Wikipedia: An unexplored resource for understanding consumer health information behavior in library and information science scholarship

Abstract
Approach: The author draws on current health information behaviour models and relevant theories from existing library and information science literature and applies them to propose a new definition of consumer health information behaviour. The author uses this definition to frame Wikipedia as an unexplored consumer health information resource in the library and information science scholarship and suggests future directions for placing such investigations within a conceptual framework from LIS.

Purpose: To date, health information behaviour models have not been applied to an exploration of Wikipedia as a consumer health information resource. Wikipedia has been situated and is well established as a valuable resource for the general layperson wishing to learn more about their health or the health of a loved one. This paper aims to identify an approach to exploring the role of Wikipedia in consumer health information behaviour that is grounded in a conceptual framework from the LIS discipline.

Findings: The paper finds that Longo's Expanded Conceptual Model of Health Information Seeking Behaviour could be valuable and useful for the exploration of consumer health information behaviour in relation to Wikipedia's health and medical content. Due to Wikipedia's online nature, research framed by this model must acknowledge and take under consideration the digital divide phenomenon and various factors that influence an individual's place within it.

Research Implications: This work builds a foundation upon which future research into the role of Wikipedia's health and medical content in CHIB can be grounded. Using Longo's model, future research might provide insight into who Wikipedia is helping and who it has left behind. LIS scholars, practicing health librarians and perhaps health workers stand to gain a deeper understanding of the potential influence of Wikipedia's health information on its consumers.

Value: This paper is novel in the fact that a HIB model has not yet been applied to the study of Wikipedia’s health content. This paper provides a foundation for this research.

Introduction
For decades, library and information science (LIS) scholars have grappled with defining and understanding the concept of information behaviour. While the scholarly conversation around information behaviour is popularly documented to have been initiated by Thomas D. Wilson (1981) in the library and information science (LIS) discipline, the effort to understand information needs, use, users and systems has been an ongoing discussion within LIS for more than seventy years (Wilson, 2006, p. 658n). Agreement on concepts such as information behaviour are sensitive to the nuanced interpretations of individual scholars. It is well documented that this disunity is not necessarily a result of direct disagreement, but rather, a noticeable deficiency of reflexive scholarship, wherein terms like information behaviour are used without elaboration. To convolute matters further, associated terminology such as “need”, “seek” and “use” are also varied in their application. For example, among LIS scholars alone, “information need” can be understood as a desire or motivation to seek answers (Taylor, 1968), to reduce one’s uncertainty (Belkin, 1980; Kuhlthau, 1993), or to make sense of a given situation or thing (Dervin et al., 2003).
However, health information behaviour (HIB) studies have less history than information behavior more broadly and in fact, as recent as 20 years ago, HIB studies were rarely grounded in any library and information science theory at all (Baker & Pettigrew, 1999, p. 444). With the emergence of a global pandemic and the ubiquitous nature of online health information, and misinformation, understanding consumer health information behaviour of those who actively use the internet as a source of health information is more important now than ever.

When it was suggested that the advent of the internet would initiate the democratization of health knowledge (Hardey, 1999, p. 827), the dramatic influence Wikipedia would have on the dissemination of health information worldwide was neither known nor predicted. In today’s online world there exists a wealth of consumer health information websites that are easy to access and use. MayoClinic and MedlinePlus are two examples of authoritative and credible consumer health information web sites with content that is produced and maintained by health experts, easy to read and understand, and do not employ advertising (Content Development Policy - About This Site - Mayo Clinic, 2020; MedlinePlus, 2020). However, Wikipedia is the most frequently accessed free health information source on the internet. It is consulted by approximately 50%-70% of physicians and more than 90% of medical students (Heilman & West, 2015). Wikipedia’s goal is to accomplish what Hardey grappled with at the cusp of the 21st century – ensure that every person has access to “the sum of all medical knowledge in their own language” (Wiki Project Med - Meta, 2020). The argument has been made that since Wikipedia’s content is so extensively read, “its authorship and reliability are important qualities” (Heilman & West, 2015) to provide and maintain. A recent study demonstrated that people with lower education levels accept Wikipedia with little critical analysis, compared to more educated individuals who tended to view Wikipedia through a critical eye or avoid it altogether (Huisman et al., 2020, pp. 7–8). This is an important concern when the context of health disparities and the digital divide is considered. It is no surprise then, especially given that Wikipedia is community-built by editors who can remain anonymous, that a review of the literature about Wikipedia’s health content found that over half the literature is concerned with assessing the quality or reliability of Wikipedia’s health related content. (Smith, 2020, pp. 6–10).

