2021-08-01

Evaluation of Patient Experience Using Natural Language Processing Algorithms

Sofia Veronica Ortega
University of Texas at El Paso

EVALUATION OF PATIENT EXPERIENCE USING NATURAL LANGUAGE PROCESSING ALGORITHMS

SOFIA VERONICA ORTEGA HARO

Master's Program in Manufacturing Engineering

APPROVED:

Sreenath Chalil Madathil, Ph.D., Chair

Thenral Mangadu, Ph.D.

Tzu-Liang (Bill) Tseng, Ph.D.

Sergio Luna, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

Dedication

To Mom, Dad, Chuy and Rodrigo:

Your love keeps me grounded, yet it gives me the courage to soar toward my dreams.

To my Nayla, Michelle, Karime, Jose and David:

Thank you for your love and support during all stages of my life.

EVALUATION OF PATIENT EXPERIENCE USING NATURAL LANGUAGE PROCESSING ALGORITHMS

by

SOFIA VERONICA ORTEGA HARO

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

DEPARTMENT OF INDUSTRIAL, MANUFACTURING AND SYSTEMS ENGINEERING

THE UNIVERSITY OF TEXAS AT EL PASO

August 2021

Acknowledgements

I would like to express my sincere gratitude to Dr. Chalil Madathil, my thesis advisor for all his guidance, patience, and assistance. His continuous support and enthusiasm has motivated me to write this thesis. Thank you for giving me the opportunity, especially as an international student, to be involved in research activities. Special thanks as well to Dr. Tseng, Dr. Mangadu and Dr. Luna, my committee members, for all your suggestions and support.

I would also like to offer a special thanks to UTEP, my dearest Alma Mater. I would not be the person and engineer I am today if it were not for all the opportunities UTEP has given me.

Abstract

INTRODUCTION: Healthcare organizations are making extensive efforts to improve the patient experience. Enhancing patient/client experience and outcomes is crucial for patient-centered care and can reveal improvement opportunities. Healthcare settings currently rely on surveys (e.g., HCAHPS) and patient feedback to measure patient experience. Studies have identified that utilizing patient journey mapping can better capture patient experience throughout all stages of the patient's journey and provide quality and process improvement recommendations at specific hotspots. However, these measurement techniques are time-consuming and resource intensive.

AIM: This research aims to measure patient experience of breast cancer patients from social media data using natural language processing algorithms.

METHODS: This study analyzes data obtained from social media (e.g., Twitter and Reddit) referent to breast cancer. Natural Language Processing (NLP) algorithms were applied to identify latent topics via Latent Dirichlet Allocation (LDA) and sentiments via Sentiment Analysis (SA) associated to specific hotspots.

DISCUSSION: The use of AI to capture patient experience during the patient's journey in the healthcare continuum provides valuable insights to improve individualized, empathetic, and respectful care in clinical systems. Such patient experience information is invaluable for healthcare quality improvement efforts and improving patient-centered care.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The COVID-19 pandemic has caused fundamental shifts in healthcare systems. Some of these shifts include patient's increasing participation in decision-making; the rapid adoption of technological and digital innovations; the advancement for interoperable data and data analytics; and an unprecedented public and private collaboration [1]. Patients are driving and accelerating the pace at which healthcare systems operate [1]. These trends, in addition to the increase in health engagement, enhanced using health monitoring devices and the wish of a trusted clinician relationship are emphasizing the importance of a patient-centered care at all levels.

1.1 Background

The Institute of Medicine (IOM) defines Patient-centered care as "providing care that is respectful of, and responsive to, individual patient preferences, needs and values, and ensuring that patient values guide all clinical decisions." [2]. According to IOM, understanding patient experience is crucial for delivering patient-centered care. The term "Patient Experience" has been defined by the Beryl Institute as "the sum of all interactions, shaped by an organization's culture, that influence patient perceptions across the continuum of care" [3]. Measures of a patient-centered care or patient experience are becoming increasingly requested by both government and non-government payers, as well as by public reporting and health care ranking organizations [4].

