is a promising learning resource for laypeople to learn about health-related information even for severe and complex diseases and for a higher level of learning objective, such as making decisions and judgments

about a disease.

## 5.4 Summary of Study 1

Table 6: Summative Results of Search Behaviors

| Variables             | Queries |    | Clicks   |          |          | Search Efficiency |
|-----------------------|---------|----|----------|----------|----------|-------------------|
|                       | NQ      | LQ | NCP      | NSP      | NUP      | SE                |
| Learning Objective    | **      | ** | **       | **       | **       | **                |
| Severity of Condition | **      | ** | Not sig. | Not sig. | Not sig. | **                |
| Interaction Effect    | **      | ** | Not sig. | **       | Not sig. | **                |

Study 1 presents a user study that examines how and how well laypeople search and learn in OHIS for two health conditions with different severity. In general, the results demonstrate that SAL behaviors and sources vary by different levels of learning objectives, and the variation is affected by the severity of health conditions.

With regard to the search behaviors, two-way within-subjects ANOVA was performed to reveal the effect of levels of learning objectives, the severity of health conditions, and the interaction effects on three groups of search behavior measures: queries, clicks, and search efficiency. Table 6 outlines the summative results. Notably, the level of learning objectives has significant main effects on all the measures of search behaviors. The severity of health conditions and the interaction effects are also discovered through many search behaviors.

As for the source selection behaviors, nine different types of online health information sources were identified based on the genre of the clicked web page—namely, search engine result page, health-specific page, wiki-based page, academic literature, personal web page, news article page, social network service, e-commerce page, and online education page. The results indicate that the types and distributions of the participants' selected online sources vary by different learning objectives and the severity of health conditions. Participants tend

to select a larger number of sources with the increase of learning levels and select more diverse types of sources in the mild condition.

The participants' knowledge change was measured through a manual assessment of written summaries. The results indicate that participants were able to achieve knowledge gain in all the levels, and the knowledge gain varies by levels of learning objectives and severity of health conditions. They tend to learn better in the easier learning tasks – the Understand level in the mild condition.

## 6.0 Study 2: OvCa Patients and Caregivers' Search as Learning Needs and Supports in Online Health Community

To answer RQ2, an analysis of an OHC of Ovca was performed to discover the information dimension of consumers' SAL in OHIS. This chapter elaborates on the data collection, analysis, and results of Study 2.

### 6.1 Data Collection

#### 6.1.1 Research Setting: National Ovarian Cancer Coalition CancerConnect Community

National Ovarian Cancer Coalition (NOCC), a nonprofit advocacy organization, has devoted itself to educating and supporting ovarian cancer survivors since its inception in 1991. The NOCC CancerConnect Community is one of the most active OvCa online health communities (OHC). It is a peer-support OHC with the goal of providing an open-access platform that encourages and enhances interpersonal learning via informational and emotional peer interactions. To this end, NOCC allows registered users to participate and contribute to the community in several ways, such as initiating a post as a *poster* and replying to the post as a *replier*, searching and reading posts and comments, creating profiles, joining groups, and sending and receiving private messages.

The goal is to understand OvCa patients and caregivers' SAL needs, support exchanged, and knowledge change in the context of OHC. One challenge is identifying the anonymous users and collecting longitude data that manifests the survivors' sustained participation and knowledge change in the community. In the NOCC community, registration is required to post and comment, so all the initial posts and comments are associated with usernames, making it possible to observe and track individuals' participation in the NOCC community.

The data collection and analysis activities were approved by the University of Pittsburgh



Human Research Protection Office (formerly IRB). The dataset contains 909 ovarian cancer threads posted from June 2010 to December 2020. Each thread is made of one initial post and corresponding comments, if any. In total, there are 909 initial posts and 14816 comments. Example of initial post and follow-up comments are shown in the next section.

### 6.1.2 OvCa Survivors' Posts and Comments

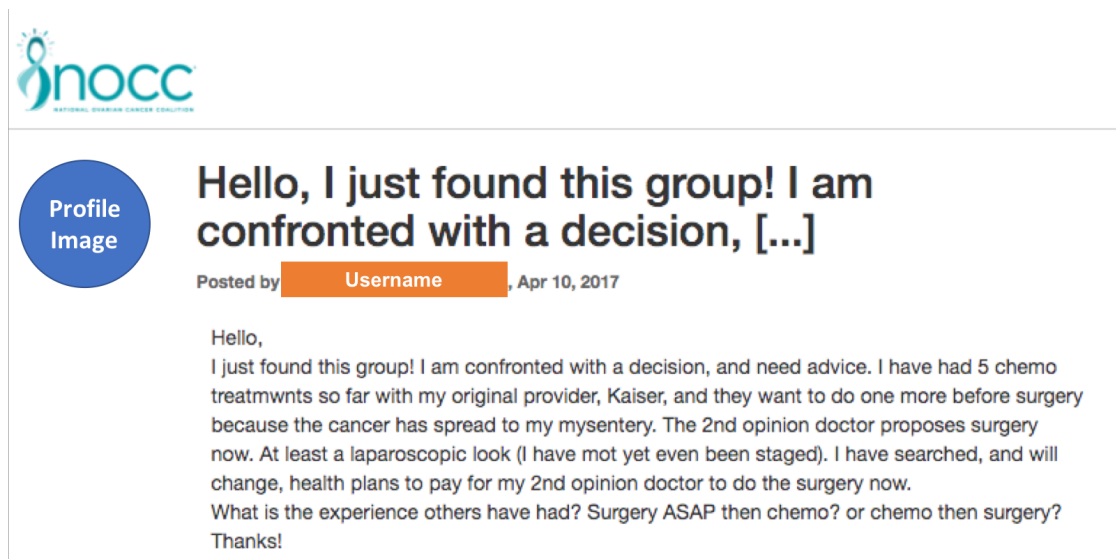



Figure 17: An Example of Initial Post

Figure 17 is an example of an initial post in a thread. The real username and the user's profile image was removed for privacy. Overall, Two categories of data were collected from a post: user-related data and post-related data. User-related data included username, number of posts. Post-related data are post title, poster username, date, URL, and content.



Figure 18 displays two comment examples replied to the post shown in Figure 17. For each comment, the collected data consisted of the replier's username, date of the comment, comment content, URL, and the post associated with the comment.

---



**Username** Apr 11, 2017


My first surgery consisted of hysterectomy, some lymph nodes and tumors removed. I then had 6 chemo treatments followed by a debulking surgery. After that was 6 more rounds of chemo.

[REPLY](#)



1 reply

---



**Username** Apr 11, 2017

Have you had a pet scan...this shows wherever the cancer is and they would know whether to do surgery first depending on the type of tumor, size etc. but sometimes they want to shrink the tumors first before going in. Ask questions of both doctors...why one wants to do it now and the other wants to wait...what are the disadvantages and advantages of each. Is the expertise of both doctors the same?

[REPLY](#)

Figure 18: Examples of Comments

## 6.2 Data Analysis

In order to identify the users' learning levels expressed in the posts and the different types of support exchanged in the comments, manual content analysis was first adopted as the method to establish different categories.

For the **initial posts**, manual annotations were performed to identify:

- The posters' *levels of learning objectives* if there was one or more explicit information needs articulated in the post (section 6.2.1);
- Ovarian cancer *topics* that the poster asked to learn (section 6.2.2);
- Two other characteristics of initial post, including *self-disclosure* and *role of the poster* (section 6.2.3).

The **comments** content was coded in terms of:

- *Types of support* provided by the commenters (section 6.2.4)

Automatic classification models were applied to predict different types of support in the comment following the manual coding due to the large number of comments.

### 6.2.1 Identifying Initial Posts with Information Needs and Levels of Learning

Though OHC is widely used to address information needs and acquire knowledge, not every consumer shares their post in OHC with the purpose of seeking information and learning knowledge. Therefore, to study consumers' search as learning in OHC, it is necessary to first distinguish posts with information needs and learning objectives from other posts.

Then, for the posts with information needs, furthermore, a coding framework guided by A & K's taxonomy (L. Anderson et al., 2001) was developed to classify the levels of learning objectives expected by the poster. Given the considerations stated in section 3.1.1, *Understand, Analyze, and Evaluate* were chosen from the six learning objectives from lower cognitive complexity to higher cognitive complexity. The description of each level and the de-identified example posts are displayed in Table 7. Note that to achieve the higher levels in the A & K's taxonomy, one must master the lower levels in the hierarchy. Therefore, the three levels of learning were coded mutually exclusively. For example:

“...Last one [CA score] showed 2 point increase and Dr wasn’t concerned as said basically save number 28 to 30. This was 1/22. Today it has went up .8. Any reason to be concerned since trend is upward?” (see complete post in Table 7)

In this post, the poster described her situation and sought judgment from peers. Therefore it was coded as the “Evaluate” level of learning objective. In addition, to make an effective judgment, she was required to “Understand” and “Analyze” the situation.

Two coders applied the coding framework to 100 sample posts out of 909 posts individually to 1) identify whether or not the post involved information needs and learning objectives; 2) if so, determine the level of learning that best described the consumer’s learning objective. One coder annotated the remaining posts after an acceptable agreement being achieved between two coders on the 100 sample posts (percent agreement = 79.4%, Cohen’s kappa = 0.72).

Table 7: Coding Framework of Levels of Learning Objectives in the Initial Post

| <b>Levels of Learning</b> | <b>Description</b>  | <b>Example</b>   |
|---------------------------|---|--|
| Understand                | Pursuit of facts, concepts, and ideas by describing, explaining, identifying, detailing, interpreting, or summarizing, etc.   | “Hi does anyone have information on AMG 386? Thank You”  |
| Analyze                   | Pursuit of connections and relationships among concepts under uncertainty by differentiating, comparing, distinguishing, contrasting, sorting, etc.                   | “I recently developed small red dots all over my legs, looks like little blood marks. I’m on Avastin and wonder if anyone has experienced these marks on their body?”  |
| Evaluate                  | Pursuit of decision or judgment given specific conditions by appraising, arguing, judging, selecting, critiquing, weighing, recommending, assessing, predicting, etc. | “Hi Sisters, I finished front line 12/8 and ca has be tested 3 times since. Last one showed 2 point increase and Dr wasn’t concerned as said basically save number 28 to 30. This was 1/22. Today it has went up .8. Any reason to be concerned since trend is upward? I’m concerned of this continuing and I’m already full of worry. I also work out every day with weights. Does intense exercise affect this with the temporary inflammation? Thanks!” |

## 6.2.2 Identifying Topics of Information Needs in the Initial Post

Table 8: Coding Framework of Topics of Information Needs in the Initial Post

| Topics               | Description  | Code |
|----------------------|--|------|
| Disease Management   | Information needs related to ovarian cancer disease management, such as diagnosis, prognosis, finding gynecologic oncologist, preparing for visit, advance care planning/advance directives, borderline malignant tumors, prophylactic surgery, secondary prevention, monitoring for recurrence, management of recurrence, supportive care/palliative care | DM   |
| Symptom Management   | Information needs related to ovarian cancer symptom management, such as fatigue, sleep, bowel, pain, neuropathy, cognitive memory, nausea, vomiting, bloating, ascites, appetite, appearance, shortness of breath, lymphedema, urinary, early menopause, ostomy management, rash, anemia, mouth sore, myelosuppression                                     | SM   |
| Treatment            | Information needs related to ovarian cancer treatment information, such as medications, surgery, radiation, chemotherapy, biologic therapies, clinical trials  | TM   |
| Treatment Decision   | Information needs related to ovarian cancer treatment decision.  | TD   |
| Emotional Management | Information needs related to emotional management, such as anxiety, depression, fear of recurrence, mood swings, coping, grief and loss  | EM   |
| Self-Management      | Information needs related to self-management, such as nutrition, spiritual support, physical activity, relationship with loved ones  | SF   |
| Practical Needs      | Information needs related to practical needs, such as finance, insurance, employment, legal, community resources   | PN   |
| Caregiving           | Information needs related to caregivers' needs, such as stress, caregiver coping, grief and loss   | CG   |
| Others               | Other ovarian cancer related information needs such as communication, sexuality, rehabilitation, complementary therapy/integrative medicine, ovarian cancer organization and facilities  | OT   |

To understand the information dimension of consumers' search as learning in OHC, topics of the information needs in the initial posts were also annotated through content analysis. The coding framework was developed inductively by a nurse practitioner by immersing herself in the posts. A coding framework with 13 topics was established initially.

Using the framework, two coders annotated all 909 posts individually, with one having a nursing degree. One initial post can include one or multiple topics. Later on, topics that appeared in less than 10 posts among 909 posts were further grouped into "Others." As a result, 9 codes were used to classify the topics of information needs in the initial post. Description of each topic is detailed in Table 8. An acceptable inter-rater agreement was obtained between the two coders, with an average percent agreement of 94.38% and Cohen's kappa coefficient of 0.717, ranging from 0.620 to 0.810 across 9 categories.

### **6.2.3 Identifying Other Characteristics of Information Needs in the Initial Post**

Self-disclosure and the role of the posters were annotated to further unveil the characteristics of the information needs in the initial posts.

Self-disclosure is the consumers' act of sharing personal information. Previous work found that self-disclosure is associated with the seek of support in the initial post (D. Yang, Yao, Seering, & Kraut, 2019; Y.-C. Wang, Kraut, & Levine, 2015). However, little is known about the relationship between self-disclosure and the levels of learning being sought in the post. Therefore, two coders annotated 100 sample posts in a binary fashion, with the question of whether there was any form of self-disclosure in the post. Sharing demographic information (e.g., age, location, family structure) and medical information (e.g., cancer trajectory, treatment, insurance) are both counted as self-disclosure. Since a good agreement was achieved between the two coders (percent agreement = 95.1%, Cohen's Kappa = 0.881), one coder coded the remaining posts.

The poster's role was elicited to discover the levels of learning of different types of consumers, i.e., patients and caregivers. The role of the poster is either explicitly stated or absent from the post. Since the annotation task is very straightforward, one coder annotated all the 909 posts using three codes: patient, caregiver, and unknown.

### **6.2.4 Identifying Support Types in the Comments**

The two most common types of social support that the consumers exchange in OHC are informational support and emotional support (Y.-C. Wang, Kraut, & Levine, 2012; X. Wang,

High, Wang, & Zhao, 2021). In this dissertation, as the interest is to investigate what information the consumers receive to fulfill their search as learning needs, the informational support provided in the comment was further classified by employing the framework proposed in Chuang and Yang (2014). Chuang and Yang (2014) identified 5 types of informational support:

- **Advice:** The comment offers ideas and suggestions actions for coping with challenges.
- **Referral:** The comment refers to information sources, e.g., books, websites, contacts.
- **Fact:** The comment offers facts or reassesses the situation.
- **Personal Experience:** The comment shares personal stories or incidents.
- **Opinion:** The comment Offers view or judgment formed about something. It is not necessarily based on fact or knowledge.

In addition, **Emotional Support** was also coded to assess if a comment provided empathy, encouragement, or appreciation (D. Yang, Kraut, et al., 2019).

All six types of support were coded in a binary fashion, and one comment can provide zero or one or multiple types of support. If no informational or emotional support was provided, the comment was coded as **Others**. For example, the comment below was coded as "1" for providing advice, fact, and opinion, and "0" for referral and emotional support:

*"Sounds difficult and i think you are taking the right approach. Stage III ovarian cancer is usually treatable with debulking surgery and chemo. You might consider getting a second opinion from a gynecologic oncologist at a large cancer center. They should have more experience in this type of situation and may have access to potentially helpful clinical trials."*

After screening out the comments of the posts without information needs, 3998 comments were remaining. The support types provided in these 3998 comments were identified through three steps. Firstly, two coders coded 150 sample comments to ensure the reliability of the coding framework. On average, an agreement rate of 94% percent agreement and 0.84 Cohen's Kappa was achieved, indicating an almost perfect agreement (McHugh, 2012). After addressing all the disagreements, one coder coded 350 more comments. Given the 500 annotated comments, BERT sequence classifiers were built to classify the six different types of support. The inter-rater agreement between the two coders and the performance of the

classification models were presented in Table 9. Finally, the models were applied to the remaining comments to predict the support types. As a result, after filtering out comments with no type of informational support or emotional support (i.e., those coded as "Others"), 3642 comments remained. Each comment provided at least one type of informational support or emotional support.

Table 9: Inter-rater Agreement between Human Annotators and Classification Score for Support Types in the Comments

| Support Type        | Inter-rater Agreement |               | Classification Score |             |             |
|---------------------|-----------------------|---------------|----------------------|-------------|-------------|
|                     | Percent Agreement     | Cohen's Kappa | Precision            | Recall      | F-score     |
| Advice              | 96%                   | 0.883         | 0.772                | 0.850       | 0.809       |
| Referral            | 98%                   | 0.935         | 0.818                | 1.000       | 0.900       |
| Fact                | 93.3%                 | 0.864         | 0.816                | 0.769       | 0.792       |
| Personal experience | 90%                   | 0.8           | 0.948                | 0.873       | 0.909       |
| Opinion             | 93.3%                 | 0.786         | 0.807                | 0.807       | 0.807       |
| Emotional support   | 91.3%                 | 0.823         | 0.910                | 0.739       | 0.816       |
| Others              | 95.3%                 | 0.762         | -                    | -           | -           |
| <b>Average</b>      | <b>94%</b>            | <b>0.84</b>   | <b>0.85</b>          | <b>0.84</b> | <b>0.84</b> |

### 6.3 Results

The results of NOCC forum analysis are elaborated in three subsections:

- Section 6.3.1 Levels of Learning in the Initial Posts: presents the prevalence and characteristics of posters' OvCa-related information needs with different levels of learning objectives (i.e., understand, analyze, evaluate) in NOCC.
- Section 6.3.2 Support for Different Levels of Learning in the Comments: reveals the types and the amounts of support exchanged in the comments for information



needs with different levels of learning objectives.

- Section 6.3.3 Knowledge Change: explores the change of individual users' levels of learning objectives in NOCC and how it relates to the current level of learning and the supports exchanged.

### 6.3.1 Levels of Learning in the Initial Posts

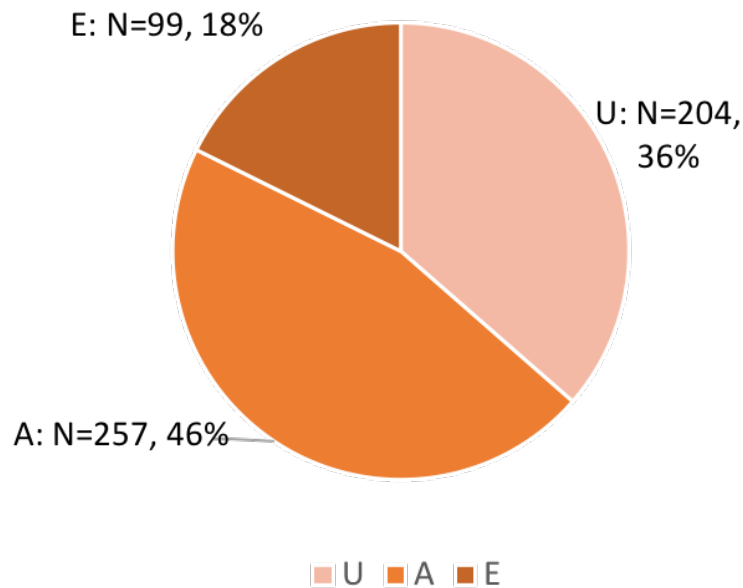


Figure 19: Distribution of Initial Posts with Different Levels of Learning Objective

U - Understand, A - Analyze, E - Evaluate

**6.3.1.1 Distribution of Different Levels of Learning** In total, 560 out of 909 initial posts were associated with learning objectives, as indicated by the explicit information needs expressed in the posts. 349 remained posts without any information needs were coded as "others," which mainly consisted of posts sharing personal updates, sharing resources, provoking discussions, providing inspirations, etc.

All the following results were based on the analysis of the 560 initial posts with identified learning levels, and 3142 comments replied to them.

The distribution of the posts with one of the three learning levels is presented in Figure 19. Among the 560 posts, Analyze was the most common learning objective, accounting for almost half of the total (46%). Following it, 36% of the posts with information needs sought the Understand level knowledge, whereas the Evaluate, the highest level of learning, only accounted for 18% of the information seeking posts. This result suggested that OvCa consumers mainly use the NOCC community to look for lower-level knowledge (i.e., facts, concepts, or relationships between facts and concepts) rather than higher-level knowledge (i.e., decisions and judgments).