This article will consider whether any theories or models of HIB, grounded in LIS, might be suitable to develop a nuanced understanding of consumer health information behaviour on the internet. Specifically, among individuals who interact with Wikipedia, which has been situated in the literature as a key consumer health information source (Heilman et al., 2011; Shafee et al., 2017; Smith, 2020). While not a scoping or systematic literature review, a discussion of previous articulations of CHIB, underlying theoretical frameworks, and relevant literature is important to provide the appropriate context for proposed future research.

**Consumer health information behaviour (CHIB)**

In an early exploration of consumer health information behaviour (CHIB) Elizabeth Lenz (1984) focuses on the search for health information and its relationship with consumer health decision making and consumer health behaviour. Lenz proposes that “the information search process is conceptualized as a subcomponent of the decision process, which precedes the enactment of discretionary health behavior”
This is perhaps the most helpful early conceptualization of CHIB when the goal is to consider it in the simplest possible fashion – that information seeking in any form is part of decision making, whatever that decision might be. It is now known that HIB in general, and CHIB specifically, are more nuanced than Lenz’s definition suggests. Fisher and Julien (2009) dedicate space to CHIB in their comprehensive ARIS&T review (Fisher & Julien, 2009, pp. 22–24), but the most robust review of HIB literature analyzes the maturity of health information-seeking behaviour (HISB) as a concept (Lambert & Loiselle, 2007). This work was subsequently revisited and updated. (Zimmerman & Shaw, 2020).

The core outcome of Lambert and Loiselle’s review and concept analysis is a characterization of health information seeking behaviour (HISB) despite a noticeable lack of an “apparent dominant definition” (Lambert & Loiselle, 2007, p. 1008) for the term and a limited use of theoretical frameworks grounded in LIS (Lambert & Loiselle, 2007, p. 1015). Despite more than a decade of research since Lambert and Loiselle’s concept analysis, their critique still stands. Zimmerman and Shaw (2020), find that the concept of HISB continues to be poorly developed. An investigation of the commonalities between 15 definitions found in 85 relevant articles exposed the ambiguous definition of HISB as “an active or purposeful behaviour undertaken by an individual with the objective of finding information about health” (Zimmerman & Shaw, 2020, p. 4). This definition is limited to the act of seeking information and does not provide space for a broader understanding of CHIB that could include active and passive information behaviour, or relevant cognitive, social, or behavioural outcomes of health information behaviour.

The implication of Lambert and Loiselle’s work, and Zimmerman’s update of the same, is that the study of HISB broadly, and by extension CHIB, applies theories and models – in the instances it employs them – to investigate a concept for which consensus on a definition does not exist. Those who attempt to review existing definitions find themselves weaving together a definition derived from various texts (Zare-Farashbandi & Lalazaryan, 2014, p. 193). In an attempt to rectify this the author intends to explore HIB as: the key variables that influence the seeking, receipt, avoidance, sharing, management, or use of health information and the outcomes that result from the behaviour. CHIB is a contextual limitation where members of the general public are the individuals engaging in HIB, or where resources designed for consumption by the layperson or general public are under consideration. In alignment with Savolainen (2019) the author also considers information seeking behaviour as a constituent of information behaviour (Savolainen, 2019, p. 115) and as such, HISB is a constituent of HIB. The author’s definition does not assume that HIB is always borne from a clear or distinct information need. For example, “information need” can be understood as a desire or motivation to seek answers (Taylor, 1968), to reduce one’s uncertainty (Belkin, 1980; Kuhlthau, 1993), or to make sense of a given situation or thing (Dervin et al., 2003). The author’s definition instead focuses on how HIB can be affected by the interplay between various cognitive, affective, physical or social factors (Abrahamson et al., 2008, p. 7).

**Theories of health information behaviour**

While the use of theory to understand information behaviour has been questioned (Robson & Robinson, 2013), theories of HIB attempt to explain the behaviour of individuals based on data that has been collected in empirical investigations. Several theories of HIB are grounded in research from other disciplines. For example, coping and adaptation theory is grounded in an attempt to understand the
actions undertaken by an individual in a stressful environment or situation (Lazarus & Folkman, 1984, p. 289), and monitoring/blunting has been discussed within the context of managing anxiety or fear by reducing uncertainty (D. Case & Given, 2016, p. 88; D. O. Case et al., 2005, pp. 355, 359). There is also a distinct conversation about motivations and drives in health information seeking, in which uncertainty, typically tied to feelings of anxiety, is a possible motive to avoid health information. (D. O. Case et al., 2005, p. 359). Similarly, Bodie and Dutta describe health motivated consumers as “health information oriented individuals”. These individuals possess an intrinsic interest in health issues or concerns and have a tendency to actively monitor their environment for health information. Conversely, those with low motivation may have a tendency to passively receive health information if it comes along (Bodie & Dutta, 2008, p. 190). Genuis (2012) uses Social Positioning Theory (SPT) to examine the self-positioning of women information seekers and its impact on their decision making about hormone replacement therapy and menopause (Genuis, 2012, p. 556). Genuis identifies three primary positions that represent “how women identified themselves in relation to formal and informal information sources and draws parallels between how women position themselves and the decision-making model they engage in with their health care practitioner (Genuis, 2012, p. 563). Drawing on Maslow’s hierarchy of needs, Boot and Meijman suggest that “people search for information in order to retrieve factual information or opinions (knowledge), belong to a group (social), to reduce uncertainty, just for fun (entertainment) or for self-actualization” (Boot C.R. & Meijman F.J., 2010, p. 154; parentheses added). They take into account individuals who do not engage in health information behaviour. They theorize that “the underlying social structure affects an individual’s... intrinsic interest in health and the perceived ability to use the Internet for health purposes” (Bodie & Dutta, 2008, p. 189).

Differential health information-seeking behaviour (Lambert et al., 2009, p. 17) integrates some of the theories already discussed. The core variable in of differential health information-seeking behaviour is “playing my part and taking care of myself” (Lambert et al., 2009, p. 15). This variable resonates social positioning theory wherein the cancer patients position themselves as personally accountable for managing their health and exercise a degree of autonomy in that regard. “Playing my part and taking care of myself”, considered on a scale from “high importance given to information-seeking” and “low importance given to information seeking” (Lambert et al., 2009, p. 17) appears to be a nod to monitoring and blunting or motivations and drives. Finally, Lambert et al. acknowledge that variation in the HIB of the cancer patients in their study could be a reflection of diverse socio-economic demographics of the participants in their study. For example, the group of participants identified as “intensive information seekers” were more likely to favour the internet over other sources of health information (Lambert et al., 2009, p. 22). This group of intensive information-seekers were identified early as “dominant” talkers in focus groups and all of held a university degree (Lambert et al., 2009, p. 17). Whether intentional or not – this study provides a glimpse into the impact the digital divide on one’s ability or willingness to position themselves as personally accountable, or autonomous, when it comes to managing their health or finding, seeking, or using health information.

**Online consumer health information and the digital divide**

At the end of the twentieth century it was claimed that the internet had the potential to dissolve “the boundaries around areas of expertise upon which the [health] professions derived much of their power”
(Hardey, 1999, p. 827). It was not acknowledged, however, that the “exposure of the health profession’s knowledge base to the public gaze” (Hardey, 1999, p. 832) might only be available to members of the public with the privilege of access. In their discussion of antecedents to health information seeking, the factors that “focus on the imperatives that motivate someone to seek answers to questions” (Johnson & Case, 2012c, p. 46), Johnson and Case describe the “classic profile” of high information seekers as “white, middle-aged women who are members of high socio-economic status (SES) groups and also highly educated” (Johnson & Case, 2012c, p. 47). They also note an association between the haves and have-nots in the information gap and those who are most at risk of not being able to access health care (Johnson & Case, 2012b, p. 140). With an ever-increasing reliance on internet-based technology to disseminate consumer health information, the disparities in our population are predicted to grow (Bodie & Dutta, 2008, p. 177) as those without access to these technologies, in conjunction with fewer skills or less knowledge and experience, are shut out from benefitting. Here lies the irony of the inverse information law – that “access to information is particularly difficult for those who need it most. (Boot & Meijman, 2010, p. 145)“.