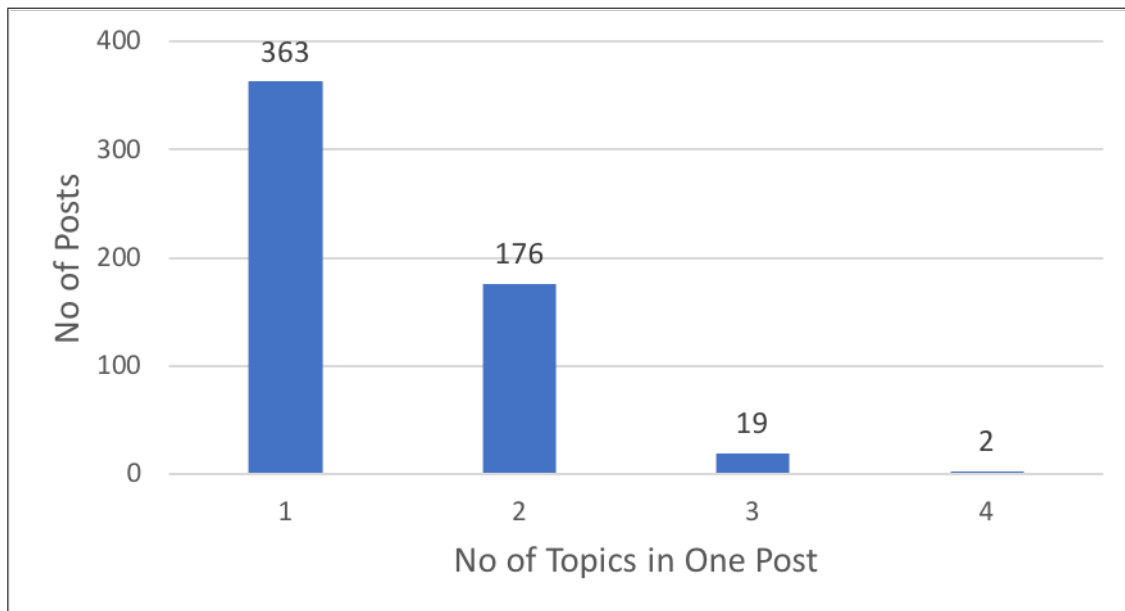


Figure 20: Distribution of Number of Topics per Post

**6.3.1.2 Number of Topics** The distribution of number of topics per post is plotted in Figure 20. Topics counted here only include topics mentioned in the consumers' information needs. In most initial posts with information needs, consumers tended to seek information and knowledge about one (N = 363, 64.82%) or two topics (N = 176, 31.43%). There were only two posts in which consumers consulted peers on four different topics.

Then, the initial posts were grouped by the three levels of learning objectives, and the average number of topics in each group was calculated (Table 10). A one-way between-

subjects ANOVA was performed on the number of topics in one post as a function of the level of learning objective. With violation of the assumption of homogeneity of variance, an F-test with Brown-Forsythe adjustment was conducted. Results suggested a statistically significant difference in the number of topics among different levels of learning goals ( $F(2, 193.364) = 72.537, p = .000$ ). A Games-Howell post hoc test revealed that there were significantly more topics in the posts with the Evaluate learning objective ( $N = 1.828, p = .000$ ) than posts with the Analyze learning objective ( $N = 1.498, p = .000$ ). The posts with the Understand learning objective consisted of the least number of topics compared to the two higher levels ( $N = 1.049, p = .000$ ).

The difference in the number of topics might indicate that to obtain the Evaluate level knowledge, consumers tend to acquire information across multiple topics. In contrast, for lower levels of learning, their information needs were more likely to focus on one specific topic. The results in the next section 6.3.1.3 elaborate on what topics the consumers were interested in for each level of learning.

Table 10: Number of Topics of Each Learning Level

| Levels of Learning | # Topics per Post (Mean) | # Posts |
|--------------------|--------------------------|---------|
| Understand         | 1.049                    | 204     |
| Analyze            | 1.498                    | 257     |
| Evaluate           | 1.828                    | 99      |
| Total              | 1.392                    | 560     |

**6.3.1.3 Category of Topics** Using the coding framework in Table 8, consumers' information needs in the initial posts were classified into 9 different categories based on cancer-related topics. In this section, two results of topic categories are presented. Firstly, topics were grouped by different levels of learning to show what information the consumers needed for each learning level. Then, for the posts with more than one topic, the frequencies of all the topic pairs were examined to further demonstrate what topics were likely to be needed together by the consumers.

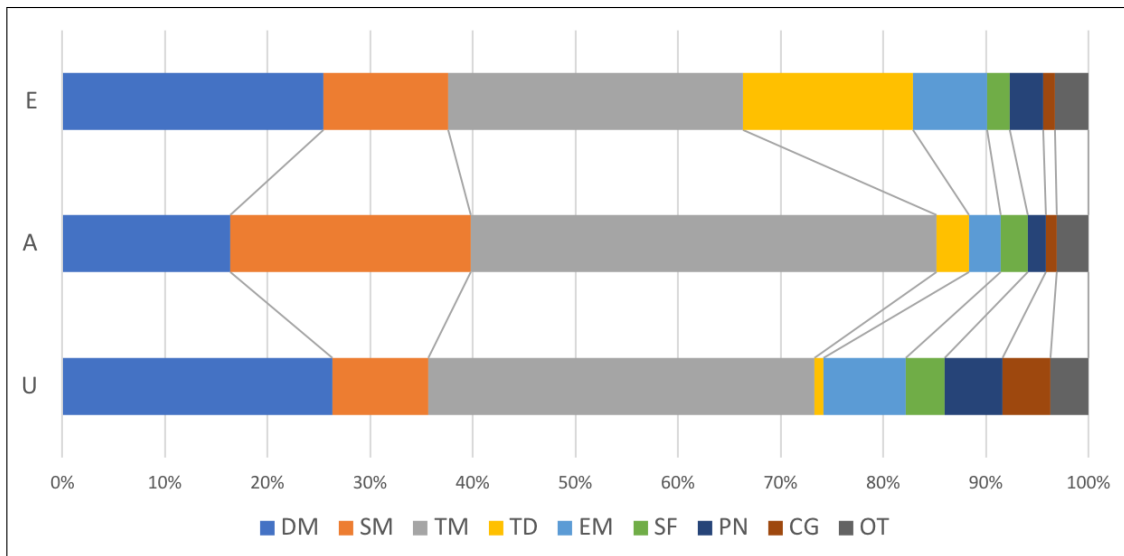


Figure 21: Distribution of Topics of Interest in Each Learning Level.  
 DM-Disease Management; SM-Symptom Management; TM-Treatment;  
 TD-Treatment Decision; EM-Emotion Management; SF-Self-management;  
 PN-Practical Needs; CG-Caregiving; OT-Others

Figure 21 shows the distribution of nine ovarian cancer topics in the posts with information needs. Each bar represents posts of one of the three learning levels, whereas a segments in a bar denote the portion of a topic among all the posts with the same level of learning. Segments of the same color are comparable.

It is evident that treatment is the most popular topic of interest in all three learning levels, with a higher portion in the Analyze level ( $N = 175, 45.45\%$ ) than the other two levels. This result indicated that comparing or differentiating treatment information under uncertainty was a common need among OvCa consumers in OHC. In addition, pursuing treatment information to understand or evaluate was also frequent, which might because the treatment information of ovarian cancer was complex and scattered, making the topic of treatment the dominant information needs across all the learning levels. Analyzing symptom management is the second prevalent information need, while understanding and evaluating symptom management information are not that popular. The result suggested that, for symptom management, the consumers struggle more with the differentiation or connection among different symptoms than learning about basic symptom information or making judgments.

On the contrary, disease management was more associated with the Understand and the Evaluate learning levels than the Analyze level, implying that consumers needed support for interpreting disease information such as diagnosis, prognosis, recurrence on both a basic fact or concept level and a high judgment or decision level. Besides, it is notable that treatment decisions accounted for a significant portion in the Evaluate level. However, it is questionable if consumers should use OHC as a resource for making treatment-related decisions. Emotional management and practical needs present similar patterns across the three levels of learning: proportions in the Understand and Evaluate level were higher than Analyze level. Caregiving information accounted for a much more share in the Understand level than the other two. Lastly, the ratios of other topics were very similar in all three levels of learning.

Chi-square results revealed that there was a significant association between levels of learning objectives and topics of DM ( $\chi^2(2) = 17.196, p < .001$ ), SM ( $\chi^2(2) = 40.260, p < .001$ ), TM ( $\chi^2(2) = 38.548, p < .001$ ), TD ( $\chi^2(2) = 85.827, p < .001$ ), EM ( $\chi^2(2) = 7.698, p < .05$ ). However, no significant association was found between levels of learning

objectives and topics of SF ( $\chi^2(2) = .004, p > .05$ ), PN ( $\chi^2(2) = .3.379, p > .05$ ), CG ( $\chi^2(2) = .4.89, p > .087$ ), and Others ( $\chi^2(2) = .690, p > .05$ ).

|    | DM     | SM     | TM     | TD    | EM    | SF    | PN    | CG    | OT |
|----|--------|--------|--------|-------|-------|-------|-------|-------|----|
| DM |        |        |        |       |       |       |       |       |    |
| SM | 6.12%  |        |        |       |       |       |       |       |    |
| TM | 13.06% | 29.39% |        |       |       |       |       |       |    |
| TD | 3.27%  | 0.00%  | 11.43% |       |       |       |       |       |    |
| EM | 4.49%  | 3.27%  | 4.49%  | 0.41% |       |       |       |       |    |
| SF | 1.22%  | 2.45%  | 1.63%  | 0.00% | 0.82% |       |       |       |    |
| PN | 0.41%  | 1.22%  | 2.45%  | 0.41% | 0.41% | 0.41% |       |       |    |
| CG | 0.82%  | 0.00%  | 1.22%  | 0.41% | 0.41% | 0.41% | 0.41% |       |    |
| OT | 2.04%  | 2.04%  | 2.86%  | 0.82% | 0.41% | 0.00% | 0.41% | 0.41% |    |

Figure 22: Co-occurrence of Topic Pairs in One Post.  
(N.B., darker color indicates larger proportions)

Figure 22 shows the proportion of different topic pairs among all the 245 topic pairs extracted from the posts with more than one topic. Notably, treatment (TM) and Symptom Management (SM) were most likely to appear together in one post (29.39%). Besides, consumers' tended to learn treatment along with disease management (DM) or treatment decisions (TD).

**6.3.1.4 Self-disclosure** The distribution of posts with or without self-disclosure grouped by learning levels was displayed in Figure 23. The number of posts without self-disclosure was similar in the Understand and the Analyze level. In the Evaluate level, however, there were only two posts without any kind of self-disclosure information. Results suggested that consumers tended to disclose some of their personal information when seeking information, especially in the learning of judgments or decisions from peers.

**6.3.1.5 Role of Posters** The roles of the consumers who posted information needs in the NOCC community were sorted by the learning levels and plotted in Figure 24 after excluding those with the unknown role. Most of the posts were posted by patients regardless of their learning level. Caregivers are also expected to learn information across all three levels.

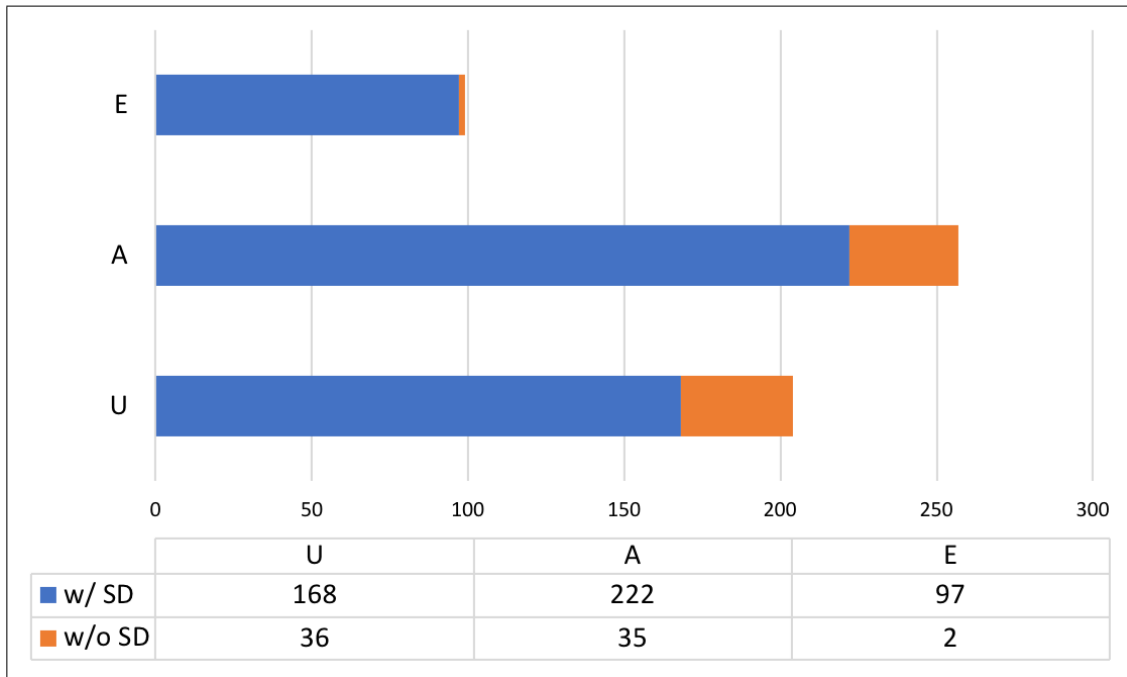


Figure 23: Distribution of Posts with or without Self-disclosure in Each Learning Level  
 (NB. w/SD - post with self-disclosure, w/o SD - post without self-disclosure)

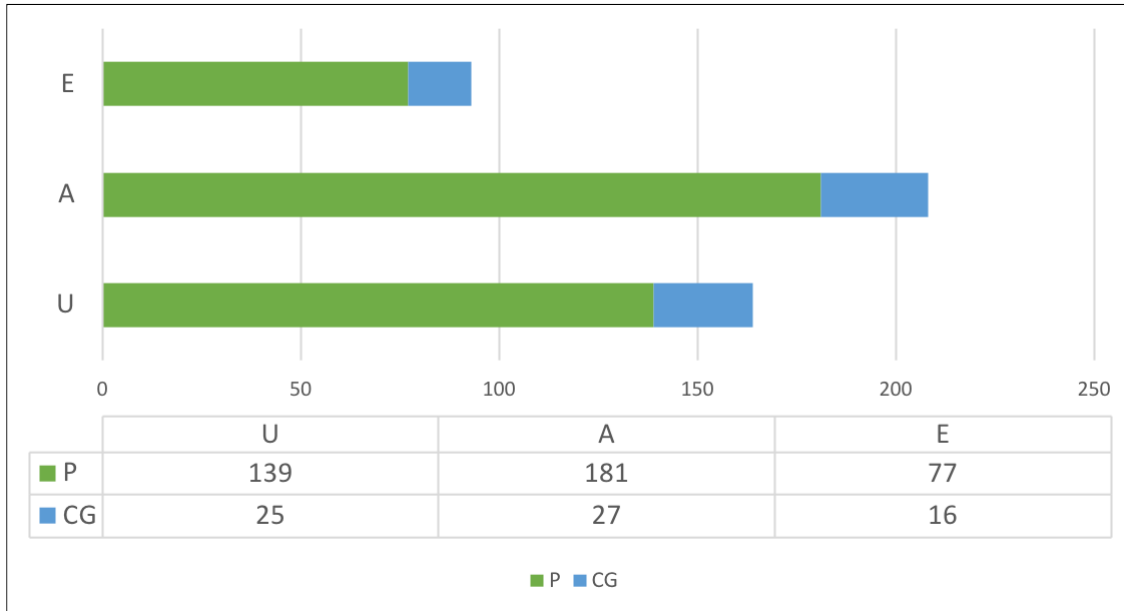


Figure 24: Distributions of Roles of the Posters in Each Learning Level.  
 (NB. P - patient, CG - caregiver)

### 6.3.2 Support for Different Levels of Learning in the Comments

**6.3.2.1 Number of Comments** The 3642 comments to the 560 initial posts with information needs were grouped by the level of learning in the post. Figure 25 shows the average number of comments of one post in each learning level. As displayed in the figure, posts with the Understand level of learning objective were likely to receive the largest number of comments from peers ( $N = 7.68$ ), followed by the Evaluate level ( $N = 7.07$ ) and the Analyze level ( $N = 5.63$ ). However, results from one-way ANOVA suggested no statistically significant difference between the average number of comments among the three learning levels ( $F(2, 451.295) = 2.712, p = .067 > 0.05$ ).

**6.3.2.2 Types of Support** As the types of support provided in one comment were coded in a non-mutually exclusive approach, which means one comment could provide one or multiple types of support among advice, referral, fact, personal experience, and emo-



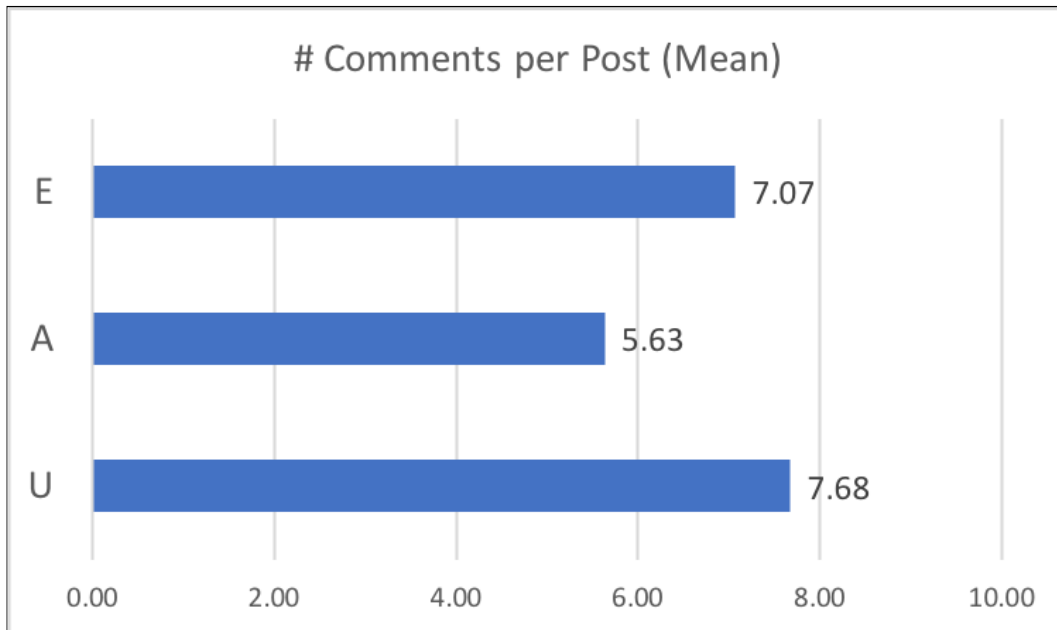


Figure 25: Average Number of Comments per Post in Each Learning Level

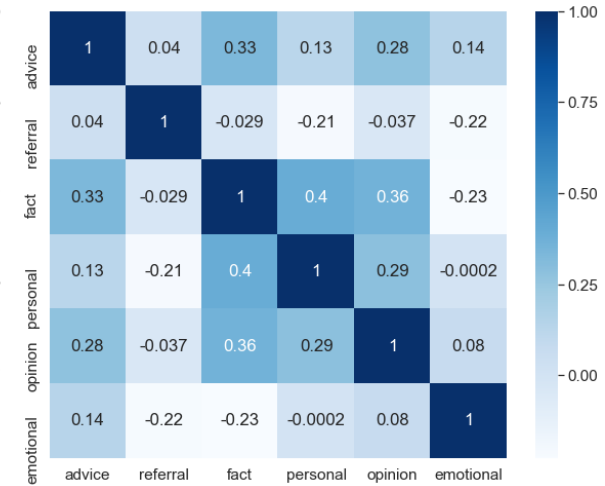
tional support, the correlation coefficient of each pair of supports was calculated to show the relationship among different types of support. Phi-coefficient was used to measure the correlation coefficient as types of support were coded as binary (i.e., 0 or 1). The correlation matrix could help to reveal which pair of supports was likely to be provided together or not in one comment. Figure 26 (a) shows the correlation matrix created based on all the 5642 comments provided to support 560 posts with learning objectives, while Figure 26 (b), (c), (d) were computed based on comments that were replied to support one of the three learning levels, i.e., understand, analyze, and evaluate, respectively.

The four correlation matrixes present similar patterns with some slight differences in the values, inferring that how each type of support related to another in one comment is independent of the learning level in the initial post. The following interpretation of the correlation matrixes takes the coefficients in Figure 26 (a) as examples, while the findings of the relationship between support types are applicable to the other three matrixes as well.

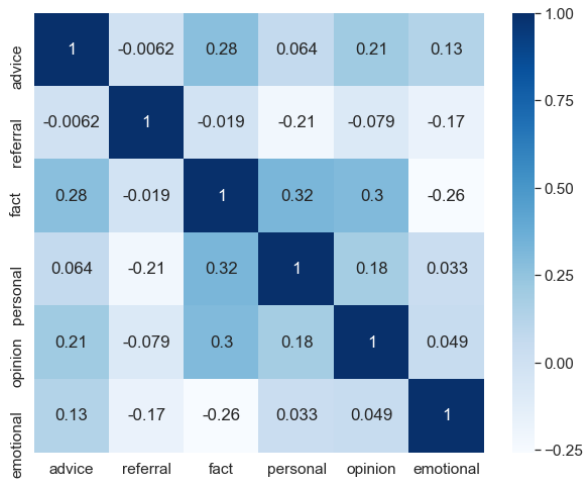
First, there is a moderate positive relationship between Advice and Fact ( $r = 0.32$ ), and a weak positive relationship between Advice and Opinion ( $r = 0.25$ ). This might because



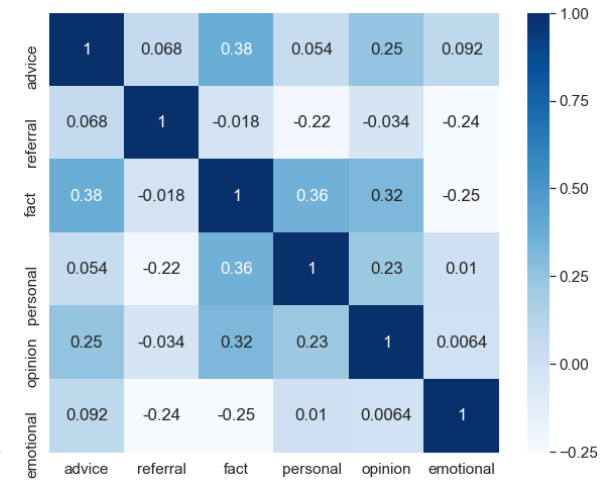
(a) All Comments to Support Learning



(b) Comments to Support Understand



(c) Comments to Support Analyze



(d) Comments to Support Evaluate

Figure 26: Correlation Matrix of Different Types of Support in One Comment (Advice, Referral, Fact, Personal Experience, Opinion, Emotional Support)

when commenters provide advice to posters in OHC, they tend to bring up some facts or express their opinions at the same time. Other than Advice, Fact also has positive moderate relationships with Personal Experience ( $r = 0.36$ ) and Opinion ( $r = 0.33$ ). Besides, a moderate positive correlation is also found between Personal Experience and Opinion ( $r = 0.24$ ). Given the moderate correlation between any two of the three, it can be inferred that when commenters share their personal experience with the posters, they might offer other informational support, i.e., fact and opinion. Interestingly, the relationship between Referral and Personal Experience ( $r = -0.21$ ), and that between Referral and Emotional Support ( $r = -0.20$ ) are both negative and weak, indicating that referral to sources is unlikely to be provided together with commenters' personal experiences or the emotional support. Additionally, other than the weak negative relationships between Emotional Support and Fact ( $r = -0.24$ ), and that between Emotional Support and Referral ( $r = -0.20$ ), no relationships between Emotional Support and other types of information support is found.

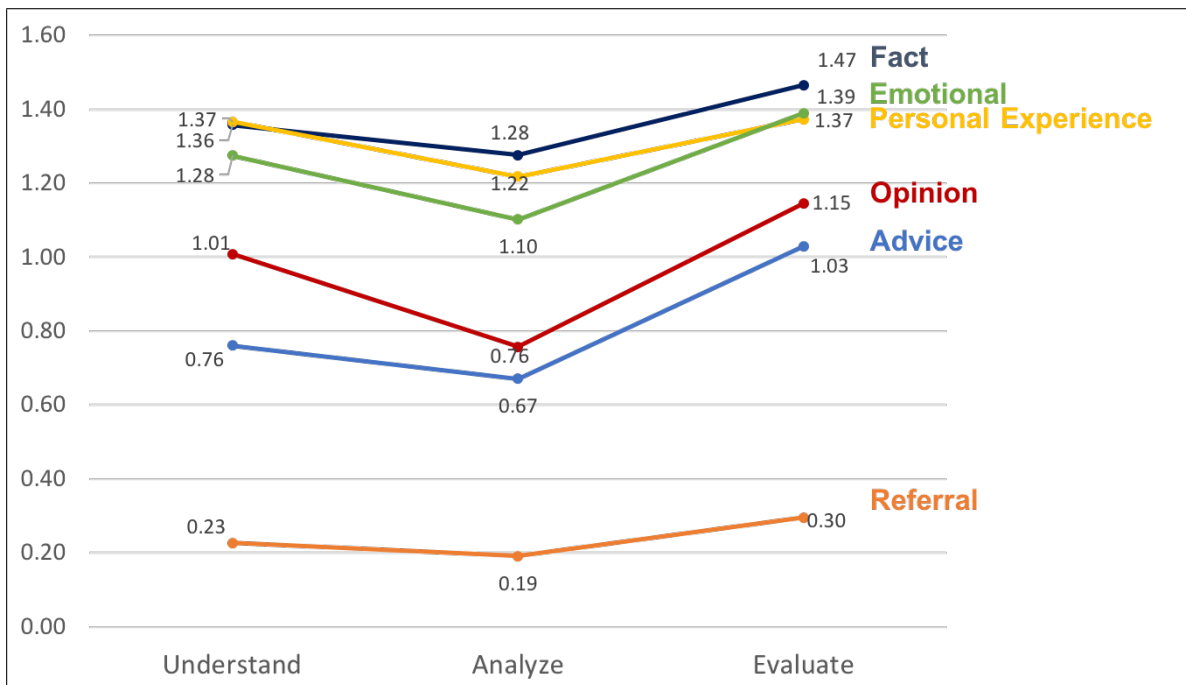


Figure 27: Amount of Different Types of Support Received in Each Learning Level

After examining the types of support provided by commenters in each comment, the

amount of support aggregated by posts was also investigated. Figure 27 shows the amount of different types of support received in one post belonging to each learning level. For different learning levels, log transformation is applied to the total number of each type of comment and plotted in the line chart.

In general, the total amount of support is the largest in the Evaluate level, followed by the Understand level, and it is the smallest in the Analyze level. As determined by one-way ANOVA, significant differences among the three levels of learning objectives were found in the Advice support ( $F(2, 437.84) = 9.687, p = .000$ ), Opinion support ( $F(2, 418.18) = 11.563, p = .000$ ), and Emotional support ( $F(2, 395.88) = 3.236, p = .014$ ). A Games-howell post hoc test revealed that posts with Analyze level learning goal received significantly less Opinion support compared to Understand ( $p = .002$ ) and Evaluate level ( $p = .000$ ). The amount of Advice support in the Evaluate level is significantly more than that in the Analyze level ( $p = .000$ ) and Understand level ( $p = .001$ ). For Emotional support, the significant result was only found between the Analyze level and the Evaluate level ( $p = .015$ ).

### 6.3.3 Knowledge Change

As some users posted more than one post with information needs and learning objectives in the forum, it allowed the researcher to unveil how the individual users' levels of learning manifested by the information needs change over time. 344 distinct users posted the 560 initial posts with information needs. Figure 28 depicts the distribution of the number of posts with information needs per user. As shown in the figure, most users ( $N = 244, 70.93\%$ ) posted only one post, while 100 users posted multiple posts. These 100 users who posted at least two posts with learning goals were further examined to uncover the change of levels of learning.

A change of levels of learning is defined as the transition between level of learning objective in post  $p_i$  and post  $p_{i+1}$  of the same user  $u$ . The changes of levels of learning were classified into three categories, based on the transition from  $p_i$  to post  $p_{i+1}$ : **Knowledge Increase**, **Knowledge Decrease**, and **No Change**. For example, if a user posted three initial posts (i.e.,  $p_1, p_2$ , and  $p_3$ ) in the NOCC forum, and the level of learning in them are  $p_1$

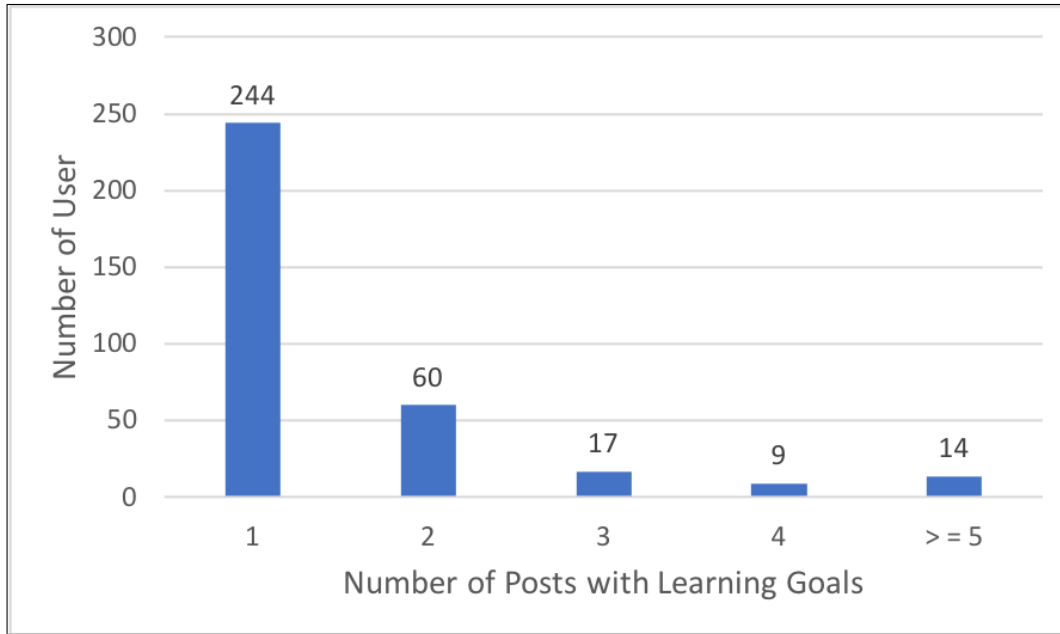


Figure 28: Distribution of Number of Posts with Information Needs per User

- Understand,  $p_2$  - Analyze, and  $p_3$  - Analyze, then the change of levels of learning from  $p_1$  to  $p_2$  is knowledge increase, and the change from  $p_2$  to  $p_3$  is no change. In total, 216 changes of levels of learning were identified from the 100 users who contributed multiple posts in the NOCC forum.

**6.3.3.1 Current Level and Knowledge Change in the Subsequent Post** The 216 changes of levels of learning were grouped by the three categories of change type, and the distribution was plotted in Figure 29. In general, 41.7% (N=90) of the pairs of two consecutive posts sought information on the same level of learning, which resulted in no change in the learning levels. Knowledge increase, in which the learning level in the subsequent post is higher than the prior one, is the second frequent change of levels of learning (N=70, 32.4%). The least frequent type of change is knowledge decrease (N=56, 25.9%). It can be inferred that users were more likely to increase or remain at the same level of learning objective as they continued posting and seeking information in the same forum.

Then the specific types of transition from different levels (e.g., from understand to un-

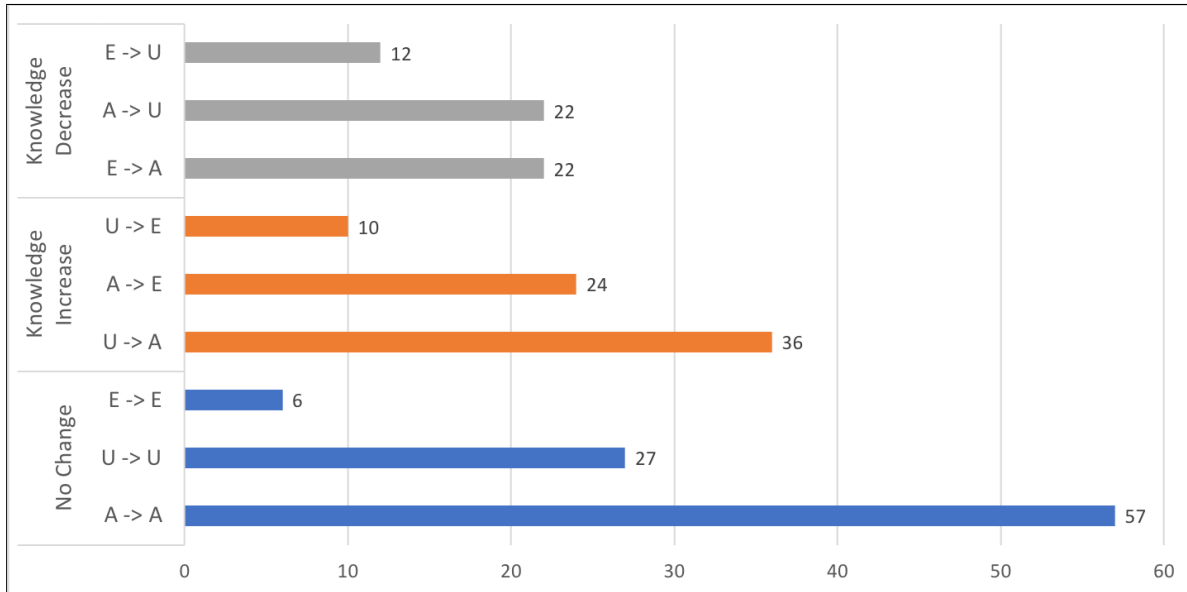


Figure 29: Change of Learning Objectives

U- Understand, A- Analyze, E- Evaluate

derstand) helped reveal how the current level of learning, which meant the level in the prior post, affected the subsequent post.

First, from the Analyze to Analyze ( $A \rightarrow A, N = 57, 26.4\%$ ) is the most common transition. Its amount and ratio are also higher than that from Analyze to Understand ( $A \rightarrow U, N = 22, 10.2\%$ ), and Analyze to Evaluate ( $A \rightarrow E, N = 24, 11.2\%$ ), suggesting that Analyze level question were likely to be followed by another Analyze question than the increase or decrease of levels of learning objective of the same user.

Second, after asking an Understand level question, users tend to increase the level of learning and ask an Analyze level question ( $U \rightarrow A, N = 36, 16.7\%$ ). The possibility is higher than asking another Understand level question ( $U \rightarrow U, N = 27, 12.5\%$ ) or Evaluate question ( $U \rightarrow E, N = 10, 4.6\%$ ). This might be attributed to the fact that the Understand level of learning was relatively easy to achieve, or the users' OvCa-related knowledge might evolve and increase over the course of time, driving them to pursue a higher level of learning.

Third, the Evaluate level posts were mainly followed by Analyze level post ( $E \rightarrow A, N =$

22, 10.2%) or Understand posts ( $E \rightarrow U, N = 12, 5.6\%$ ). Only rarely would the users ask another Evaluate level question ( $E \rightarrow E, N = 6, 2.8\%$ ). Additionally, users were more likely to increase learning objectives by one level (e.g.,  $U \rightarrow A, \rightarrow E$ ) or decrease by one level (e.g.,  $E \rightarrow A, A \rightarrow U$ ) in two consecutive posts than increase or decrease by two levels (e.g.,  $U \rightarrow E, E \rightarrow U$ ), indicating that change of learning level was a graduate evolving process.

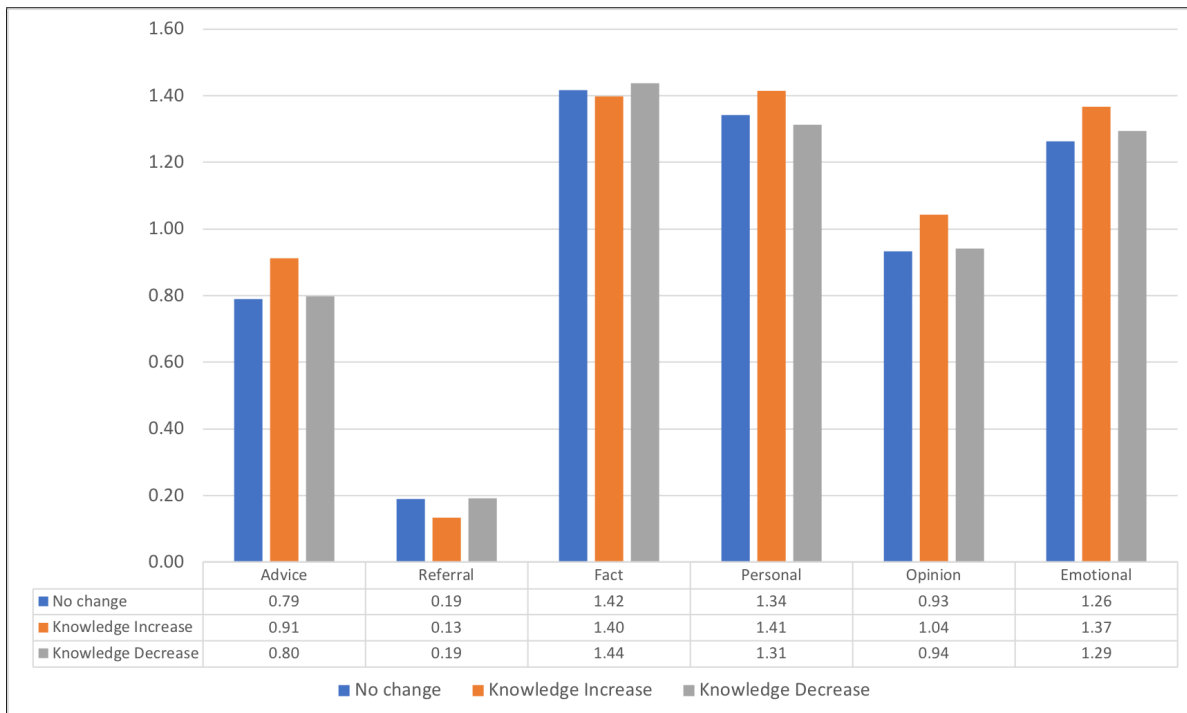


Figure 30: Support Received and the Change of Learning Objectives

**6.3.3.2 Support Received and Knowledge Change** Next, how the types and the amounts of support received for  $p_i$  influenced the levels of learning objectives in  $p_{i+1}$  is explored. As depicted in Figure 30, on average, for most types of support, when users received more support in amount, including Advice, Personal Experience, Opinion, and Emotional Support, they were more likely to increase the learning objective in the next post, rather than decrease or keep the same level of learning objective. Note that no statistically significant differences were found among the three types of change.

## 6.4 Summary of Study 2

Study 2 reveals the information dimension of health consumers' search as learning process in OHIS through the analysis of a well-known online health community for ovarian cancer. The results demonstrate that consumers use online health communities to address information needs with different levels of learning objectives, and at the same time, they can acquire various types of informational support in the comments from peers. Receiving more information support drives the consumers to seek information with higher levels of learning objectives.

Based on the underlying learning objective, 560 initial posts with explicit information needs were annotated and mapped to one of the three learning levels based on the underlying learning objective associated with the information needs. The coding framework and example posts were presented in Table 7. The results suggest that ovarian cancer patients and caregivers are more likely to use OHC to address lower-level information needs. The information needs with different learning levels are also different in the frequency of self-disclosure behaviors and topics of interest in the questions.

Regarding the support received by the consumers for information needs with different learning levels, it is found that the largest number of supports is provided for the Understand level information needs, followed by the Evaluate level. Generally, more advice, opinion, and emotional supports were obtained for the information needs with the Evaluate level of learning. Consumers' knowledge state was measured unobtrusively by annotating the change of learning levels in the posts from the same user. The results reveal that consumers' knowledge change varies by the amount of support received.



## 7.0 Study 3: Laypeople's Outcomes of OHIS

This chapter focuses on the interview findings of laypeople's OHIS and outcomes in their real life. This semi-structured interview was a follow-up study of the user study reported in Chapter 5. Chapter 5 presents the user study examining consumers' search as learning behaviors in completing different pre-designed OHIS tasks, while this interview study complements Study 1 because the information needs described by the lay participants in this interview were their real needs rather than assigned ones in Study 1.

### 7.1 Data Collection

#### 7.1.1 Setting and Sampling

In total, 24 participants (Female: 15 and Males: 9; Undergraduates: 17 and Graduates: 7; Age: 18-33), who were recruited via flyers from the University of Pittsburgh and Carnegie Mellon University, were interviewed. The details of the participant recruitment and screening have been presented in section 5.1.3. The participants were considered as laypeople, because they had no health- or medical-related background. Besides, the interview results suggested that most of them ( $N = 20$ ) were unfamiliar with the health conditions they searched for.

Each participant came for approximate two hours for the whole study, and the interview accounted for about 30 minutes. All the interviews were audio-taped. Each participant was compensated \$20, and the interview design was approved by the Human Research Protection Office of the University of Pittsburgh.

#### 7.1.2 Interview Design

The interviews were guided by a semi-structured interview protocol (Appendix C). The protocol mainly consists of two sections. In the first section, the participants were asked to express their perspectives on OHIS, centering around *How do you decide whether or not to*

*select an online health information source, and why?*

After these conversations, the interview processed to elicit the participants' real OHIS experiences based on the principles of Critical Incident Technique (CIT) (Flanagan, 1954). The CIT approach directly asks the participants to recall the context and the task, which can recover a more realistic process. For decades, CIT has been actively used in the information behavior studies, including the research on health information seeking behaviors (Urquhart et al., 2003; Marcella, Rowlands, & Baxter, 2013; T. Jiang, Liu, & Chi, 2015; Courtright, 2005). CIT was adopted because it allowed the researchers to analyze the OHIS process in the context of participants' real OHIS tasks. Since the topic was health issues, the participants were first informed that they could choose not to tell the exact health condition if they felt uncomfortable to share, and the focus of the study was their search process. If they were glad to share, they were prompted to start from their most impressive or most recent OHIS experience.

Participants gave responses by recalling the most recent or most impressive OHIS experience. To reveal the details in the whole OHIS process, a list of questions and prompts were designed that would help the participants to recall and unfold the specific incidents from memory. In general, the conversation was mainly led by the following questions: *“What motivated your OHIS? Why did you want to search for it? How did you search for it? How did you start your search? Could you recall what kind of websites you visited? Were they helpful? Why or why not? How did you feel before and after the search or during the search? Why?”* and *“Did your OHIS process involve any decision-making activities? If so, what decision did you make after seeking the information? Did you take any action? Why or why not?”*

## **7.2 Data Analysis**

All 24 interview audio clips were transcribed, and the data analysis was conducted by the author and three other researchers through an iterative thematic content analysis process (Hsieh & Shannon, 2005) which was consisted of three main phases. In the first phase,

the research team worked together to develop an initial codebook. The initial themes and codes were mainly originated from two sources: the semi-structured interview protocol and the results of the first-round open coding. As described in the last section, the interview protocol has already composed some themes. As for the first-round open coding, three researchers coded individually on two randomly chosen transcripts and generated some new themes. The team discussed the themes and the subcategories from the two sources and established a revised codebook.