Consequently, it is important to assume that existing studies of CHIB in the online environment might only represent the segment of our population that firmly belong to the “haves” group. Those who have access to the Internet first and foremost, but beyond that, those who have the basic skills required to feel comfortable using the Internet as a resource for health information. This understanding of CHIB research allows one to make sense of Case and Johnson’s identification of health information seekers as predominantly white, middle-class, highly educated women. Additionally, this knowledge should be considered an incentive to broaden the scope of CHIB to those who do not use online health information in order to develop a better understanding of whether this practice is a choice, or the result of barriers borne from socioeconomic, cognitive, physical, or affective factors.

**CHIB for the study of Wikipedia as a health information resource**

While there is some evidence to suggest that online health information is often used in conjunction with other health information sources, such as physicians, Wikipedia’s influence is not yet fully understood. It has been assumed that “the public relies on free online medical information for making health decisions” (Shafee et al., 2017a, p. 1122) but it has not yet been documented whether there are outcomes, such as decision making, related to the CHIB of individuals who access Wikipedia’s health and medical content.

While the literature has positioned Wikipedia as a key consumer health resource, it is insufficient to advocate for the improvement and growth of Wikipedia’s health content based on only the knowledge that it is accessed frequently and the assumption that online health information is relied on for health decision-making. An investigation into the CHIB of individuals who use Wikipedia’s health and medical content could produce evidence that would provide LIS scholars, health experts, and Wikipedia advocates, particularly Wiki Project Medicine, with a richer understanding of health information seeking, sharing, use, or encounters on Wikipedia and the outcomes, if any, of the CHIB of Wikipedia users. Further, it could improve our current understanding of how Wikipedia is perceived by the consumer.
As proposed in the introduction, the author has defined CHIB as a concept that seeks to understand the key variables that influence the seeking, receipt of, avoidance, sharing, management, or use of health information and the outcomes that result from the behaviour of the lay person or where resources designed for or used by the general lay person are under investigation. In the context of online health information, there are three existing and ongoing conversations.

First, there is consideration of whether or not the lay person uses the internet to search for, retrieve, or use health information and if so, which specific sources are consulted (Cotten & Gupta, 2004, p. 1799). A 2013 Pew Research Center study indicated that 72% of U.S adults who use the internet have used it to look for health information and overall, 35% of Americans have gone online at least once to “figure out what medical condition they or someone else might have” (Pew Research Center’s Internet and American Life Project, 2013, p. 6).

Second, the question of who is using the internet when engaging in CHIB. A 2017 study supported Johnson and Cases’ profile of the health typical information seeker with the finding that individuals with “more education, higher [socioeconomic status], being younger, and having internet skill were factors associated with health information seeking on the internet” (Jacobs et al., 2017, p. 8). While this profile is limited to active seekers of health information, global health disparities based on socio-economic status, gender, ethnicity, language, or simply geography are widely known. Subsequently, it makes sense that these same disparities would present themselves in explorations of online CHIB.

The Digital Divide – while accurate in its concern with inequitable access to internet and communication technologies, plays a much larger role than its baseline definition suggests. In addition to the need to ensure all groups have equitable access to internet technologies and devices, it is of equal importance “to present information that can be understood... by a population that is less prone to search for information in general and less able to comprehend that information when found” (Bodie & Dutta, 2008, p. 178). Only once online consumer health information is accessible in myriad ways, can the gap created by the digital divide and the consequential health disparities begin to be addressed.

Third, efforts to understand the motivation for online CHIB has yielded evidence that this behaviour is not intended to replace the expert knowledge of health care providers. It has altered the nature of the patient-physician relationship and has encouraged the advent of shared decision-making in health care. CHIB is considered complementary to the expertise of health care providers by offering individuals a sense of empowerment over the decisions made with respect to their health and health care or the care of their families (Anker et al., 2011, p. 347; Powell et al., 2011). Patients have sought information online to “seek clarity and confirmatory information in depth” either before or after a consultation with a medical practitioner, for reassurance, or simply to take advantage of the anonymity of the web (Powell et al., 2011). These behaviours have been reported in these types of investigations repeatedly and consistently, particularly within the context of the effect of the availability of online health information on the patient-physician relationship. Lenz, in the 1980s, indicated that information-seeking is a part of health decision making, so it follows that with the ubiquity of health information on the internet, patients might be compelled to take a more active role in managing their health. Furthermore,
groups that are underrepresented in health care might feel they must be self-reliant when considering a health decision, which can present as consulting multiple sources of information, including those online, not just the health care provider (Wathen, 2006, p. 482).