In the second phase, three researchers coded individually the same two transcripts again using the revised codebook, and reached an average pairwise percent agreement of 0.891 and average pairwise Cohen's Kappa of 0.583, indicating a reliable agreement (Viera, Garrett, et al., 2005). After discuss and address the disagreements, the code team then applied the codebook to the remaining 22 transcripts in the third phase. As a result, each transcript was coded by two of the three coders. The percentage agreement was 0.879, and Cohen's kappa coefficient was 0.539. Throughout the process, the team met weekly to discuss problems and address disagreements, and the whole analysis process lasted approximately two months.

### **7.3 Results: Characterize Laypeople's Own OHIS Story**

This section presents the results of analyzing the participants' own search incidents. Due to various needs, goals, or search strategies that laypeople may have, their descriptions vary greatly. However, four important aspects that commonly existed in each story were identified. The first two are *I. information needs* and *II. searching starting-point*, and they are focused on the OHIS behaviors. A consumer's search is always triggered by a particular health-related information need, and the consumer would have to pick up an online website to start the search.

The following two aspects are *III. emotion changes* and *IV. follow up decisions*, which occur offline as an outcome of the OHIS. The results will focus on the analysis of the four aspects and the transitions between them. As shown in Figure 31, 24 valid incidents collected from the participants' real search experiences are visualized with a Sankey diagram (Csala,

2014). A set of subcategories is generated for each aspect, and the width of the flow branches is proportional to the number of incidents.

### 7.3.1 Information Needs

The most common type of information needs that motivated the participants to conduct OHIS is to self-diagnose (N=10). It usually happened when the participant found a suspicious symptom and was eager to check out what exactly it was. This needs usually come together with high uncertainty. For example, S06 described: *“I started feeling itchy in my throat, so I just looked up, what’s going on? Is it a virus going on?”* In another example, S23 recalled that: *“I felt something in my rib popped, and it was sort of have a shape that I could see it in the mirror, so I searched online. . .”*

The second most frequent type of information needs is to learn about a health issue (N=7). In these cases, the information need may be aroused from prior activities on the Internet such as browsing, searching or social interactions, during which, the participants encountered some stimuli regarding health issues and then conducted the following OHIS. For example, S04 recalled the reason that she searched for health information was: *“there was a case which I had read in the paper, online paper, where it was given why exactly the problems happen, why the girls don’t have the regular cycles and all.”* The desire to learn about a health issue might also happen when the participant wanted to search for someone else, such as their relatives or friends. For example, S16 shared an incident: *“one of my friends...he was actually diagnosed with clinical depression. Because I was concerned for him, I would just do research on symptoms, like how I can help without being a burden and stuff like that.”*

When a participant had a pre-existing condition, one would search to relieve a condition (N=4). As the participants in this study were all young adults, none of them reported any incident relates to severe chronic conditions, but four participants reported their searches for pre-existing conditions. For example, two participants shared how they struggled against overweight, and they searched online to learn how to lose weight. S20 confirmed the cost-free advantage of OHIS: *“...the internet definitely helped me a lot like how to lose weight.*

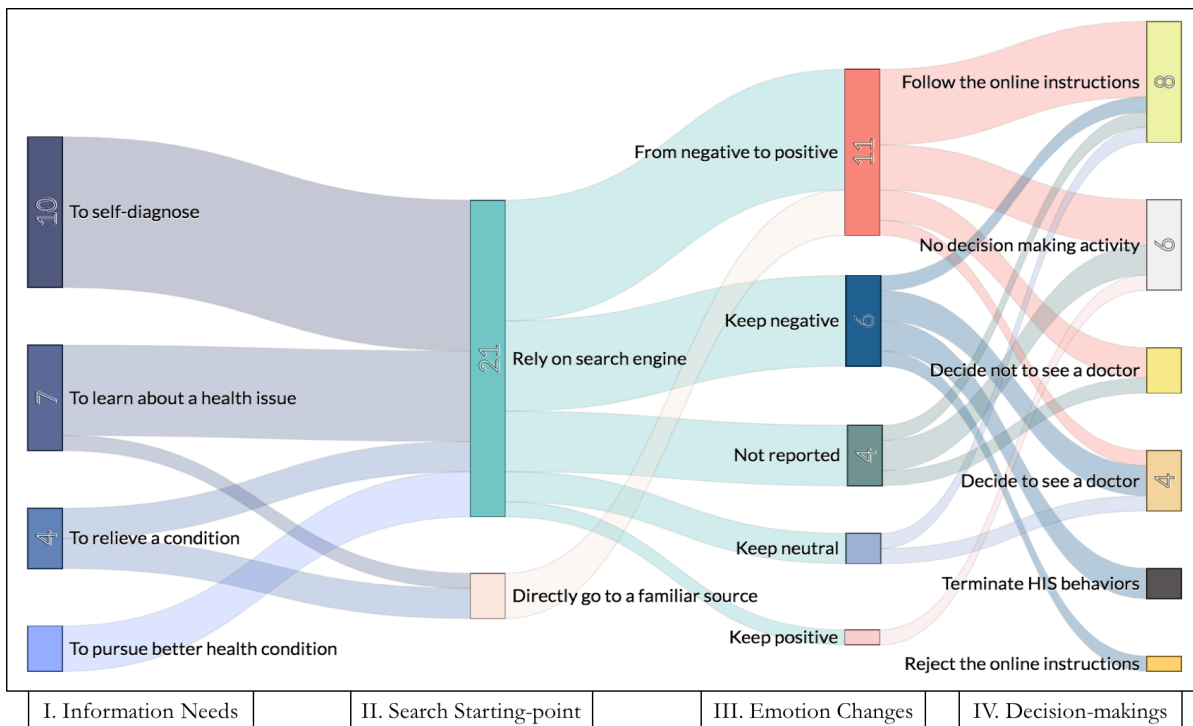


Figure 31: Four Important Aspects of OHIS Process and the Transitions Between the Adjacent Aspects

*Because I don't really go to a doctor or a physical trainer because I don't want to pay money for that..."*

In some cases, the participants, especially those who were currently healthy, might conduct OHIS to pursue better health conditions (N=3), such as reconstructing the workout plan, or searching for a new diet, etc. This result confirmed that wellness and nutrition information is a real information need of lay consumers.

### **7.3.2 Search Starting-points**

The participants were requested to describe more about how they started the OHIS process in their own search story. It is found that the majority just relied on the search engines (N=21), and only a few directly went to familiar sources (N=3). Such finding justified the design of the user study (Study 1), in which search engine interface was displayed to the participants at the beginning of the search.

As for the search strategy with general-purpose search engines, S16 shared a common experience. She described that she just started from general keywords, and further refined the search with the suggested search queries. When she examined the search results, she preferred the sources that present the same information: *"I would just use the keywords and a lot of it when I typed it in, there would be suggested queries. Yeah, so then I would use those. Usually, the good is like the most used search engines are at the top when you search for something. So, I would usually click on those first and get the general gist and then click the other ones to see what kind of common information would be on it."* It is notable that the participants' clicks of the search results were affected by the ranking positions. A similar case was reported by S11: *"Probably, like what the search engine brings up... like the first ten that come up. I will look at those ten and pick whichever one is most specific to the question that I'm asking."*

In only three incidents, the participants claimed they would directly go to their familiar online health information sources, and two of them (i.e., S02 & S20) conducted the search to relieve a pre-existing condition (see Figure 31). S02 used Mayo Clinic to learn about what to do to relieve her sciatic pain, and S20 reported using Reddit and other similar online forums

to deal with overweight. So, different from other consumers, those who conducted OHIS to relieve a condition were more likely to have prior knowledge about online health-related sources, and preferred to directly consult and trust their familiar sources. While for other lay consumers, they tended to simply start from search engines and filter the sources with the help of search result rankings.

### 7.3.3 Emotion Changes

The participants were asked to recall how they felt before and after they conducted the OHIS respectively. The comments about their feelings were categorized into five categories of emotion change outcomes: 1) changing from negative to positive (N=11); 2) keeping negative (N=6); 3) keeping neutral (N=2); 4) keeping positive (N=1); and 5) being not sure about the feeling before and after the search, which is coded as “not reported” (N=4). None of the participants reported feeling changing from positive to negative.

When searching for health information, especially searching to relieve a condition or to self-diagnose, it is very common that the consumers start the OHIS with a negative feeling. Unsurprisingly, 17 out of 24 participants expressed negative feelings before they conducted the OHIS, such as worried or anxious. Among these 17 participants, 11 reported their emotions changed to positive after searching online, while 6 remained feeling negative. This indicated that online information made them feel relieved or less worried about the condition. For example, S02 recalled: *“Before the search, I was kind of like, ‘Oh, I’m in trouble, I need to see the doctor but the doctor is not available right now, and I want to go to bed but I can’t because I’m in pain.’ Then after, I was like, ‘I’m still in pain, but maybe this is a way to help improve it for the future.’ ”*

Another reason that the participants felt better after OHIS was that they knew that their condition was common. For example, S05 stated that *“Before I was scared just to my knowledge of what I was experiencing but if you read more about how common it is online then I feel a little bit better”*. Sometimes, the consumers would also feel positive as they feel *“more knowledgeable”*(S09), or *“more confident”*(S20).

For those who keep feeling negative, one common reason was that they were not satisfied

with the information. The frustration could come from the failure to understand the information. For example, S17 commented that *“I was just frustrated, also not really understood the information given to me. All these, like the words that were being used were not just like caffeine, it was like really big scientific words.”*

Knowing health-related information, especially the content of severe conditions, would also trigger the participants’ negative feelings such as panic or worried. For example, S18 complained that *“I think it’s strange that reading the information can affect you. A lot of people look for the information because they hate not knowing, but also sometimes when you read it, it actually makes you panic more.”*

Two participants (S10 and S14) reported they did not feel any strong emotions involved in their search, maybe because one searched for relatively mild cold symptoms, and the other searched for gym routines. Only one participant (S04) kept positive during the whole HIS. She was searching to satisfy her curiosity about a health issue she encountered during browsing articles.

#### **7.3.4 Decision-makings**

At the last section of the interview, the participants were asked about whether they conducted any decision-making activity after the search experiences they just shared with us. This question was asked to capture the potential behavioral outcomes of OHIS. Decision-making refers to the decision that was made by the participants regarding the online instructions and a visit to a doctor. All the decision-making activities collected from the interviews were coded mutually exclusive. As a result, six categories were established, and the ranking based on the count is: 1) Follow the online instructions (N=8, 33.33%); 2) No decision-making activity (N=6); 3) Decide to see a doctor (N=4); 4) Decide not to see a doctor (N=3); 5) Terminate HIS behaviors (N=2); and 6) Reject the online instructions (N=1).

Surprisingly, one-third of the participants claimed that they decided to follow what they found online after the search. Among them, 5 applied the online instructions to change their workout plans or diet plans.

For those who reported there were no decision-making activities after the online search,



we further asked them why. It turned out they were relatively more cautious about online health information. S15 told us: *“You just cannot trust on all the information. Basically, I agree on information for symptoms, but for treatments, like I don’t trust it.”*

Only four participants confirmed that after their own OHIS experiences, they decided to see a doctor, because they felt that the online information was not sufficient for them to confirm their conditions, and the professional opinions were needed. For example, S10 said that *“Now I need to go to my doctor to confirm that and to make sure that I have that or I don’t, because they will be able to run tests and stuff like that. Whereas, like online, I can’t diagnosis with myself, I have to go to doctor for that.”*

However, some participants decided not to see a doctor after finding their symptoms were common through an online search. For example, S05 was experiencing headaches and doubted if she had diabetes. So she conducted an OHIS with the purpose of self-diagnosis and decided not to see a doctor because: *“I was scared just to my knowledge of what I was experiencing, but if you read more about how common it is online, then I feel a little bit better and that’s why afterwards I was like, ahh it’s not as alarming as I thought it was, so I’m not going to go to my doctor.”*

Two participants decided to never conduct OHIS again as it could be so frustrating and frightening. For example, S12 stated that *“I took a decision that never search for any other symptoms on the Internet. It starts with something small; it turns out to be something big. That’s how these websites work.”*

In only one case, S03 thought that the online information was so ridiculous that she decided to not follow the instruction. The instruction was to ask her to drink “Bulletproof coffee” to boost performance in the CrossFit.

### **7.3.5 Transitions between the Stages**

Figure 31 characterizes the transitions between every two adjacent aspects of the two stages. Firstly, as presented by the transition from Aspect I to Aspect II, which are the two aspects of the beginning stage, most participants relied on search engines to start their OHIS, and only a small portion of participants with the information needs to relieve a condition or

to learn about a health issue directly went to a familiar source. Then, these participants' emotions tended to change from negative to positive.

With respect to the outcomes, the transition from Aspect III “the emotion changes” to Aspect IV “the decision-making” is diverse. The emotion change from negative to positive would result in many different decisions, with a main stream following the instructions online. The continued negative emotion is likely to push the participants to see a doctor, but in some extreme situations, it would stop the OHIS forever. The neutral emotions would also drive the consumers to either follow the online instructions or see a doctor.

#### **7.4 Summary of Study 3**

Online Health Information Seeking has become pervasive with critical impacts on consumers' health. Yet, little is known about the connections between consumers' OHIS process and subsequent outcomes. Therefore, a semi-structured interview was conducted by adopting the Critical Incident Technique to understand the real-world search experiences from 24 laypeople participants.

Participants' OHIS process was characterized around information needs, search starting-point, then further their impacts on the emotion outcomes and decision-making outcomes were analyzed. Specifically, self-diagnosis is a common need for OHIS, where search engines are dominantly used as the starting point. More surprisingly, although mostly being viewed as helpful, OHIS might lead to consumers' extremely negative emotions and decisions. These findings deepen the understanding of consumer-centered health information seeking behaviors and provide insights for designing better interactive technologies to facilitate desirable online-offline transitions and thus promoting the outcomes of healthcare.

## 8.0 Study 4: OvCa Patients and Caregivers' Outcomes of OHIS

This interview study was conducted as part of an ongoing research project, Health E-Librarian with Personalized Recommendations (HELPeR), which has an overarching goal to build and implement a personalized health recommendation system for OvCa patients and caregivers. The analysis and results presented in this chapter focus on OvCa patients and their family caregivers' perceived outcomes of OHIS.

### 8.1 Data Collection

The specific goal of this interview study was to identify potential users' needs and expectations for the HELPeR system. This study was funded through National Library of Medicine (R01 LM013038). Therefore, the semi-structured interview protocol consisted of two major sections: users' own information seeking experience and the perceptions about the recommendation system being developed (See Appendix D for the full list of interview questions).

In the first section, the conversation was framed by the questions about the subjects' illness-related information and their own information seeking experiences, such as the information needs, channels, evaluation criteria, outcomes of information seeking activities, etc. Then the subjects were asked to identify their interested ovarian cancer-related topics of interest, and the potentially relevant information from the HELPeR resource library was presented to the subject to elicit their perceptions and suggestions.

Our inclusion criteria of participants were: (1) patients with a diagnosis of ovarian, fallopian, or primary peritoneal cancer (any stage of OvCa) and any time after initial diagnosis, including recurrence) or family caregivers of patients diagnosed with ovarian, fallopian, or primary peritoneal cancer, (2) 18 years or older, (3) ability to read and write in English, and (4) access to a computer (or mobile device) and the Internet.

From May 2020 to November 2020, a total of 20 subjects who were either ovarian cancer

patients (n=18) or family caregivers (n=2) were interviewed over Zoom due to COVID restrictions. The average age of participants was 58, and 60% of the participants (n=14) were diagnosed at stage III or stage IV. All but one had received chemotherapy, and half of the participants had finished the initial treatment. Each interview lasted approximately 1 hour.

## 8.2 Data Analysis

20 interview audio clips were fully transcribed and anonymized with pseudonyms. Then the research team proceeded with a 4-step content analysis process (Hsieh & Shannon, 2005).

In the first step, a tentative coding framework was developed with a combination of inductive and deductive approaches. The semi-structured interview protocol was applied in a deductive manner, while inductively, 3 coders were immersed in the 20 transcripts and performed open coding to generate codes derived from the data.

Secondly, 3 coders annotated 3 randomly chosen transcripts individually with the tentative coding framework. NVivo was used as the tool for the analysis. NVivo allows coders to highlight any length of text in the transcript and assign one or multiple codes from the pre-defined codebook. Once completed, the team met and discussed and established a revised coding framework.

Thirdly, using the revised codebook, 3 coders each coded all the 20 transcripts separately and reached an average percent agreement of 0.996 and Cohen's kappa coefficient of 0.746, ranging from 0.596 to 0.950 across 27 subcategories. Note that the inter-rater agreement is calculated by NVivo and the units of measure are characters. Following NVivo's guidelines, the current Cohen's kappa indicates a fair to good agreement.

Finally, the overarching themes and findings were synthesized through discussion among team members.

The whole coding process lasted approximately 3 months, and during which the research team met on a weekly or bi-weekly basis to discuss disagreements and come to resolutions.

### 8.3 Results: OvCa Patients and Caregivers' Outcomes of OHIS

The results reported here focus on the ovarian cancer patients and their family caregivers' perspectives towards online health information seeking and their associated outcomes, including **affective outcome, cognitive outcome, and behavioral outcome**. Knowledge change is a type of cognitive outcome.

The findings are framed by the themes that emerged from the interview analysis: 1) what leads to positive affective and cognitive outcomes (Section 8.3.1); 2) what causes negative affective and cognitive outcomes (Section 8.3.2); 3) behavioral outcomes associated with affective and cognitive outcomes (Section 8.3.3); and 4) online health community: a double-edged sword (Section 8.3.4).

#### 8.3.1 What Leads to Positive Affective and Cognitive Outcomes?

In the interview, the participants commented on their feelings and thoughts after conducting OHIS. In general, these affective and cognitive outcomes can both be either positive or negative. Surprisingly, the participants reported more positive cognitive outcomes, whereas OHIS was more associated with negative feelings.

Increased knowledge is the most frequently discovered positive outcome, with several participants reporting a gain of knowledge through OHIS. Patients and caregivers affirmed that they acquired knowledge about ovarian cancer from the Internet, where a wealth of information was provided. For example, as one of the interview questions, we asked the participants to assess their knowledge about ovarian cancer-related topics. S03, who rated herself 8.5 on a scale from 0-10, further told us *"I would say that I probably only got 3 of that 8 from my doctors. I got the 5.5 from the internet"*. S04 felt that after OHIS *"I've kind of started learning more about medical jargon around cancer and genetics."*

It is also notable that though searching online information did not answer all the questions of the patients, the belief of being knowledgeable did make them feel more informed and confident, as stated by S02: *"what I notice from the people I've talked to is I would say 30 or 40% of families dealing with this have the degree of the knowledge we do and the majority*

*do not. We're able to grasp it [information on the Internet] and translate it into what does it mean. It doesn't answer all the questions, but it makes you a more informed patient."*

From experience shared by S01, we also found that having appropriate comprehensive information could make the patient feel less worried (affective outcome), more confident, and assured the disease is curable (cognitive outcome). S01 told us that when she found and read NCCN guidelines for ovarian cancer, a resource that covers comprehensive aspects of ovarian cancer, she thought *"I'm very fortunate. And I knew going in I had a high degree of confidence in my team. I was assured that this was curable. I had really no reason to worry."*

### **8.3.2 What Causes Negative Affective and Cognitive Outcomes?**

Different from the cognitive outcomes, most of the affective outcomes reported by the participants in this interview were negative rather than positive. We summarized the reasons for participants' negative feelings and thoughts and grouped them into three categories: knowledge-related, quality-related, and disease-related.

Firstly, **lack of cancer-related knowledge** could engender extremely negative feelings such as hopelessness, fear, and frustration when the OvCa patients and caregivers conducted OHIS. For example, S13 told us, *"If it gets too complicated, then I get frustrated, and then I get more confused, and then I have to look up more words. And by the time I'm done, I'm getting totally upset."*

This can be explained that OvCa is, by nature, a complicated disease that requires a high degree of information and knowledge to be understood. For example, Donovan, Hartenbach, and Method (2005) found that women with OvCa under treatment suffer from 12 concurrent symptoms on average. OvCa patients and caregivers were required to search and learn a lot to manage their complex symptoms. In a recent study, Reid et al. (2021) found that 69% of the surveyed OvCa patients had not heard of or knew nothing about OvCa before their diagnosis. Lack of disease awareness makes the learning process extremely difficult and causes negative feelings toward the OHIS process.