Wikipedia seeks to dismantle some of the barriers to health information by making health information available to everyone, everywhere, regardless of their background. Whether or not it can overcome socioeconomic factors influencing CHIB has not yet been documented. Moreover, of the three academic conversations discussed above, Wikipedia’s place has not yet been clearly defined. As such, it is imperative to investigate Wikipedia’s place in CHIB. It would be particularly beneficial to consider Wikipedia and CHIB through the lens of the digital divide: what factors motivate or enable individuals to include Wikipedia in their CHIB? Is there a relationship between one’s place in the digital divide and whether Wikipedia is actively sought, passively encountered, or avoided altogether? While there are several possible applications of the theories and models in LIS to look for answers to the above questions, two particularly stand out as useful: Updated Integrative Model of eHealth Use and Longo’s Expanded Conceptual Model of Health Information Seeking Behaviors.

**Using HIB models to investigate the HIB of Wikipedia users**

Much work has been done to review the existing models of HIB documented in the literature. To avoid duplication of these efforts, Table I. LIS Models of HIB, documents each known model in chronological order. The models included in Table 1 were gathered from review articles, primary studies, and survey books (Anker et al., 2011; Baker & Pettigrew, 1999; D. Case & Given, 2016; Johnson & Case, 2012a; Lambert & Loiselle, 2007; Marton & Wei Choo, 2012; Zare-Farashbandi & Lalazaryan, 2014; Zimmerman & Shaw, 2020). Johnson and Case (2012) have done similar work (Johnson & Case, 2012a), and there is overlap with Case & Given (2016). The table is not a comprehensive list of all information behaviour models in LIS, but is limited to models of health information behaviour that specifically consider the general consumer.

The Updated Integrative Model of eHealth Use (Bodie & Dutta, 2008) draws on the potential impact of the digital divide on CHIB and is framed within ehealth marketing. This model explicitly considers the context and acknowledges the influence of an individual’s situation, personality, demographic and internet use history as factors that may influence their HIB. It also indicates that beyond these factors sits health literacy and computer literacy as additional factors that can influence CHIB. The presentation of their model includes important demographic factors correlated with health literacy, which itself is determined by the characteristics of the seekers themselves such as their situation, personality, demographics and internet use history (Bodie & Dutta, 2008, p. 189). Bodie and Dutta’s model aligns with other models in its understanding that individual characteristics can predict CHIB, and it attests that before one engages in CHIB two interrelated groups of factors should be considered as possible predictors of CHIB: (1) health literacy and computer literacy, and; (2) motivation to use the internet for health information and perceived ability to use the internet for health information (Bodie & Dutta, 2008, p. 191). They propose that these two inter-related pathways could lead to “a whole host of health outcomes” (Bodie & Dutta, 2008). This model could be applied to develop a richer understanding of the delicate relationship between health outcomes, personal characteristics, ability, motivation, and
health literacy in the context of understanding the role of Wikipedia’s health and medical content in CHIB. While the Updated Integrative Model of eHealth Use is built with an awareness of the determinants that place individuals on one side or another of the digital divide, it is not built from, nor does it appear to be influenced by, models and theories of HIB currently engaged with in LIS.

On the other hand, Longo’s Expanded Conceptual Model of Health Information Seeking Behaviors (ECMHISB) (Longo, 2005; Longo et al., 2010) is influenced by the Health Belief Model from social psychology that has influenced the development of other models of HIB (Johnson & Case, 2012c), and it targets patients and other general consumers of health information specifically. The ECMHISB divides variables that influence patient/consumer information seeking behaviour into two distinct categories: contextual (e.g. health status, information environment) and personal (e.g. demographics, socioeconomic status, language) (Longo, 2005, p. 192). These two classifications of variables allow for consideration of a vast number of potential variables that might influence CHIB among either active seekers or passive recipients. Finally, the ECMHISB includes the possible outcomes of CHIB such as empowerment, satisfaction, activities of daily living and health outcomes.