S13 shared a story of how online information made her feel terrified due to the lack

of knowledge about what the content means: *“I can’t understand it [information on the Internet], when I first found out it was cancer, and then I saw that there was this five-year thing. I thought it meant that I only had five years to live. I had no clue what it meant. And they weren’t exactly descriptive on the Internet about it. So, I panicked and thought, “Oh, my gosh, I only have five years to live. But then my doctor explained that it’s just during five years, this is how many people survive. But at first it was terrifying. I would get on the Internet and find all sorts of stuff, and I wouldn’t know what it was, and it was terrible. I would lose sleep.”*

Besides, learning through OHIS is an independent and unassisted process, unlike consulting a doctor. S12 described her challenges when she tried to search and learn the medical terms on the Internet: *“I’d take a lot of words, medical words in particular, that I didn’t understand in the doctors’ notes and google those and get more information. And that was overwhelming because a lot of it was— it just seemed like doom and gloom. And you really don’t have someone there to turn to and say, “How is this?” And you’re just reading something. It’s not interactive. So, it was frustrating. I didn’t feel like I had a handle on all I could know about my disease other than cancer being— you’re fearful, obviously. But yeah, I didn’t feel like I ever got enough information.”*

Second, the *poor quality of some online information* caused a sense of frustration among participants. As stated in the literature (Chi, He, & Jeng, 2020), the quality of online health information is highly variable. Due to the low prevalence of OvCa, the online information on OvCa is generally scarcer, and the quality was less validated. S11 said *“what I initially read online was really frightening and wasn’t necessarily true.”* Worse yet, some participants were not confident in evaluating the quality of the online information that they searched. S17 stated that *“I’m a little leery about reading things that are more— it’s hard to tell sometimes what’s drug-pushed articles and what’s actual medical exam articles. And I get nervous about not knowing whether or not this relates to my specific process in my journey. . . I would just be kind of nervous about where I’m getting the information of, and knowing whether or not it’s valid information or if it’s just trying to sell me a drug or something like that.”*

Two specific quality issues of online content were reported by the participants: too generic and incomprehensive. Too generic made it hard for the patient to identify relevant

information (S03), and the incomprehensive online information about clinical trials eligibility impeded the progress of application and enrollment. S05 stated that *“once you’re turned down a couple of times, you don’t want to be bothered. It’s too much of a headache.”*

Third, the inherent characteristics of the OvCa led to feelings of fear, depression, anxiety, and hopelessness about OHIS for the participants. Women with OvCa experience high rates of recurrence, high symptom burden, and high intensity of treatment (Hagan & Donovan, 2013). In addition, most OvCa patients are diagnosed in advanced stages. A population-based study found one in four women with advanced OvCa die within 3 months of diagnosis (Urban et al., 2016). The majority of our participants were diagnosed with advanced stages. Thus, knowing the complexity and the prognosis of the disease can be very stressful to patients and family caregivers, regardless of the source of information.

Our participants described that information provided by reputable cancer organizations did not ease their anxiety on the poor prognosis of OvCa. S13 told us her feeling toward the information from the American Cancer Society, an authoritative cancer resource: *“Everything’s negative on there. I mean, it says you have a very low chance of living. That ovarian cancer is very hard to go through. They say it’s one of the worst, and it’s just a terrible way they have it set up. They give you absolutely no hope. Believe me... they’re terrifying. They’re so negative it’s ridiculous. It scares me. So, I mean, that site is one of the main reasons why I stopped looking stuff up”*. S13 was not the only one who stopped OHIS ever since being scared by OHIS. S15 quitted OHIS and turned to her doctor and children for information because *“I find I get depressed and more anxious if I read the bad [from Internet].”*

### **8.3.3 Behavioral Outcomes Associated with Affective and Cognitive Outcomes**

**Consulting with doctors** was the solution to deal with the lack of knowledge challenge and the associated negative feelings, as mentioned by several participants. For example, if S10 had questions about the online information, she would bring them to the doctor visit: *“It [online information] seems to be understandable. If we have any questions, we’ll ask whenever we go to the office”*. Very similarly, S16 said: *“Well, if I don’t know some of*



*the stuff [on the Internet], can't figure it out, I will ask the doctor".* Consulting with the doctor and seeking help about OHIS from professionals is a desirable behavior as it eases the patients' anxieties led by the lack of knowledge. It, however, also highlights the importance of getting stable and reliable communications with healthcare teams or providers for women with OvCa.

Another desirable subsequent behavior of OHIS is the **curation of the information sought**. S05 shared with us her personal strategies in organizing the information from her email subscriptions. Though it was a passive information seeking process, she managed to save the online information for later use: *"What I typically do when I get my email subscriptions, I'll put them in separate folders from the different companies or whatever, and I'd probably have at least 20 different folders"*

Beyond obtaining information and knowledge from the Internet, surprisingly, OvCa patients could result in changing their behaviors after OHIS. Surprisingly, three participants told us how they **followed the online instructions** about nutrition and diet after searching online by themselves. S11 told us that she used black raspberry powder as a consequence of her own research about the impact of concentrated black raspberry powder on cancer. S16 needed nutrition-related information due to her too low platelets for chemotherapy, and she also ended up following online instructions: *"my platelets were too low to get chemo, and I looked up and seen what kind of foods you could eat to help bring it up. It gave you this food in there which you could eat to bring it up. So I found it, and then started doing that food too."* Likewise, S03 went on fasting during chemotherapy as a result of her own research through OHIS.

It is notable that the nutrition and diet information is a commonly unmet information need of women with OvCa, as mentioned by several participants in this study. The participants would follow the online instructions to prepare foods or go on a diet, which could be problematic if the information found online is not clinically verified or not applicable to the patients' situation.

OHIS could also lead to **physical activity**. S11 shared with us that *"I have researched a lot about physical activity and joined the application Noom for managing what you eat and managing your exercise. So I learned a lot from that."*

It is worth noting that another OHIS behavioral outcome due to negative emotions and cognitions is the **information avoidance behavior**. Several participants who were scared or frustrated after OHIS tended to be more cautious of the online information. For example, S07 said: *“I try not to do it [OHIS] a lot because I did it first, and it was pretty depressing. So, I started to only when something came to my mind that I either didn’t feel comfortable asking my doctor because I was afraid, I would get emotional or it was something that I just felt like I needed to know right then.”*

While some patients told us they would even intentionally avoid any information from online resources, instead, they turned to other information resources, such as doctors. As S13 said: *“Now, if I have questions, I just ask the doctor. I don’t get on the Internet because it’s too terrifying. There’s way too much negative on there.”*

#### **8.3.4 Online Health Community: A Double-edged Sword**

The online health community provides various benefits to cancer patients and caregivers thanks to its around-the-clock availability, anonymity, immediate and time-delayed responses (Cline & Haynes, 2001; Grandinetti, 2000). The impact of participating in OHC, However, remains inconsistent in literature as it is hard to quantitatively and homogeneously measure the outcomes of OHC (van Eenbergen et al., 2017; Houlihan & Tariman, 2017). Aligning with previous work, in this study, the interviewed OvCa patients and caregivers reported inconsistent outcomes of participating in OHC. The reasons that led to the extremely different outcomes deepened the understanding of OHC.

On the one hand, some participants acknowledged that OHC provided them with informational support and emotional support. They feel informed and confident. S09 described OHC as *“a huge place of learning”*. Particularly, reading the experiences and perspectives of the other patients informed the participants of what to expect at certain stages of the disease trajectory. For example, S10 learned knowledge about OvCa in OHC when she was first diagnosed, the period that patients were more likely to experience an aggressive and rapid information seeking phase because of the diagnosis with cancer (Hayes et al., 2008). Learning in OHC could also lead to active communications with the doctors, as stated by

S12: *“Reading people’s stories was truly important to me because you could see that their doctor maybe took a different approach. It created questions for me, which was good, that I could go back and ask my doctor.”* Besides, the emotional support in OHC is also very critical. Learning about other patients’ stories could engender a feeling of strength. S18 told us: *“When somebody tells you their story, and you can pick up their strength and maybe improve your own strengths. like, they share stuff about how the drugs affected them. It’s a tough fight.”*

On the other hand, however, sharing and reading personal stories of fighting with OvCa could be a painful experience due to the negative feelings associated with the disease. S07 told us: *“when you hear that somebody’s been in chemotherapy for seven years or something, it gets a little depressing.”* S17 felt that the information in OHC was a constant reminder of her negative situation, which she was trying to move forward with positive, so she left the online community, even though she acknowledged the benefits of OHC.

#### 8.4 Summary of Study 4

Ovarian cancer (OvCa) patients and caregivers have constant and evolving information needs. To meet their needs, they seek information from various resources, including online health information. Although about 60% of cancer patients are now using the Internet to meet their information needs, little is known about online health information seeking (OHIS) among OvCa patients and caregivers. Therefore, 18 OvCa patients and 2 family caregivers of OvCa patients were interviewed, and content analysis was performed on the transcribed data.

Outcomes of OHIS were identified, including cognitive, affective, and behavioral outcomes. Knowledge change is a common positive cognitive outcome of OHIS. Subsequently, the results reveal what leads to positive affective and cognitive outcomes and what causes negative affective and cognitive outcomes. It is also notable that OHIS could also result in some debatable behavioral outcomes such as following the online instructions and information avoidance. The findings deepen the understanding of OvCa patients and caregivers and

provide insights to support their OHIS experience better.

## 9.0 Discussions and Implications

This chapter discusses the empirical, practical, theoretical, and methodological implications of this dissertation. First, the answers to the research questions are summarized by highlighting the major results from the four sub-studies. The associated empirical implications are discussed based on the alignment with previous literature in Section 9.1. Then, practical implications for design and theoretical and methodological implications are proposed in Section 9.2 and Section 9.3, respectively.

### 9.1 Discussion of Results

#### 9.1.1 Method Dimension

**RQ1: How do health consumers search and learn in OHIS to achieve different levels of learning objectives?**

The examination of the method dimension intends to answer the above research question. Given the research design, Study 1 (the user study with laypeople) mainly contributes to the method dimension, as no constraints on search approaches or online sources were imposed in the user study, resulting in a wealth of data about how the consumers search and learn. In general, Study 1 demonstrates that SAL behaviors and sources vary by different levels of learning objectives, and the variation is affected by the severity of health conditions.

Specifically, the results of search behaviors and source selections were presented separately to answer the two research sub-questions of RQ1.

- **RQ1a: How do health consumers' search behaviors vary by different levels of learning?**

Given the significant results of queries, clicks, and search efficiency presented in Chapter 5, it could be inferred that to achieve the Analyze level of cognitive learning objective in the severe condition (SC), participants used the longest but fewest queries ( $NQ = 1.83, LQ =$

5.53). Meanwhile, using those queries, participants saved the greatest number of relevant pages, compared to two the other two levels ( $NSP = 8.17$ ), and they believe these saved pages are also more relevant to the task, which could be seen from the highest search efficiency among all the levels and conditions ( $SE = 24.20$ ). This might indicate that, after searching the first Understand task in SC, in the Analyze level, coming up with the appropriate query and recognizing relevant information at that stage.

In contrast, to achieve the Evaluate level in SC, participants used the shortest but the largest number of queries ( $NQ = 4.46, LQ = 3.60$ ). The search efficiency, however, was relatively low in the Evaluate level ( $SE = 7.56$ ). This might suggest that the participants used a trial-and-error strategy. In other words, the participants tried many short queries with few terms and changed to new queries if the search results did not appear to be very relevant. Evidence also came from the examination of the number of unique queries and terms. Participants issued the largest number of unique queries and terms in the Evaluate level of SC, compared to the Understand and Analyze levels, implying that, unlike the other two levels, participants issued diverse queries and terms beyond the task description to overcome the Evaluate level task.

The trends of participants' behaviors among the three levels in SC align with the findings in Jansen, Booth, and Smith (2009), in which the search behaviors in the lower levels (i.e., Remember and Understand) are found to be similar to the higher levels (i.e., Evaluate and Create). Conversely, in Jansen et al. (2009), the middle levels (i.e., Apply and Analyze) are found to be most difficult since they were associated with more searching behaviors than other levels. In the current study, with the measure of the search efficiency, we believe that the participants were more satisfied with the queries issued and the information found for the Analyze level in SC. On the contrary, the participants appeared to experience more difficulty in the Evaluate level, as implied by the low search efficiency.

Concerning the mild condition (MC), descriptive statistics revealed that participants' search behaviors — including the number of queries, number of clicks, and search efficiency — increased with learning objectives, though some pairwise comparisons were not statistically significant. This behavioral pattern is consistent with several existing studies (Brennan, Kelly, & Arguello, 2014; Wu, Kelly, Edwards, & Arguello, 2012; Kelly et al., 2015; Thomas,

Bailey, Moffat, & Scholer, 2015), , in which higher levels were associated with more search behaviors.

Therefore, the effects of levels of learning objectives on the search behaviors are affected by the severity of the health conditions. For the MC, which is more common and less complex, participants generally have more search behaviors with increased learning levels. On the other hand, in the search and learn for the SC, which is a serious disease involving complex concepts and procedures, participants tend to face difficulties in creating appropriate queries and collecting relevant information for search tasks with high levels of learning objectives.

- **RQ1b: How do health consumers' selected sources vary by different levels of learning?**

To answer this question, a follow-up content analysis of participants' clicked pages was conducted. Among the nine different types of sources, attention should particularly be paid to two dominant types of sources: the health-specific page and search engine result page.

Notably, the health-specific page (HSP) is the prominent source, with a leading position, especially in the Understand level in SC and the Evaluate level in MC. This might be because that HSPs, such as WebMD.com, display multiple aspects of a specific health issue, making them particularly helpful for consumers to learn about an issue quickly. Search engines might consider these webpages highly relevant when consumers lack sufficient knowledge about the health issue. Though we assume webpages of the same genre imply similar source characteristics in terms of content, presentation, purpose, and so on, it is important to note that variances in quality may exist even in the same type of web page. Thus, some HSPs may be of higher quality than others. Given its popularity, it is crucial to monitor and improve the quality of HSPs.

The search engine result page (SERP) is the second most selected source in SC and the first in MC across the whole OHIS process across the three learning levels. It is obvious that the share of SERP increases when consumers select more diverse types of online sources, and it occurred in the higher levels in SC and all the three levels in MC. With respect to SC, SERP's increasing proportion might indicate that consumers have to collect information from various sources to address the tasks with higher level of learning objectives. Regarding

the large amount of SERP in MC across all the learning levels, one possible reason stems from the specific health condition (i.e., weight loss) that we selected to design the search scenario and tasks. As one of the most common health information needs that drives OHIS (Yilma et al., 2017; Jong & Drummond, 2016), it is not surprising that online information about weight loss is scattered across different types of webpages. As a result, search engines are critical in organizing scattered information for such mild and common health conditions. Practical implications for promoting SERP are further discussed in Section 9.2.

### 9.1.2 Information Dimension

**RQ2: What information do health consumers search and learn in OHIS to achieve different levels of learning objectives?** While there exists an extensive body of literature investigating online health communities (OHCs), and it has been proven that patients and their caregivers would use OHCs to post questions and address information needs (D. Yang, Kraut, et al., 2019; Y.-C. Wang et al., 2015; Chuang & Yang, 2014), little has been done to differentiate the information needs with different levels of learning objectives and the associated informational support provided by peers in OHCs. Study 2 demonstrates that the consumers' interaction in OHCs can be described and studied from a learning perspective. Not all the information needs are identical regarding the underlying complexity of cognitive learning objectives.

Through content analysis of the National Ovarian Cancer Coalition (NOCC), a leading OHC for ovarian cancer, it is confirmed that ovarian-cancer patients and their caregivers resort to OHCs for information needs with different levels of learning objectives. The information needs of each learning level are featured with different characteristics, including one or multiple cancer-related topics, the role of the poster, and the self-disclosure behavior. These findings provide answers to *RQ2a: What are health consumers' information needs of different levels of learning?* Additionally, based on the classification of five types of informational support in the comments, the results demonstrate the types of information available in OHCs to support the learning of different levels, which answers *RQ2b: What types of information do health consumers acquire for different levels of learning?*



The major findings of the two sub-questions are discussed below.

- **RQ2a: What are health consumers' information needs of different levels of learning?**

First of all, the NOCC was mainly used by the OvCa patients and their caregivers to address information needs with low- to middle-level learning objectives. The Analyze level was found to be the most prevalent learning objective in the NOCC ( $N = 257, 46\%$ ). The consumer initiated a post to pursue connections and relationships among concepts under uncertainty, followed by the Understand level ( $N = 204, 36\%$ ), which referred to the posts seeking basic facts and concepts. It is notable that only 18% of posts with information needs were categorized as Evaluate level learning objective, in which the consumers asked other OHC members to help them to make decisions or judgments based on their specific conditions. These results are partially different from the findings in Savolainen (2011), in which the author found that more than 70% of the posted questions in the online discussion forums were seeking an opinion or evaluation of an issue, resembling an Evaluate- or Analyze-level learning objective, while needs for factual information and procedural information were less common. The conflicting results could possibly be attributed to the different health conditions being studied. In Savolainen (2011), the topic of interest in the threads was depression, while in the current study, it was OvCa, which is listed as one type of rare cancer by the National Institutes of Health (*Ovarian Cancer*, n.d.). As a result, the general public lacks disease awareness and education regarding OvCa, and the information is complex and scattered. This might cause the consumers in OHCs to seek some basic facts and concepts of the Understand level more often. Besides, the prevalence of Analyze-level information needs could be explained by the fact that OvCa is a very complex disease. Women with OvCa and their caregivers are required to learn and sort out many concepts under uncertainty at the same time. As an example, on average, women with OvCa under treatment need to manage 12 concurrent symptoms (Donovan et al., 2005).

To capture the characteristics of the information needs of different learning levels, a coding framework of nine OvCa-related topics was adopted to annotate the topics of the information needs in the initial posts. The topics include disease management (DM), symp-

tom management (SM), treatment (TM), treatment decision (TD), emotional management (EM), self-management (SF), practical needs (PN), caregiving (CG), and others (OT).

First, the descriptive results of the content analysis show that the number of topics increases with learning levels. In the Evaluate level, on average, one initial post is associated with 1.828 topics, while the number of topics was just slightly above one in the Understand level ( $N = 1.049$ ), and around one and a half in the Analyze level ( $N = 1.498$ ). This result is somewhat predictable, as the learners are expected to master multiple concepts and knowledge components to achieve higher learning objectives. But, notably, treatment is the most popular topic of interest among all the information needs, regardless of learning levels. The finding is in accordance with the results in Madathil et al. (2013), in which treatment-related information was found to be the most sought-after information by patients (41.3%), compared to ovarian cancer-specific and coping information. The data analysis was conducted in the Ovarian Cancer National Alliance (OCNA), another OHC for OvCa. In this current study, given the fine-grained topic classification framework, nine different topics were identified, and the posts were classified in a non-mutually-exclusive manner. Treatment was still found to be the most popular topic. Such a finding further underlines OvCa patients' and their caregivers' high demand for treatment-related information and support.

Other than treatment-related information, the numbers of posts seeking information on topics of DM, SM, treatment decision (TD), and EM varied by levels of learning objectives, as indicated by Chi-square results. It is noteworthy that TD accounted for a large share in the Evaluate level, despite the concern that an OHC might not be the appropriate resource in which to ask for treatment-related decisions.

Beyond topics of interest, information needs were further captured by examining the role and the self-disclosure behavior of the question posters. The roles of posters (i.e., patients or caregivers) were found to be independent of levels of learning objectives, and among the 560 initial posts with explicit information needs, 70.89% were posted by OvCa patients, compared with that in Madathil et al. (2013) (68.9%). In addition, self-disclosure of personal information was found to happen more frequently in the Evaluate level. This might indicate some consumers believed it was necessary to disclose personal information, to a certain extent, to learn peers' responses on decisions or judgments.

- **RQ2b: What types of information can health consumers acquire for different levels of learning?**

It has been demonstrated in previous work that online knowledge building and collaborative knowledge production take place through the discourse among members in online communities (Griesbaum, Mahrholz, von Löwe Kiedrowski, & Rittberger, 2015). After posting questions in the OHC, consumers can acquire a range of information, such as factual information and personal opinions, from comments to address their information needs and close knowledge gaps (Chuang & Yang, 2014; Harkin et al., 2017; Gill & Whisnant, 2012). Therefore, examining the type and amount of informational support in the comments provides a means to study the quantity and quality of the information that health consumers can acquire for different levels of learning.

In general, consumers in the NOCC received the largest amount of support in the Understand level ( $N = 7.68$ ), followed by the Evaluate level ( $N = 7.07$ ) and the Analyze level ( $N = 5.63$ ), as indicated by the average number of comments. However, the number of comments itself was not enough to reflect the quality and quantity of support in OHC because one comment could provide none or one or multiple types of informational support. With respect to the types of support in the comments, the coding framework in Chuang and Yang (2014) was adopted to classify comments into five types of informational support, including Advice, Referral, Fact, Personal Experience, and Opinion. Additionally, Emotional support and Others were added to capture other types of support beyond informational.

The correlation matrixes (Figure 26) confirm that some types of support are more likely to be offered together and some are not. The co-occurrence is independent of levels of learning objectives in the initial posts. For example, personal experiences and facts were most likely to be shared together ( $r = .36$ ), and advice and opinion could be provided together in one comment ( $r = .25$ ).

With respect to the amount of support received in different learning levels, descriptive results indicated that in general, the total amount of support of all types was the largest in the Evaluate level, followed by the Understand level, and it was the smallest in the Analyze level. As for each type of support, Fact was found to be acquired the most, compared to other types of support. This result is consistent with the results in Chuang and Yang (2014),

which were based on a manual analysis of an alcoholism OHC. Regarding the effect of levels of learning objectives, significant differences among the three levels were found in the Advice support ( $F(2, 437.84) = 9.687, p = .000$ ), Opinion support ( $F(2, 418.18) = 11.563, p = .000$ ), and Emotional Support ( $F(2, 395.88) = 3.236, p = .014$ ), as determined by one-way ANOVA. Generally, more advice, opinion and emotional supports were obtained for the information needs with the Evaluate level of learning. One possible explanation about this finding was that some subjective knowledge to a certain extent was needed to support cancer patients and caregivers' information needs of the Evaluate level. As justified by the interviewees in Harkin et al. (2017), online practical online advice shared by peers in OHCs was welcomed by many interviewees, as such information led them on a "journey to become informed".

It is also notable that though the questions with Analyze level information needs were the most frequently posted in the OHC, there were the lowest number of average comments and the least amount of almost all the types of support in the comments. Other measures beyond the number of comments and supports are needed to explore this problem in the future further.

### 9.1.3 Knowledge Change

#### **RQ3: How well do health consumers learn in OHIS (knowledge change)?**

This research question was investigated in all the four studies of this dissertation from different perspectives. Specifically, Study 1 and Study 2, which focused on method dimension and information dimension, respectively, revealed how information and methods related to knowledge change, and the results suggested that *knowledge change varies by levels of learning objectives, health conditions, and the amount and types of support received*. Additionally, the two interviews about the outcomes of OHIS also provides answers to this question, given that knowledge change is a particular type of OHIS outcome. Therefore, the results of the two interviews were triangulated and discussed from two points: *knowledge change as a type of cognitive outcome of OHIS* and the relationship between *knowledge change and emotion change of OHIS*.

**9.1.3.1 Knowledge Change Varies by Levels of Learning Objectives** Knowledge change (KC) was measured as the difference between the prior-search knowledge stage (PK) and after-search knowledge state (AK) in Study 1. Notably, a gain in the knowledge state was observed in all the learning levels in both health conditions. But the knowledge gain and the PK and AK behind it vary by the three levels of learning, which means that participants could learn at certain levels better than others.

Participants' KC was the highest in the Understand level, especially in the SC; meanwhile, their PK was also the highest in the Understand level. Such findings suggest that the participants could achieve the best learning outcomes in the lower levels, though they had already known the most for the tasks in the lower levels, compared to higher levels. Additionally, the statistically significant differences were only found between the Understand and Evaluate levels, as indicated by the simple pairwise comparisons. This implies that the differences in the knowledge change were more likely to happen between the extreme low and high levels in the A & K's hierarchy rather than between low and middle levels or between middle and high levels.

It is interesting to note that, in general, KC decreased with the increase of learning objectives in both conditions, as shown in Figure 16. These findings show that the participants established a higher knowledge gain initially, and the gain gradually slowed down and finally reached a plateau if participants' searching and learning for the three levels were viewed as a continuous OHIS process. These results align with the findings in previous work (Han, 2018; J. Liu et al., 2013), and the plateau in the knowledge gain was defined as the "ceiling" effect in J. Liu et al. (2013). For future studies, it will be interesting to explore the design that mitigates the ceiling effect and enhances health consumers' searching and learning for complex health-related tasks in an OHIS.

**9.1.3.2 Knowledge Change Varies by Health Conditions** Participants' knowledge changes were different in the two health conditions examined in Study 1. In particular, it was found that participants were more likely to achieve more knowledge gain and higher AK in SC, regardless of learning levels. On the contrary, in MC, their AK and KC appeared to be lower. The results suggested that OHIS is a promising learning resources for health

consumers to learn about health-related information, even for serious and complex diseases.

One possible explanation for the varied KC in the two conditions is the different PK of the two conditions. As indicated by the two-way ANOVA results, health conditions had a larger effect on PK than that of learning levels. This might be because the participants' PK and familiarity of the two health conditions were quite different, as one of them was a serious and complex disease (i.e., multiple sclerosis) while the other was a health-related issue (i.e., weight loss) that they were more likely to be familiar with. Evidence also came from a previous study (Yu et al., 2018), in which the authors found that the more individuals tended to obtain the lesser knowledge gain in online search, the more they were familiar with a topic, though the topics were not constrained in the context of health information seeking.

**9.1.3.3 Knowledge Change Varies by Amount of Support Received** The analysis of the posts and comments in the NOCC in Study 2 exhibited how consumers' knowledge changes and evolves in the real world without manipulations and interruptions in the lab settings. The knowledge state was reflected by the level of learning objective in the initial post with information needs, and the knowledge change was measured as the change from the current post's learning level to the subsequent post's level of the same user. Specifically, posts from the same users were tracked to reveal two questions: *1) how do the users' knowledge changes in general, knowledge increase, decrease or no change?* and *2) how is the amount of support received from the comments related to knowledge change?*

First, most of the users who posted more than one initial post with information need in the NOCC tended to increase (32.4%) or remain the same (41.7%) regarding the level of learning objective as they continued posting and seeking information in the same forum. Second, it should be noted that for users who increased the level of learning objective in the next post, a larger amount of support in Advice, Personal Experience, Opinion, and Emotional Support was observed in the current post (Figure 30). In other words, receiving more social support was associated with the knowledge increase. Though the result was not statistically significant, this finding adds to previous studies that have demonstrated the effect of social support on member retention and engagement (X. Wang et al., 2021; Xing, Goggins, & Introne, 2018; Y.-C. Wang et al., 2012) and contributes new evidence on

the potential effects of social support in collaborative knowledge building and generation in online communities (Griesbaum et al., 2015). In-depth future research promises to investigate the relationship between receiving social support, especially informational support and the knowledge change in OHC.

**9.1.3.4 Knowledge Change as a Type of Cognitive Outcome** Though Study 1 and Study 2 have investigated two groups of representative health consumers' (laypeople's and OvCa patients') searching and learning in OHIS and the associated knowledge change, both adopted quantitative methods. To complement Studies 1 and 2, two qualitative semi-structured interviews were conducted to directly elicit inputs from health consumers' perspectives and generate in-depth insights in natural settings. Laypeople and women with OvCa were interviewed, respectively, in Study 3 and Study 4.

Though the participants of the two interviews were different, both of them focused on the outcomes of OHIS. Knowledge change, especially knowledge gain, was reported as an important cognitive outcome of consumers' OHIS, along with manifold outcomes of OHIS, such as emotional outcomes and behavioral outcomes.

It is interesting to note that consumers' cognitive outcomes of OHIS were mainly positive, and they mainly came from the feeling of being more informed after OHIS. In other words, knowledge gain is the most frequently reported cognitive outcome of consumers interviewed in this dissertation. OvCa patients and their family caregivers affirmed that they acquired knowledge about ovarian cancer from the Internet, where a wealth of information was provided. One OvCa patient (S03) even told us she gained more knowledge about OvCa from her OHIS than from her doctors. According to what we found in Study 1, laypeople were able to gain knowledge about multiple sclerosis after a short-term OHIS. It makes sense that the OvCa community could benefit from an extensive online resource to gain knowledge about cancer and address their assorted informational needs.

In addition, according to the laypeople interviewed, searching to learn was found to be a critical motivation that leads to OHIS. Most of the participants conducted OHIS to fill a knowledge gap, either learning for self-diagnosing an uncertain symptom or learning about a health issue. This finding confirms that one's knowledge state affects OHIS, and the

knowledge change is the effected outcome (A. T. Chen, 2021).

**9.1.3.5 Knowledge Change and Emotion Change** It is noteworthy that knowledge change is intertwined with emotional change in OHIS. This finding was observed in both interviews. In most cases, a lack of knowledge was the reason that led to laypeople’s negative feelings in OHIS. For example, as stated by S17, the failure to understand the “scientific words” made her feel very frustrated. This also happened to OvCa patients and caregivers. Most OvCa patients had not heard of or knew nothing about OvCa before their diagnosis (Reid et al., 2021), making them the equivalent of laypeople. However, facing OvCa, a fatal and aggressive disease, posed extra burdens on women with OvCa.

Few subsequent positive emotion outcomes were found in the interview with OvCa patients and caregivers, though many participants reported feeling more knowledgeable after OHIS. The rich online information did make the participants feel more knowledgeable and empowered, but at the same time, the overwhelming information, or the learned knowledge, about the disease also could be associated with terrible feelings. This could be attributed to the inherent characteristics of OvCa, such as the high complexity and aggressive progression of the disease (Hagan & Donovan, 2013; Donovan et al., 2005). Madathil et al. (2013) suggested that the presentation of OvCa-specific information should be carefully adjusted to take the information seekers’ negative emotions into account.

## 9.2 Implications for Design

### **RQ4: What are the implications for system design?**

Based on the results and discussions of this dissertation, the researcher proposes implications for design that can better facilitate health consumers’ search as learning activities in OHIS. In general, the proposed practical implications can be grouped under three critical components of OHIS: 1) adapting to consumers’ *needs and contexts*, 2) advancing search *tools and platforms*, and 3) promoting desirable *OHIS outcomes*.



### 9.2.1 Adapting to Consumers' Information Needs and Search Contexts

First, the findings in this dissertation highlight the complexity and importance of health consumers' information needs and contexts. In particular, search systems should be able to take into account the fact that consumers' health-related information needs are not identical. Rather, they feature different levels of cognitive learning objectives and certain health conditions. The complexities of consumers' information needs should be recognized by the search system and adjusted to accordingly.

**9.2.1.1 Adapting to Learning Objectives** This study has exhibited how health consumers' information needs feature different levels of learning objectives. To address these needs, consumers resort to the internet and go through the search as learning process, which varies in its information and method dimensions. As a consequence, consumers' knowledge change also varies by learning objectives.

Given the fact that the information and method dimensions fluctuate throughout the learning process, search systems should be designed to adapt to the consumers' level of learning objective at a given stage to support and enhance the search as learning in OHIS. For example, if the consumers are searching for low-level knowledge, e.g., Understand level, they are likely to look for information about one specific topic, and prefer health-specific webpages (HSP) as information sources. Therefore, a high-quality HSP focusing on the topic of interest should be prioritized. In contrast, consumers tend to look for information about multiple topics for high-level learning, e.g., Evaluate. They may experience difficulty in coming up with accurate search queries, so query suggestions are particularly needed.

Additionally — as supported by the evidence about knowledge change — while the search process might not be long, consumers gain knowledge when interacting with search systems and different sources. Thus, search systems are not expected to treat consumers as being completely naive about health conditions and may gradually adjust the content to meet patients' knowledge states and learning curves. The search as learning systems that accommodate consumers' learning objectives should be more intuitive and interactive and provide sufficient support, particularly for medical terms and jargon.

**9.2.1.2 Adapting to Health Conditions** Health-related search systems should also be able to adapt to consumers' health conditions, as consumers' OHIS and SAL behaviors differ by the conditions they search for. In Study 1, laypeople were requested to search and learn about two conditions with different severities, and the results showed that they faced different challenges in the two conditions.

In the severe condition (SC), laypeople mainly struggle at the Evaluate level, as inferred by the extremely low search efficiency. Moreover, they need help from the system to issue search queries. Additionally, in terms of the source selection, laypeople with severe conditions (SC) are more likely to trust health professionals' information. This raises a concern over whether or not laypeople can understand and digest information from professionals given their limited knowledge about complex and often urgent health conditions.

More importantly, regarding mild health conditions which are nonetheless much less severe, the apparently relevant information comes from diverse sources, including many non-professional sites that sell health-related products in addition to providing health advice. Given this complication, can laypeople acquire adequate health knowledge and information literacy to separate low-quality or biased information from truly useful information? For example, a website selling protein powder suggests a much larger quantity of protein is needed per day for a teenager than in normal practice.

The results indicate that certain support mechanisms are needed to help laypeople search and learn based on the severity of a health condition. For example, when dealing with mild or common health issues such as nutrition and diet, search systems must play a more important role in organizing scattered information at the beginning of the search process to help laypeople judge the sources' reliability. However, when handling a search for a serious disease, laypeople focus on information from fewer sources; thus, broadening the search results with different sources would help them better explore the relevant information.

In addition, OvCa patients and caregivers exhibit constant and dynamic information needs that change based on disease trajectory. When first diagnosed, most OvCa patients could be considered laypeople given their low awareness of OvCa. Their OHIS behaviors and the supports they need would be comparable with those in severe condition. But as the disease trajectory proceeds, both their information needs and knowledge about the disease

evolve gradually. A more personalized search system would be beneficial to provide adjusted and dynamic support.

## **9.2.2 Advancing Search Tools and Platforms**

Practical design implications for search engines and online health communities are proposed in accordance with the major findings of this dissertation.

**9.2.2.1 Advancing Search Engines** Most of our participants, including laypeople in Study 1 & 3 and OvCa patients in Study 4 relied on search engines to initiate their OHIS and SAL by issuing queries and navigating the search results. This finding aligns with prior work (Eysenbach & Köhler, 2002; Fox, 2013; D. L. Hansen et al., 2003). In D. L. Hansen et al. (2003), 87% of the adolescent participants visited sites directly presented by the search results or the search engine’s recommendations. Eickhoff et al. (2014) found that 3% of web search sessions were associated with a learning intent.

There are three concerns and suggestions regarding this finding. First, it is questionable whether or not consumers are capable of issuing correct or accurate queries to describe their health information needs, as such an action often requires a certain level of health-related background knowledge. Consumers’ OHIS queries are revealed to suffer from problems such as misspellings (D. L. Hansen et al., 2003) and being too general (Eysenbach & Köhler, 2002), resulting in search failure. For example, in Study 1, the participants adopted a trial-and-error strategy by using the shortest but largest number of queries and collected the least-relevant information for each query. Search engines should therefore implement query suggestions, particularly to help consumers with tasks for which they lack knowledge. In addition, features that could promote learning should be explored. For example, using social tags to filter search results was found to be an effective feature that could enhance learning in previous work (Kammerer et al., 2009).

Second, some consumers may be unaware that the ranking positions can be biased by marketing, especially when searching for health information that may lead to a purchase (Kim et al., 2011). For example, if one is searching for “weight loss,” they may be

directed to advertisements or e-commerce websites via the top-ranked results. It is important to highlight the search engines' ranking and scoring mechanisms to the health consumers.

Third, many participants expressed the challenge and frustration they experienced when reading complicated medical-related content in OHIS. For example, S13 in Study 4 said: *“Just layman terms are great with me. If it gets too complicated, then I get frustrated”*. Such findings provide important insights for designing search engines. For example, for search result ranking, the content in layman's terms should perhaps be assigned more weight.

**9.2.2.2 Advancing Online Health Communities** The analysis of NOCC demonstrated that OHC is a promising resource for health consumers to address information needs with different levels of learning objectives, and OHC can help consumers improve their knowledge if their information needs are well-supported with informational and emotional support.

On the other hand, it is also notable that conflict perspectives towards OHC were expressed by interview participants in Study 4. OHC is a double-edged sword. With its social feature, OHC amplifies the benefits from exposure to a wealth of information as well as the negative emotions shared by peers. In addition, there are also concerns about the quality of the narratives shared by patients in OHC (Bekker et al., 2013; J. Zhang, 2019). False information and rumors could give members false expectations (Harkin et al., 2017). To deal with the potential downside of OHC, it is suggested that the content be carefully administered by moderators who have a professional background. Attention should be particularly devoted to posts with extremely negative tones and information-seeking posts with high learning objectives such as pursuing judgments and decisions from peers. In addition, some high-quality learning materials could be developed and disseminated via OHC, as it has been proven to be an active informal learning platform.

### **9.2.3 Promoting Desirable Outcomes**

Last but not least, it should be mentioned that knowledge change or knowledge gain is one particular type of outcome of OHIS and is the most frequently reported positive

cognitive outcome by the participants in the two interviews. Other than knowledge change, however, affective and behavioral outcomes communicated in the interviews were not always desirable. Both laypeople and OvCa patients and caregivers brought up extreme negative emotions and concerned behaviors they experienced after OHIS. Therefore, the researcher believes that attention should be drawn to these “side effects” of OHIS, as these negative outcomes deserve mention.

Some participants in both interviews terminated OHIS due to extreme negative emotions. In many cases, consumers conduct OHIS to calm or eliminate their anxiety, but conversely, the search may result in even worse fear or information avoidance. However, the information that causes the exceedingly negative feeling is sometimes simply misinterpreted or even irrelevant to the consumer’s situation (El Sherif, Pluye, Thoër, & Rodriguez, 2018). In our findings, this negative feeling could lead to terminating HIS forever, emphasizing consumers’ terrible sense of frustration and stress.

Understanding the causes of these negative emotions would help to reduce such “side effects” and promote a positive consumer experience of OHIS. Interestingly, Study 3 showed that consumers who directly visited a familiar source all transformed their negative emotions into positive ones. Similarly, White and Horvitz (2009) found that search engines potentially escalate medical concerns, making consumers feel more anxious. Given the relatively small sample size, the current study is not sufficient to conclude that health consumers are more likely to feel good after OHIS if they directly visit their familiar source, but a promising research question in the future would be to explore and examine the different impact of a familiar source vs. search engines.

We expect this paper to be a starting point for designing interactions to encourage health consumers to engage in desirable online-offline behaviors, for example, promoting the seeking of care if people encounter health problems, relieving consumers who feel uneasy about their health condition. It is critical to investigate further how to create a better experience for consumers who decided to never conduct OHIS again due to escalated health anxiety.

As demonstrated in Lauckner and Hsieh (2013), even a small manipulation of the presentation of the search result page influences health consumers’ emotional outcome. However, research seeking to understand the outcomes of OHIS and the factors leading to different out-

comes remains scarce. It is important to fill this gap and provide consumers with the desired help. As demonstrated in this study, knowledge gain is generally reported as a dominant positive outcome of OHIS. Therefore, this study highlights the question of how to design, from a search as learning perspective, a better search engine and online health information resources to ensure positive outcomes.

### **9.3 Theoretical and Methodological Implications**

Aside from the empirical and practical implications, there are several theoretical and methodological implications of this dissertation study.

#### **9.3.1 HearSAL Framework**

HearSAL is the conceptual framework that guides the design of this study, but it is also a stand-alone research output of this work. Built upon the literature, models and theories from OHIS and SAL, HearSAL outlines an approach to investigating consumers' search as learning experiences and perspectives in OHIS that could be adopted by future researchers.

The HearSAL framework highlights three levels of learning objectives, from low to middle to high. Information and method are emphasized as two critical dimensions in the process, and knowledge change is the subsequent outcome.

In this study, the framework was utilized for two populations, each focusing on examining one of the two dimensions in the framework given the characteristics of the population and the research methods employed in each sub-study. Specifically, laypeople were recruited to a user study that examined the method dimension in HearSAL as the obtrusive user study allowed the researcher to control the search tasks and collect rich behavioral data that was later used to uncover the search behaviors and sources at each level of learning. With respect to the information dimension, an OHC for OvCa was sampled to investigate the characteristics of the consumers' information needs at each learning level. Such unobtrusive data collection methods aided in studying the real information needs of consumers with health concerns,

such as the OvCa patients and caregivers selected in this study. The sampling of the two populations enables better testing of the validity and generalizability of the framework.

The empirical studies also add certain details to the framework, such as what to examine in the two dimensions, how to design search tasks with the three learning levels, and how to measure knowledge change. The results also expand the framework by pointing out the importance of consumer-related characteristics, such as prior knowledge and the severity of the condition. More empirical evidence is needed in the future to further expand, validate and polish the framework. In addition, three future expansions of the HearSAL framework that could be explored are: 1) the validity of the framework with respect to different health consumer groups other than laypeople and cancer patients; 2) the long-term learning effect and knowledge change with longitudinal studies; and 3) the other three cognitive learning levels in A & K's taxonomy.

### **9.3.2 Coding Frameworks**

Several coding frameworks are originated in this study. Some frameworks can serve as the theoretical basis for future researchers who look into similar problems and contexts. Among them, the ones worth noting are elaborated below:

**9.3.2.1 Classification of Online Health Sources Based on Genre** There are numerous existing ways to classify sources (L. Lu & Yuan, 2011; Krikelas, 1983). However, to the best of our knowledge, there have been few discussions on health information source classification based on the genre of the online source. In order to fill this gap, in Study 1, a genre-based classification framework for online health information sources is proposed through a bottom-up analysis of the participants' search logs. Such a framework can provide future researchers with an approach to classify online health information sources.

**9.3.2.2 Approaches and Frameworks of Knowledge State Assessment** It is hard to measure an individual's intangible knowledge state. In this study, two different knowledge state assessment approaches are used to measure participants' knowledge state in the user

study and in the NOCC. In the user study, the participants were asked to compose written summaries before and after the search sessions, which were later used to assess their knowledge state within a framework that defines the criteria. The method of assessing written summaries originated from previous work (Kammerer et al., 2009), but the assessment criteria are first proposed by this study. As for the posts in NOCC, a coding framework of levels of learning with descriptions and examples is developed to assess the knowledge state in the initial posts with information needs. The two approaches and corresponding frameworks can be examined and expanded upon by future researchers.

**9.3.2.3 Framework of Task Design** By mapping the OHIS intentions (Cartright et al., 2011) to the levels of cognitive learning objectives, a framework that guides the design of search tasks is proposed and detailed in Table 2. Correspondingly, the scenario-based task design, which consists of three sub-tasks with increasing levels of learning and two health conditions, can be adopted by future researchers to study similarly complex and exploratory online seeking behaviors.

### **9.3.3 Research Design**

This study employs a mixed-method research design, which consists of qualitative and quantitative data collection and analysis methods. In addition, two different groups of consumers are selected and sampled to emphasize some parts of the conceptual framework and answer specific research questions. Such a design confers benefits as well as specific challenges to the study. This section explains the benefits and rationale of adopting the data from different health consumer groups, and the mixed-method approach.

**9.3.3.1 Mixing Results from Two User Groups** Given the same overarching research goal, the relationship between the two consumer groups is expansion, because the two studies of the two groups highlight different inquiry components and adopt different methods, thus extending the inquiry's breadth and range (Creswel, 2009).

The major advantage of a laypeople user study is the plentiful and unconstrained be-



havioral data on the method dimension, while the OHC dataset adds more capabilities to uncover consumers' self-motivated SAL information needs. The data collection methods for studying laypeople and OvCa patients are determined by the inherent characteristics of the two user groups. Specifically, a user study with two pre-assigned OHIS scenarios is designed to elicit search behaviors from laypeople, as they were found to mainly rely on search engines and conduct OHIS in search of health issues with a very broad range of topics. Thus, the user study allows us to automatically record laypeople's search behaviors and control the health conditions. In contrast, for the OvCa patients, scraping OHC datasets provides the possibility to reveal their relatively long-term SAL experience in a natural context.