The ECMHISB has been applied in empirical studies (Longo et al., 2010, p. 335), and has been shown to be appropriate for use when applied to diabetes self-management. In one study it was found that diabetes patients both passively receive health information from news or other media, and they actively “weave their own information web” using internet sources, print media such as books, and advice from health care practitioners (Longo et al., 2010, p. 337). As a result of this study, Longo not only noticed the important role of health literacy in influencing CHIB (Longo et al., 2010, p. 338), he also updated the ECMHISB model to account for the “non-linear interplay of both active information-seeking behavior and passive information receipt, as well as the important role relationships play as the patients incorporate the information that works for them” (Longo et al., 2010, p. 339).

Employed in acknowledgement and awareness of the digital divide phenomenon, Longo’s Expanded Conceptual Model of Health Information Seeking Behaviours (Longo, 2005, p. 192) creates space to consider both contextual factors, including those of an individual’s information environment, and personal factors, including demographics, socioeconomic status and education level as potential predictors of CHIB. Longo’s model also provides space for both active information seeking and the passive receipt of information – an important distinction to consider in an investigation of the CHIB of Wikipedia users. Finally, Longo’s model culminates with a consideration of the outcomes of the CHIB that can be either cognitive or behavioural, including a sense of empowerment, as has been reported (Anker et al., 2011; Powell et al., 2011). Dissimilar to Bodie and Dutta’s Updated Integrative Model of eHealth Use, Longo’s model is influenced by previous LIS work in HIB research and it has been developed and improved upon based on empirical evidence.

Since little is known about the CHIB of Wikipedia users, Longo’s model permits a broad approach to developing an initial understanding of Wikipedia use through this viewpoint. It could be applied to considerations of the contextual or personal factors that might be related to Wikipedia use, whether there are any patterns in the phases of information use, and what outcomes could potentially result from
Wikipedia use. Further, Longo’s model facilitates an investigation into whether there is a relationship between personal and contextual factors and the outcomes reported by Wikipedia users. These factors, considered through the lens of the digital divide as a determinant of disparities in health, offer the potential to understand Wikipedia’s role in the landscape of online health information as a free, publicly available and easy to read resource, its impact on those who use it actively or passively engage with it, and whether it stands to close the digital divide or intensify it.

**Future research**
Using Longo’s ECMHISB as a conceptual framework, the author is prepared to conduct a mixed methods qualitative study to explore the role of Wikipedia in CHIB as it intersects with Bodie and Dutta’s model of the key factors that influence one’s place in the digital divide. This study will be framed by the digital divide as a social determinant of health and as such, will explore the following questions:

1. Who has access (in all aspects of the term) to Wikipedia? Of those who access it, how is Wikipedia’s medical and health content encountered (e.g., passively or actively)?
2. Are there benefits or concerns to having access to Wikipedia’s health content and if yes, what are they? Is there a relationship between these benefits or concerns and the social, physical, personal, or cognitive factors that can place an individual on either end of the digital divide?

This investigation can potentially provide insight into who Wikipedia is helping and who it has left behind. This knowledge can build upon work already done to situate Wikipedia as a valuable health information resource (Smith, 2020) in the promotion of global health (Heilman et al., 2011), a growing resource of reliable health information (Brezar & Heilman, 2019; Heilman & West, 2015; Shafee et al., 2017a), and attract more contributors who wish to improve its health and medical pages. For LIS scholars, this study will be novel in the fact that a HIB model has not yet been applied to the study of Wikipedia’s health content. Finally, practicing health librarians and perhaps health workers stand to gain a deeper understanding of the potential influence of Wikipedia’s health information on its consumers.

**Conclusion**
In 2019 Savolainen called out the “need to extend the research setting by examining how information seeking, encountering, and sharing are related to information use particularly in social media forums such as question and answer (Q&A) sites, blogs and online discussion groups and ResearchGate” (Savolainen, 2019, p. 532). Wikipedia’s health and medical content is a crowd-sourced online resource that has yet to be explored in depth using a CHIB framework from the LIS discipline. An overview of theories and models of CHIB has demonstrated the availability of appropriate models upon which to frame this kind of examination. It has also demonstrated, however, a clear lack of consensus on how consumer health information behaviour is defined, conceptualized, or operationalized. This paper has taken existing definitions under consideration and forged a new one to apply to the author’s future research.
References