Additionally, the data from these two groups validates the HearSAL framework in two settings, increasing the external validity of the study. This allows the possibility of generalizing and transferring the findings to other health consumer groups in a similar context such as laypeople or cancer patients.

**9.3.3.2 Mixing Results from Different Types of Data** A mixed-method approach consisting of quantitative and qualitative data is adopted in this study. Mixing both types of data contributes strengths that offset the weaknesses of either the quantitative or qualitative approach alone (Creswell & Clark, 2017). For example, in Study 1, though the search scenarios and tasks are designed to be close to real practice, the artificially designed information needs are not the participants' real needs. Study 3 then serves as a complement to Study 1 by adopting the critical incident technique in the qualitative interview to unfold laypeople's own OHIS stories, including the real needs that motivate laypeople's OHIS.

## 10.0 Conclusion

### 10.1 Limitations

Despite the benefits and rationale of the research design, it is critical to understand that this study has several limitations.

First, one should be aware that the current study was conducted on laypeople and the OvCa community, two particular health consumer groups. Consequently, the results of this dissertation might be biased towards specific consumer groups. In addition, the convenience sampling used to recruit participants and the relatively small sample size in each consumer group may further hurt the generalizability and transferability of the results.

Second, the inherent limitations of the research methods adopted in each sub-study should be noted, though the data triangulation was expected to offset some of the limitations. For example, given the limitations of controlled user studies, the usage of some types of sources might be affected in Study 1. The “save to workspace” function has the benefit of capturing source adoption behavior, but it may also introduce unnatural factors into this study — in real life, people might take other actions to adopt a source, such as taking notes, texting friends or just trying to remember the information. In Study 2, meanwhile, the analysis did not take into account the patients’ disease trajectories, which could potentially affect their information needs and levels of learning objectives. Besides, it is notable that women living with gynaecological cancer have various supportive care needs, such as interpersonal/intimacy needs and daily living needs (Maguire, Kotronoulas, Simpson, & Paterson, 2015), however, in study 2, the OvCa-related topics were only classified based on the information needs. Additionally, in the two qualitative interviews, the comments on the outcomes of OHIS were collected based on participants’ self-reports, which were not validated or directly observed by the researcher.

Third, the measurement of consumers’ knowledge state and knowledge change in this study is limited by its scope. Overall, only three representative cognitive learning levels are selected to elicit participants’ search as learning behaviors and assess the learning ob-

jectives in the information needs. Such a design is based on the rationale explained in the methodology, but one should recognize that consumers' learning and knowledge evolution could be oversimplified. In addition, the measurement of knowledge change is conducted within a certain context, though the researcher has attempted to use both unobtrusive and obtrusive methods to collect data. For instance, in Study 2, little is known about how much the participants have learned via other information sources beyond the information seeking and support in the sampled OHC.

## 10.2 Future Work

In addition to the findings regarding consumers' search as learning in OHIS, several other interesting questions that are worth future research arise from this study.

First, given the scope of the current study, much attention has been devoted to knowledge change among a variety of outcomes of OHIS. However, it is notable that beyond knowledge change, there are other types of cognitive outcomes, such as informed decision making and perception of control and coping (Lambert & Loiselle, 2007). Moreover, emotional outcomes and behavioral outcomes are frequently mentioned by the participants in the two interviews. What is noteworthy about these outcomes is that many negative emotions and concerning behaviors are discovered. As an example, several OvCa patients interviewed told us that they followed online nutrition instructions to prepare their food, a practice which can be risky. Laypeople participants were also found to use online information to decide whether or not to see a doctor or merely follow the online instructions. In spite of the benefits of OHIS, consumers should be aware that OHIS is supposed to be a complement to, rather than a replacement for, offline face-to-face interaction with healthcare providers (Zeng et al., 2004). However, little is known about the connections between OHIS and the negative emotional outcomes and debatable behavioral outcomes. Revealing the connections among these aspects would shed new light on how to better design OHIS tools and systems and empower consumers with a healthier online-offline transition.

Second, as mentioned in the discussion of limitations, the current study did not dis-

tinguish the OvCa patients according to their disease trajectory given the scarce data in NOCC and the convenience sampling in the interview. But there exists literature to suggest that OvCa patients' information needs change with the disease trajectory (Simacek, Raja, Chiauzzi, Eek, & Halling, 2017). The question is then whether or not trajectory has a significant effect on the learning objectives and behaviors. If so, the extent of that effect should be determined. The answer to this question may help researchers and clinicians to design interventions that better support OvCa patients along their disease trajectory. This is an interesting and promising direction for future research.

Third, health conditions (severe vs. mild) are found to be critical factors that influence laypeople's search behaviors and source selection behaviors, and the interaction effects between health conditions and learning objectives are also discovered on many behavioral measures. This highlights the importance of task-related factors in modeling consumers' search as learning behavior in OHIS. Although more studies are needed to examine the generality of this finding, it does suggest that other aspects of health conditions besides severity could be examined for their influence in the future.

### 10.3 Conclusion

This dissertation is one of the first to investigate health consumers' OHIS from a cognitive learning perspective.

A conceptual framework of health information search as learning (HearSAL) is proposed by combining the work from two areas: online health information seeking (OHIS) and search as learning (SAL). The framework is used to guide the design of the four-studies and is verified by the empirical evidence. Laypeople and OvCa patients are selected as two representative health consumer groups.

Based on the synthesized results of the four sub-studies, it can be concluded that health consumers are able to increase knowledge through OHIS, and that the searching and learning process features unique method and information dimensions.

With respect to the method dimension, health consumers' search as learning behaviors

and sources vary with different levels of learning objectives, and the variation is affected by the severity of health conditions. As for the information dimension, consumers' expressed information needs are characterized by underlying learning objectives of different cognitive levels. The characteristics of the needs and the types of informational support at each level are identified. Consequently, consumers' knowledge change in OHIS varies by levels of learning objectives, health conditions, and the amount of support received. Implications to system design are provided by reflecting consumers' needs, advancing search tools, and promoting desirable outcomes. The theoretical and methodological implications are further summarized. Finally, limitations and future directions are discussed.

## Appendix A

### User Study Introductory Script

The purpose of this research study is to investigate how users, especially laypeople seek health-related information online. For that reason, we are going to run a scenario-based user study and a follow-up interview. You will be asked to complete a brief questionnaire before, during and after each search task. You will conduct 2 search tasks in this study (approximately 30 min one task), and each task is broken down to 3 sessions. When you are searching to complete the task, measures relating to information seeking behavior will be collected. Your activities such as reading text, annotating passages, opening document, search queries, etc. will be logged and analyzed afterwards in order to study your behavior and performance. After the main session, a short interview will be conducted. The Mac built-in audio recorder will be used to record the interview. The whole study will take approximately 1.5 hours.

All participants must be 18 years of age or older. If you are willing to participate, our questionnaire will ask about your background (e.g., age, academic status, gender), as well as your level of familiarity with web search. The foreseeable risks are breach of confidentiality which is an inherent risk in most research studies and the fatigue for study longer than one hour. To deal with the risks, a subject number will be assigned to you as the only identifier. The identifier will be saved separately from the record dataset, and results will be kept well and safely. We will arrange a short break between the tasks to reduce fatigue. You will get \$12.5 compensation per hour for your participation.

Your participation is voluntary, and you may withdraw from this project at any time, without any penalty or loss of benefits. This study is being conducted by Yu Chi, who can be reached at [yuc73@pitt.edu](mailto:yuc73@pitt.edu), if you have any questions.

Authorized representatives of the University of Pittsburgh Research Conduct and Compliance Office may review your identifiable research information for the purpose of monitoring

the appropriate conduct of this research study. You may contact the Human Subjects Protection Advocate of the IRB Office, University of Pittsburgh (1-866-212-2668) to discuss problems, concerns, and questions; obtain information; offer input; or discuss situations that have occurred during participation.

Do you consent to participant in this study?

Please check:

Yes

No

## Appendix B

### User Study Entry Questionnaire

1. Subject ID: \_\_\_\_\_

2. Your gender: \_\_\_\_\_

3. Your age: \_\_\_\_\_

4. Your program of study:

Undergraduate

Graduate

Others

\_\_\_\_\_

5. Your major: \_\_\_\_\_

6. How would you rate your search experience? (5 indicates strongly agree):

I am very experienced in online search. 1 2 3 4 5

I search online very frequently. 1 2 3 4 5

For the next several questions, I would like to ask you for your opinion and about your experience using the Internet to search for health-related information. Health-related information can be the information about a disease, condition, nutrition, everyday healthy living, etc.

7. When was the last time you search health-related information online?

Within last week

Within last month

within last 6 months

within last year

More than 1 year ago

8. How often do you search the Internet for health-related information?



- Daily
- Weekly
- Biweekly
- Monthly
- Less than monthly

9. How useful do you feel the Internet is in helping you in making decisions about your health?

- Not useful at all
- Not useful
- Unsure
- Useful
- Very useful

10. How important is it for you to be able to access health resources on the Internet?

- Not important at all
- Not important
- Unsure
- Important
- Very important

10. What resources have you used for seeking health-related information?

- Search engine, e.g., Google: (Rarely) 1 2 3 4 5 (Frequently)
- Health portal, e.g., MedlinePlus: (Rarely) 1 2 3 4 5 (Frequently)
- wikis, e.g., wikipedia: (Rarely) 1 2 3 4 5 (Frequently)
- Online communities, e.g., Patientlikeme: (Rarely) 1 2 3 4 5 (Frequently)
- Social media, e.g., Facebook: (Rarely) 1 2 3 4 5 (Frequently)
- Other online resources: \_\_\_\_\_

## Appendix C

### Laypeople Interview Protocol

#### Reflection on the search tasks (asked based on observation)

- For the two health conditions and six tasks you just searched, Which one(s) do you is most challenging? Why?
- Did you use any secret technique/trick to solve the task?
- Based on your impression, what kind of website did you go? Were they helpful? Why or why not?
- Did you get stuck in anywhere? How did you solve it?

#### Participant's attitude towards OHIS

- What is your impression when thinking of the online health information? Why?
- What kind of health-related information do you think people might want to search online?
- Who will search online more frequently for health-related information than others?
- Would you like to share your health-related knowledge to others online? Why or why not?

#### Participants' personal OHIS experience

Did you ever search health-related information on the Internet? Would you like to share a real experience/story to me, for example, your most recent or most memorable OHIS experience? You don't need to tell me the exact condition you searched for if you find it uncomfortable, and my focus is only your search process. (Or, I noticed that in the entry questionnaire, you replied last month was the last time you searched for health information,

can we talk more about that? You don't need to tell me the specific condition you search for, because our focus is your search process.)

- (Optional, only when not covered) Why do you want to search for it?
- How did you search for it?
- Could you briefly recall; what kind of website you visited? Were they helpful? If yes, could you name at least one? If no, why not?
- How do you tell a website or the information is good or not?
- How did you feel before and after the search or during the search process?
- Did the search task involve a decision making activity? If yes, what decision did you make after seeking the information? Did you take any action? Why or why not?

## Appendix D

### OvCa Patients and Caregivers Interview Protocol

#### Background Information

- When were you diagnosed with ovarian cancer?
- What stage was the cancer when you were diagnosed?
- What treatments have you received: (check all that apply – surgery, chemotherapy, radiation therapy, immune therapy, hormone therapy)
- What is the status of the cancer now?
- Do you live with family? If so, who would that be? Who in the family are helping you? Do you have another caregiver? If so, who is the caregiver? How often does the caregiver help you?
- How would you say that your overall knowledge about the ovarian cancer is... (Likert scale? 0-10? Not at all to expert?)

#### Search Experience

- What are one or more topics that you would want to know more about?
- Why do you want to know about these topics?
- How would you say that your current knowledge about topic XXX [answer of the Q6] is... (Likert scale? 0-10? Not at all to expert?)
- Generally, how do you find information for topic XXX? Can you describe in details the steps you would take to find the information? - Prompts: do you search on your own or someone else will search for you?
- Do you have particular sources that you go to get information about cancer? How did you learn about these sources? Were they useful? Why or why not?
- How do you judge whether the information is relevant/important/trustworthy or not? What criteria do you use to judge relevance?

- How do you plan to use the information you get about topic XXX?

### **Perspectives on the Recommendations**

- We have selected online information that seems helpful for you. Could you take a look and let us know if the recommended information meets your needs?
- Could you tell us whether this suggested information is helpful/useful for you?
- Why do you think it is useful/ not useful to you?
- Is the recommendation relevant to your current situation? Could you explain how the recommendation is relevant to your current situation?
- Would you recommend this information to others with the same problem?
- How would you say the knowledge level in this recommended article is... (Likert scale? 0-10? Not at all to expert?)
- Do you think the recommended information is too hard or too easy for you? Or just at the right level? Why?
- How often would you want to get follow-up recommendations? - Prompts: any time when there is a new relevant information, every time they come for a treatment, once every month, or others
- Do you think you would use the recommendation in your daily life?
- Could you explain how you will use the information to manage your cancer?

## References

- Agosti, M., Fuhr, N., Toms, E., & Vakkari, P. (2014). Evaluation methodologies in information retrieval dagstuhl seminar 13441. In *Acm sigir forum* (Vol. 48, pp. 36–41).
- Ahmad, F., Hudak, P. L., Bercovitz, K., Hollenberg, E., & Levinson, W. (2006). Are physicians ready for patients with internet-based health information? *Journal of medical internet research*, 8(3).
- Allan, J., Croft, B., Moffat, A., & Sanderson, M. (2012). Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012 the second strategic workshop on information retrieval in lorne. In *Acm sigir forum* (Vol. 46, pp. 2–32).
- Anderson, C., Carter, J., Nattress, K., Beale, P., Philp, S., Harrison, J., & Juraskova, I. (2011). "the booklet helped me not to panic": a pilot of a decision aid for asymptomatic women with ovarian cancer and with rising ca-125 levels. *International Journal of Gynecologic Cancer*, 21(4), 737–743.
- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (2001). A taxonomy for learning, "teaching, and assessing-a revision of bloom's taxonomy of educational objectives", (eds.) addison wesley longman. *New York*.
- Anker, A. E., Reinhart, A. M., & Feeley, T. H. (2011). Health information seeking: a review of measures and methods. *Patient education and counseling*, 82(3), 346–354.
- Åström, F. (2007). Changes in the lis research front: Time-sliced cocitation analyses of lis journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology*, 58(7), 947–957.
- Ayers, S. L., & Kronenfeld, J. J. (2007). Chronic illness and health-seeking information on the internet. *Health*, 11(3), 327–347.
- Baggott, R., Allsop, J., & Jones, K. (2014). *Speaking for patients and carers: health consumer groups and the policy process*. Macmillan International Higher Education.
- Barsevick, A. M., & Johnson, J. E. (1990). Preference for information and involvement,

- information seeking and emotional responses of women undergoing colposcopy. *Research in Nursing & Health*, 13(1), 1–7.
- Bekker, H. L., Winterbottom, A. E., Butow, P., Dillard, A. J., Feldman-Stewart, D., Fowler, F. J., ... Volk, R. J. (2013). Do personal stories make patient decision aids more effective? a critical review of theory and evidence. *BMC medical informatics and decision making*, 13(S2), S9.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1), 133–143.
- Bhattacharya, N., & Gwizdka, J. (2019). Measuring learning during search: differences in interactions, eye-gaze, and semantic similarity to expert knowledge. In *Proceedings of the 2019 conference on human information interaction and retrieval* (pp. 63–71).
- Bhavnani, S. K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *Chi'02 extended abstracts on human factors in computing systems* (pp. 610–611).
- Bloom, B. S., et al. (1956). Taxonomy of educational objectives. vol. 1: Cognitive domain. *New York: McKay*, 20–24.
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of documentation*, 56(1), 71–90.
- Brennan, K., Kelly, D., & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *proceedings of the 5th information interaction in context symposium* (pp. 165–174).
- Broom, A. (2005). Virtually he@lthy: the impact of internet use on disease experience and the doctor-patient relationship. *Qualitative health research*, 15(3), 325–345.
- Bundorf, M. K., Wagner, T. H., Singer, S. J., & Baker, L. C. (2006). Who searches the internet for health information? *Health services research*, 41(3p1), 819–836.
- Cancer statistics*. (n.d.). National Cancer Institute. Retrieved from <https://www.cancer.gov/about-cancer/understanding/statistics>
- Cartright, M.-A., White, R. W., & Horvitz, E. (2011). Intentions and attention in exploratory health search. In *Proceedings of the 34th international acm sigir conference on research and development in information retrieval* (pp. 65–74).

- Case, D. O., Johnson, J. D., Andrews, J. E., Allard, S. L., & Kelly, K. M. (2004). From two-step flow to the internet: The changing array of sources for genetics information seeking. *Journal of the American Society for Information Science and Technology*, 55(8), 660–669.
- Chen, A. T. (2021). Interactions between affect, cognition, and information behavior in the context of fibromyalgia. *Journal of the Association for Information Science and Technology*.
- Chen, Y., Zhao, Y., & Wang, Z. (2020). Understanding online health information consumers' search as a learning process. *Library Hi Tech*.
- Chi, Y., Han, S., He, D., & Meng, R. (2016). Exploring knowledge learning in collaborative information seeking process. In *Ceur workshop proceedings* (Vol. 1647).
- Chi, Y., He, D., & Jeng, W. (2020). Laypeople's source selection in online health information-seeking process. *Journal of the Association for Information Science and Technology*.
- Chou, W.-y. S., Liu, B., Post, S., & Hesse, B. (2011). Health-related internet use among cancer survivors: data from the health information national trends survey, 2003–2008. *Journal of Cancer Survivorship*, 5(3), 263–270.
- Chuang, K. Y., & Yang, C. C. (2014). Informational support exchanges using different computer-mediated communication formats in a social media alcoholism community. *Journal of the Association for Information Science and Technology*, 65(1), 37–52.
- Chung, J. E. (2014). Social networking in online support groups for health: how online social networking benefits patients. *Journal of health communication*, 19(6), 639–659.
- Cline, R. J., & Haynes, K. M. (2001). Consumer health information seeking on the internet: the state of the art. *Health education research*, 16(6), 671–692.
- Collins-Thompson, K., Hansen, P., & Hauff, C. (2017). Search as learning (dagstuhl seminar 17092). In *Dagstuhl reports* (Vol. 7).
- Collins-Thompson, K., Rieh, S. Y., Haynes, C. C., & Syed, R. (2016). Assessing learning outcomes in web search: A comparison of tasks and query strategies. In *Proceedings of the 2016 acm on conference on human information interaction and retrieval* (pp. 163–172).
- Courtright, C. (2005). Health information-seeking among latino newcomers: an exploratory



- study. *Information Research*, 10(2), 10–2.
- Creswel, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. *Los angeles: University of Nebraska–Lincoln*.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Csala, D. (2014). *Sankey diagram generator based on the sankey plugin for d3 by mike bostock*. Retrieved from <https://sankey.csaladen.es/>
- Cutrona, S. L., Mazor, K. M., Vieux, S. N., Luger, T. M., Volkman, J. E., & Rutten, L. J. F. (2015). Health information-seeking on behalf of others: characteristics of “surrogate seekers”. *Journal of cancer education*, 30(1), 12–19.
- Davis, S. N., O’Malley, D. M., Bator, A., Ohman-Strickland, P., & Hudson, S. V. (2020, Feb). Correlates of information seeking behaviors and experiences among adult cancer survivors in the usa. *Journal of Cancer Education*. doi: 10.1007/s13187-020-01758-6
- De Choudhury, M., Morris, M. R., & White, R. W. (2014). Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1365–1376).
- Denlinger, C. S., Carlson, R. W., Are, M., Baker, K. S., Davis, E., Edge, S. B., ... others (2014). Survivorship: introduction and definition. *Journal of the National Comprehensive Cancer Network*, 12(1), 34–45.
- Dervin, B. (1983). An overview of sense-making research: concepts, methods and results to date..
- Donovan, H. S., Hartenbach, E. M., & Method, M. W. (2005). Patient–provider communication and perceived control for women experiencing multiple symptoms associated with ovarian cancer. *Gynecologic Oncology*, 99(2), 404–411.
- Eickhoff, C., Teevan, J., White, R., & Dumais, S. (2014). Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of the 7th acm international conference on web search and data mining* (pp. 223–232).
- Ekwall, E., Ternstedt, B.-M., Sorbe, B., & Graneheim, U. H. (2011). Patients’ perceptions of communication with the health care team during chemotherapy for the first recurrence of ovarian cancer. *European Journal of Oncology Nursing*, 15(1), 53–58.

- El Sherif, R., Pluye, P., Thoër, C., & Rodriguez, C. (2018). Reducing negative outcomes of online consumer health information: Qualitative interpretive study with clinicians, librarians, and consumers. *Journal of medical Internet research*, *20*(5).
- Ewerth, R., Dietze, S., Hoppe, A., & Yu, R. (2019). Salm'19: First international workshop on search as learning with multimedia information. In *Proceedings of the 27th acm international conference on multimedia* (pp. 2724–2725).
- Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, *324*(7337), 573–577.
- Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Jama*, *287*(20), 2691–2700.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin*, *51*(4), 327.
- Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New media & society*, *9*(2), 319–342.
- Flanagin, A. J., Metzger, M. J., & Hartsell, E. (2010). *Kids and credibility: An empirical examination of youth, digital media use, and information credibility*. MIT Press.
- Fox, S. (2013). *Health fact sheet, pew research centre*.
- Fox, S., & Purcell, K. (2010). *Chronic disease and the internet*. Pew Internet & American Life Project Washington, DC.
- Freimuth, V. S., Stein, J. A., & Kean, T. J. (1989). *Searching for health information: The cancer information service model*. University of Pennsylvania Press.
- Ghosh, S., Rath, M., & Shah, C. (2018). Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 conference on human information interaction & retrieval* (pp. 22–31).
- Gibbons, M. C., Fleisher, L., Slamon, R. E., Bass, S., Kandadai, V., & Beck, J. R. (2011). Exploring the potential of web 2.0 to address health disparities. *Journal of health communication*, *16*(sup1), 77–89.
- Gill, P. S., & Whisnant, B. (2012). A qualitative assessment of an online support community

- for ovarian cancer patients. *Patient related outcome measures*, 3, 51.
- Goldner, M. (2006). How health status impacts the types of information consumers seek online. *Information, Community & Society*, 9(6), 693–713.
- Grandinetti, D. A. (2000). Doctors and the web. help your patients surf the net safely. *Medical economics*, 77(5), 186–8.
- Griesbaum, J., Mahrholz, N., von Löwe Kiedrowski, K., & Rittberger, M. (2015). Knowledge generation in online forums: a case study in the german educational domain. *Aslib Journal of Information Management*.
- Groen, W. G., Kuijpers, W., Oldenburg, H. S., Wouters, M. W., Aaronson, N. K., & van Harten, W. H. (2015). Empowerment of cancer survivors through information technology: an integrative review. *Journal of medical Internet research*, 17(11), e270.
- Gwizdka, J., Hansen, P., Hauff, C., He, J., & Kando, N. (2016). Proceedings of the second international workshop on search as learning. In *Second international workshop on search as learning-co-located with the 39th international acm sigir conference on research and development in information retrieval (sigir 2016), pisa, italy, july 21st, 2016*.
- Hagan, T. L., & Donovan, H. S. (2013). Ovarian cancer survivors' experiences of self-advocacy: A focus group study. In *Oncology nursing forum* (Vol. 40, p. 140).
- Halkett, G., Kristjanson, L., Lobb, E., O'driscoll, C., Taylor, M., & Spry, N. (2010). Meeting breast cancer patients' information needs during radiotherapy: what can we do to improve the information and support that is currently provided? *European Journal of Cancer Care*, 19(4), 538–547.
- Han, S. (2018). *Understanding, modeling and supporting cross-device web search* (Unpublished doctoral dissertation). University of Pittsburgh.
- Hanbury, A. (2012). Medical information retrieval: an instance of domain-specific search. In *Proceedings of the 35th international acm sigir conference on research and development in information retrieval* (pp. 1191–1192).
- Hansen, D. L., Derry, H. A., Resnick, P. J., & Richardson, C. R. (2003). Adolescents searching for health information on the internet: an observational study. *Journal of medical Internet research*, 5(4), e25.
- Hansen, P., & Rieh, S. Y. (2016). *Recent advances on searching as learning: An introduction*

- to the special issue*. SAGE Publications Sage UK: London, England.
- Harkin, L. J., Beaver, K., Dey, P., & Choong, K. (2017). Navigating cancer using online communities: a grounded theory of survivor and family experiences. *Journal of Cancer Survivorship*, *11*(6), 658–669.
- Hartoonian, N., Ormseth, S. R., Hanson, E. R., Bantum, E. O., & Owen, J. E. (2014). Information-seeking in cancer survivors: application of the comprehensive model of information seeking to hints 2007 data. *Journal of health communication*, *19*(11), 1308–1325.
- Hayes, G. R., Abowd, G. D., Davis, J. S., Blount, M. L., Ebling, M., & Mynatt, E. D. (2008). Opportunities for pervasive computing in chronic cancer care. In *International conference on pervasive computing* (pp. 262–279).
- Hölscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer networks*, *33*(1-6), 337–346.
- Hong, Y., Pena-Purcell, N. C., & Ory, M. G. (2012). Outcomes of online support and resources for cancer survivors: a systematic literature review. *Patient education and counseling*, *86*(3), 288–296.
- Hoppe, A., Holtz, P., Kammerer, Y., Yu, R., Dietze, S., & Ewerth, R. (2018). Current challenges for studying search as learning processes. *Proceedings of Learning and Education with Web Data, Amsterdam, Netherlands*.
- Houlihan, M. C., & Tariman, J. D. (2017). Comparison of outcome measures for traditional and online support groups for breast cancer patients: An integrative literature review. *Journal of the advanced practitioner in oncology*, *8*(4), 348.
- Howlader, N., Noone, A., Krapcho, M., Noone, A., Neyman, N., Aminou, R., ... others (2012). *Seer cancer statistics review, 1975–2009 (vintage 2009 populations)*.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277–1288.
- Huh, J., Kwon, B. C., Kim, S.-H., Lee, S., Choo, J., Kim, J., ... Yi, J. S. (2016). Personas in online health communities. *Journal of biomedical informatics*, *63*, 212–225.
- Jacobs, M., Johnson, J., & Mynatt, E. D. (2018). Mypath: Investigating breast cancer patients’ use of personalized health information. *Proceedings of the ACM on Human-*

*Computer Interaction*, 2(CSCW), 1–21.

- Jansen, B. J., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management*, 45(6), 643–663.
- Jenkins, C., Corritore, C. L., & Wiedenbeck, S. (2003). Patterns of information seeking on the web: A qualitative study of domain expertise and web expertise. *IT & society*, 1(3), 64–89.
- Jiang, J., He, D., & Allan, J. (2014). Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. In *Proceedings of the 37th international acm sigir conference on research & development in information retrieval* (pp. 607–616).
- Jiang, T., Liu, F., & Chi, Y. (2015). Online information encountering: modeling the process and influencing factors. *Journal of Documentation*.
- Johnson, J. D. (1997). *Cancer-related information seeking*. Hampton Press.
- Johnson, J. D. (2003). On contexts of information seeking. *Information Processing & Management*, 39(5), 735–760.
- Jong, S. T., & Drummond, M. J. (2016). Exploring online fitness culture and young females. *Leisure Studies*, 35(6), 758–770.
- Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009). Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 625–634).
- Kanaris, I., & Stamatatos, E. (2009). Learning to recognize webpage genres. *Information Processing & Management*, 45(5), 499–512.
- Kelly, D., Arguello, J., Edwards, A., & Wu, W.-c. (2015). Development and evaluation of search tasks for iir experiments using a cognitive complexity framework. In *Proceedings of the 2015 international conference on the theory of information retrieval* (pp. 101–110).
- Kharazmi, S., Karimi, S., Scholer, F., & Clark, A. (2014). A study of querying behaviour of expert and non-expert users of biomedical search systems. In *Proceedings of the 2014 australasian document computing symposium* (pp. 10–17).
- Kim, H., Park, S.-Y., & Bozeman, I. (2011). Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health

- experts. *Health Information & Libraries Journal*, 28(3), 188–199.
- Kintsch, W., & Walter Kintsch, C. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.
- Koch-Weser, S., Bradshaw, Y. S., Gualtieri, L., & Gallagher, S. S. (2010). The internet as a health information source: findings from the 2007 health information national trends survey and implications for health communication. *Journal of health communication*, 15(sup3), 279–293.
- Krikelas, J. (1983). Information-seeking behavior: Patterns and concepts. *Drexel library quarterly*, 19(2), 5–20.
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5), 361–371.
- Lambert, S. D., & Loiselle, C. G. (2007). Health information—seeking behavior. *Qualitative health research*, 17(8), 1006–1019.
- Lauckner, C., & Hsieh, G. (2013). The presentation of health-related search results and its impact on negative emotional outcomes. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 333–342).
- Lee, C.-J. (2009). The role of internet engagement in the health-knowledge gap. *Journal of Broadcasting & Electronic Media*, 53(3), 365–382.
- Lenz, E. R. (1984). Information seeking: a component of client decisions and health behavior. *Advances in Nursing Science*.
- Liu, J., Belkin, N. J., Zhang, X., & Yuan, X. (2013). Examining users' knowledge change in the task completion process. *Information Processing & Management*, 49(5), 1058–1074.
- Liu, J., Liu, C., & Belkin, N. J. (2016). Predicting information searchers' topic knowledge at different search stages. *Journal of the Association for Information Science and Technology*, 67(11), 2652–2666.
- Liu, N., Tong, Y., & Chan, H. C. (2017). Information seeking in online healthcare communities: the dual influence from social self and personal self. *IEEE Transactions on Engineering Management*, 64(4), 529–538.
- Longo, D. R. (2005). Understanding health information, communication, and information seeking of patients and consumers: a comprehensive and integrated model. *Health*

- Expectations*, 8(3), 189–194.
- Lu, L., & Yuan, Y. C. (2011). Shall i google it or ask the competent villain down the hall? the moderating role of information need in information source selection. *Journal of the American Society for Information Science and Technology*, 62(1), 133–145.
- Lu, Y., & Hsiao, I.-H. (2017). Personalized information seeking assistant (pisa): from programming information seeking to learning. *Information Retrieval Journal*, 20(5), 433–455.
- Madathil, K. C., Greenstein, J. S., Juang, K. A., Neyens, D. M., & Gramopadhye, A. K. (2013). An investigation of the informational needs of ovarian cancer patients and their supporters. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, pp. 748–752).
- Maddock, C., Lewis, I., Ahmad, K., & Sullivan, R. (2011). Online information needs of cancer patients and their organizations. *Ecancermedicalscience*, 5.
- Maguire, R., Kotronoulas, G., Simpson, M., & Paterson, C. (2015). A systematic review of the supportive care needs of women living with and beyond cervical cancer. *Gynecologic oncology*, 136(3), 478–490.
- Marcella, R., Rowlands, H., & Baxter, G. (2013). The critical incident technique as a tool for gathering data as part of a qualitative study of information seeking behaviour..
- Marchionini, G. (1997). *Information seeking in electronic environments* (No. 9). Cambridge university press.
- Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41–46.
- Mayer, D. K., Terrin, N. C., Kreps, G. L., Menon, U., McCance, K., Parsons, S. K., & Mooney, K. H. (2007). Cancer survivors information seeking behaviors: a comparison of survivors who do and do not seek information about cancer. *Patient education and counseling*, 65(3), 342–350.
- McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England journal of medicine*, 338(3), 171–179.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.

- McMullan, M. (2006). Patients using the internet to obtain health information: how this affects the patient–health professional relationship. *Patient education and counseling*, *63*(1-2), 24–28.
- Metzger, M. J., & Flanagin, A. J. (2015). Psychological approaches to credibility assessment online. *The handbook of the psychology of communication technology*, *32*, 445.
- Mistry, A., Wilson, S., Priestman, T., Damery, S., & Haque, M. (2010). How do the information needs of cancer patients differ at different stages of the cancer journey? a cross-sectional survey. *JRSM short reports*, *1*(4), 1–10.
- Nagler, R. H., Gray, S. W., Romantan, A., Kelly, B. J., DeMichele, A., Armstrong, K., . . . Hornik, R. C. (2010). Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. *Patient education and counseling*, *81*, S54–S62.
- NCI. (2018). Cancer stat facts: ovarian cancer.
- Ogan, C. L., Ozakca, M., & Groshek, J. (2008). Embedding the internet in the lives of college students: Online and offline behavior. *Social Science Computer Review*, *26*(2), 170–177.
- Oh, Y. S. (2016). *Predictors of online health information seeking behavior and health information seeking experience of elderly cancer survivors using the internet* (Unpublished doctoral dissertation). Case Western Reserve University.
- Ovarian cancer*. (n.d.). U.S. Department of Health and Human Services. Retrieved from <https://rarediseases.info.nih.gov/diseases/7295/ovarian-cancer>
- Pang, P. C.-I., Chang, S., Pearce, J. M., & Verspoor, K. (2014). Online health information seeking behaviour: Understanding different search approaches. In *Pacis* (p. 229).
- Pang, P. C.-I., Chang, S., Verspoor, K., & Pearce, J. (2016). Designing health websites based on users’ web-based information-seeking behaviors: A mixed-method observational study. *Journal of medical Internet research*, *18*(6), e145.
- Pang, P. C.-I., Verspoor, K., Pearce, J., & Chang, S. (2015). Better health explorer: Designing for health information seekers. In *Proceedings of the annual meeting of the australian special interest group for computer human interaction* (pp. 588–597).
- Piaget, J. (1976). Piaget’s theory. In *Piaget and his school* (pp. 11–23). Springer.



- Pian, W., Khoo, C. S., & Chang, Y.-K. (2016). The criteria people use in relevance decisions on health information: an analysis of user eye movements when browsing a health discussion forum. *Journal of medical Internet research*, *18*(6), e136.
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review*, *106*(4), 643.
- Pozzar, R. A., & Berry, D. L. (2019). Preserving oneself in the face of uncertainty: A grounded theory study of women with ovarian cancer. In *Oncology nursing forum* (Vol. 46, pp. 595–603).
- Ramsey, I., Corsini, N., Peters, M. D., & Eckert, M. (2017). A rapid review of consumer health information needs and preferences. *Patient education and counseling*, *100*(9), 1634–1642.
- Reid, F., Bhatla, N., Oza, A. M., Blank, S. V., Cohen, R., Adams, T., ... others (2021). The world ovarian cancer coalition every woman study: identifying challenges and opportunities to improve survival and quality of life. *International Journal of Gynecologic Cancer*, *31*(2).
- Rice, R. E. (2006). Influences, usage, and outcomes of internet health information searching: multivariate results from the pew surveys. *International journal of medical informatics*, *75*(1), 8–28.
- Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H.-J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, *42*(1), 19–34.
- Robinson, T. N., Patrick, K., Eng, T. R., Gustafson, D., et al. (1998). An evidence-based approach to interactive health communication: a challenge to medicine in the information age. *Jama*, *280*(14), 1264–1269.
- Rumelhart, D. E., & Norman, D. A. (1976, Jan). Accretion, tuning and restructuring: Three modes of learning.  
doi: 10.21236/ada030406
- Rutten, L. J. F., Arora, N. K., Bakos, A. D., Aziz, N., & Rowland, J. (2005). Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). *Patient education and counseling*, *57*(3), 250–261.
- Rutten, L. J. F., Squiers, L., & Hesse, B. (2006). Cancer-related information seeking: hints

- from the 2003 health information national trends survey (hints). *Journal of health communication*, 11(s1), 147–156.
- Savolainen, R. (2011). Requesting and providing information in blogs and internet discussion forums. *Journal of Documentation*.
- Seçkin, G. (2010). Cyber patients surfing the medical web: Computer-mediated medical knowledge and perceived benefits. *Computers in Human Behavior*, 26(6), 1694–1700.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7–30.
- Silberg, W. M., Lundberg, G. D., & Musacchio, R. A. (1997). Assessing, controlling, and assuring the quality of medical information on the internet: Caveant lector et viewer—let the reader and viewer beware. *Jama*, 277(15), 1244–1245.
- Simacek, K., Raja, P., Chiauzzi, E., Eek, D., & Halling, K. (2017). What do ovarian cancer patients expect from treatment?: perspectives from an online patient community. *Cancer nursing*, 40(5), E17–E27.
- Stewart, D. E., Wong, F., Cheung, A., Dancey, J., Meana, M., Cameron, J., . . . Rosen, B. (2000). Information needs and decisional preferences among women with ovarian cancer. *Gynecologic oncology*, 77(3), 357–361.
- Stvilia, B., Mon, L., & Yi, Y. J. (2009). A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology*, 60(9), 1781–1791.
- Sundar, S. S. (2008). *The main model: A heuristic approach to understanding technology effects on credibility*. MacArthur Foundation Digital Media and Learning Initiative.
- Suri, V. R., Majid, S., Chang, Y.-K., & Foo, S. (2016). Assessing the influence of health literacy on health information behaviors: a multi-domain skills-based approach. *Patient education and counseling*, 99(6), 1038–1045.
- Tao, D., LeRouge, C., Smith, K. J., & De Leo, G. (2017). Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. *JMIR human factors*, 4(4), e25.
- Tardy, R. W., & Hale, C. L. (1998). Getting “plugged in”: a network analysis of health-information seeking among “stay-at-home moms”. *Communications Mono-*

*graphs*, 65(4), 336–357.

- Thomas, P., Bailey, P., Moffat, A., & Scholer, F. (2015). Towards nuanced system evaluation based on implicit user expectations. In *Airs* (pp. 332–344).
- Thon, F. M., & Jucks, R. (2017). Believing in expertise: How authors' credentials and language use influence the credibility of online health information. *Health communication*, 32(7), 828–836.
- Urban, R. R., He, H., Alfonso, R., Hardesty, M. M., Gray, H. J., & Goff, B. A. (2016). Ovarian cancer outcomes: Predictors of early death. *Gynecologic oncology*, 140(3), 474–480.
- Urquhart, C., Light, A., Thomas, R., Barker, A., Yeoman, A., Cooper, J., ... Spink, S. (2003). Critical incident technique and explicitation interviewing in studies of information behavior. *Library & Information Science Research*, 25(1), 63–88.
- Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1), 7–18.
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information processing & management*, 39(3), 445–463.
- van Eenbergen, M. C., van de Poll-Franse, L. V., Heine, P., & Mols, F. (2017). The impact of participation in online cancer communities on patient reported outcomes: systematic review. *JMIR cancer*, 3(2), e15.
- Viera, A. J., Garrett, J. M., et al. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5), 360–363.
- Wang, X., High, A., Wang, X., & Zhao, K. (2021). Predicting users' continued engagement in online health communities from the quantity and quality of received support. *Journal of the Association for Information Science and Technology*, 72(6), 710–722.
- Wang, Y.-C., Kraut, R., & Levine, J. M. (2012). To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the acm 2012 conference on computer supported cooperative work* (pp. 833–842).
- Wang, Y.-C., Kraut, R. E., & Levine, J. M. (2015). Eliciting and receiving online support:

- using computer-aided content analysis to examine the dynamics of online social support. *Journal of medical Internet research*, 17(4), e99.
- Weaver III, J. B., Mays, D., Weaver, S. S., Hopkins, G. L., Eroğlu, D., & Bernhardt, J. M. (2010). Health information-seeking behaviors, health indicators, and health risks. *American journal of public health*, 100(8), 1520–1525.
- White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second acm international conference on web search and data mining* (pp. 132–141).
- White, R. W., & Horvitz, E. (2009). Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)*, 27(4), 23.
- Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the american society for information science and technology*, 55(3), 246–258.
- Wildemuth, B. M., Kelly, D., Boettcher, E., Moore, E., & Dimitrova, G. (2018). Examining the impact of domain and cognitive complexity on query formulation and reformulation. *Information Processing & Management*, 54(3), 433–450.
- Wilson, M. J., & Wilson, M. L. (2013). A comparison of techniques for measuring sense-making and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology*, 64(2), 291–306.
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of documentation*, 55(3), 249–270.
- Wilson, T. D. (2000). Human information behavior. *Informing science*, 3(2), 49–56.
- Wu, W.-C., Kelly, D., Edwards, A., & Arguello, J. (2012). Grannies, tanning beds, tattoos and nascar: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th information interaction in context symposium* (pp. 254–257).
- Xie, B., Wang, M., Feldman, R., & Zhou, L. (2010). Health information and decision-making preferences in the internet age: a pilot study using the health information wants (hiw) questionnaire. In *Proceedings of the 1st acm international health informatics symposium* (pp. 610–619).
- Xing, W., Goggins, S., & Introne, J. (2018). Quantifying the effect of informational sup-

- port on membership retention in online communities through large-scale data analytics. *Computers in Human Behavior*, 86, 227–234.
- Yang, D., Kraut, R. E., Smith, T., Mayfield, E., & Jurafsky, D. (2019). Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–14).
- Yang, D., Yao, Z., Seering, J., & Kraut, R. (2019). The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–15).
- Yang, M., & Xu, X. (2019). Understanding the patient health information behavior to guide their health self-management: Findings from online lupus erythematosus communities..
- Yilma, T. M., Inthiran, A., Reidpath, D., & Orimaye, S. O. (2017). Health information seeking and its associated factors among university students: a case in a middle-income setting. In *Pacific asia conference on information systems (pacis)*.
- Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P., & Dietze, S. (2018). Predicting user knowledge gain in informational search sessions. In *The 41st international acm sigir conference on research & development in information retrieval* (pp. 75–84).
- Zeng, Q. T., Kogan, S., Plovnick, R. M., Crowell, J., Lacroix, E.-M., & Greenes, R. A. (2004). Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. *International journal of medical informatics*, 73(1), 45–55.
- Zhang, J. (2019). *Supporting diabetes patient decisional needs through online health communities* (Unpublished doctoral dissertation). UC San Diego.
- Zhang, P., & Soergel, D. (2016). Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, 42(1), 59–78.
- Zhang, Y. (2012). Consumer health information searching process in real life settings. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–10.
- Zhang, Y. (2014). Beyond quality and accessibility: Source selection in consumer health

- information searching. *Journal of the Association for Information Science and Technology*, 65(5), 911–927.
- Zhang, Y., Sun, Y., & Kim, Y. (2017). The influence of individual differences on consumer's selection of online sources for health information. *Computers in Human Behavior*, 67, 303–312.
- Zhang, Y., Sun, Y., & Xie, B. (2015). Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. *Journal of the Association for Information Science and Technology*, 66(10), 2071–2084.
- Zimmermann, M., & Jucks, R. (2018). How experts' use of medical technical jargon in different types of online health forums affects perceived information credibility: Randomized experiment with laypersons. *Journal of medical Internet research*, 20(1), e30.