Furthermore, it is also imperative to explain the terms in the definition of "patient experience" such as interaction, culture, perception, and continuum of care. Firstly, 'interaction' indicates the orchestrated touchpoints of people, processes, policies, communication, and environment. Secondly, 'culture' implies the vision, and values of people at all levels of the organization and community. Thirdly, 'perceptions' refers to what is recognized, understood, and remembered by patients and support people; perceptions may vary based on individual experiences such as beliefs, values and cultural background. Lastly, 'continuum of care' is the before, during and after the delivery of care. Moreover, one study reported that hospitalized patients see an average of 17.8 health professionals during a single hospitalization [5], which can create a complex and confusing system to capture patient experience, reliably.

Patient satisfaction surveys provide baseline information about patient experience from which quality can be assessed, analyzed, and interpreted to determine ways in which a service can be developed. Several techniques such as patient satisfaction surveys, Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) surveys [3], Press Ganey Surveys [4], semi-structured interviews, patient-reported experience/outcome measures (PREMs/PROMs) [6], and patient journey mapping (PJM) [7] can measure patient experience. Nonetheless, it is currently unclear whether these types of tools really reflect what is important for patients.

1.2 MOTIVATION

Ever since the Institute of Medicine (IOM) published "*Crossing the Quality Chasm: A New Health System for the 21st Century*" in 2001[2], the attention to patient-centered care has been accentuated. This, coupled with fundamental shifts arisen from the COVID-19 pandemic, have caused patients to drive and be more participative in their care than ever before. Their preferences are pushing the improvement and development of digitally enabled and on-demand connected interactions with clinicians, and their expectations are driving a transition from a health care encounter to a holistic human-centered experience [1]. Providing these patient experiences can only be accomplished by collaborating with patients to understand their perspectives and needs towards their care. Existing methods of measuring patient experience are resource-intensive, time consuming, and cumbersome. Other issues in patient satisfaction and experience measurement are: validity and reliability of surveys; approach (qualitative vs. quantitative); survey design and format; administration technique (in-person, telephone surveys, self-administered, online); and, timing [8]. With the advent of machine learning and artificial intelligence, such efforts can be automated to improve the staff productivity, provide better care perception to the patients, identify gaps in healthcare service perception by the patients, and improve overall healthcare.

1.3 Problem Statement

Current PREMs rely on simple feedback with a low number of participants. This is unlikely to produce significant improvements in care, making these measurements time and resource-intensive. Automating and improving a way to measure patient experience with Natural Language Processing Algorithms following the Patients' Journey Maps will reduce the time and resources spent on patient experience measurements.

1.4 Thesis Objectives

The main objective of this thesis is to implement Machine Learning and Natural Language Processing Algorithms to efficiently capture patient experiences of breast cancer patients using data from social media, i.e., Twitter and Reddit. Furthermore, this data will be analyzed to evaluate patient experiences and sentiments at various hotspots of the breast cancer patients' journey. After identifying patient experiences and sentiments are evaluated, recommendations will be provided targeting specific hotspots.

1.5 Organization of the Thesis

This thesis is structured in six chapters. Chapter 1 starts with the introduction and background, motivation, problem statement, and thesis objectives. Chapter 2 continues with the literature review on patient experience, text analytics, and social media. This chapter also includes the study contributions. Chapter 3 describes the methodology followed and includes the research questions, study, and model design and outlines the proposed model. Chapter 4 includes the results from the model deployment with an emphasis on the topic modeling, sentiment analysis and the relation of the comments with specific hotspots. Chapter 5 covers the discussion and findings from the analyses and includes social media and study limitations. Finally, Chapter 6 contains the conclusions and recommendations for future work.

Chapter 2: Literature Review

A systematic review on patient experience was conducted to understand the key elements that constitute patient experience, how has patient journey mapping been used to capture this experience, and to summarize tools and techniques that capture patient experiences. The review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to summarize the article selection process used for this systematic review and synthesize current knowledge on patient experience using patient journey mapping (PJM) as shown in figure 2.1. This section also includes a literature review on text analytics with an especial focus on topic modeling and sentiment analysis. Additionally, research on social media emphasizing the social networks of Twitter and Reddit is also included in this literature review.

2.1 Bibliographic Search Process

This search queried four databases: Embase, ScienceDirect, PubMed and CINAHL using keywords such as “Patient Journey”, “Patient Journey Mapping”, “Journey Mapping”, “Patient Experience”, “Process Mapping”, “Patient-centered Approach” and “Patient-centric Approach”. Table 2.1 summarizes the web search results. A total of 5092 records were identified through database searching and screened for eligibility.

Table 2.1: Initial web search results.

	"Patient Journey"	**"Patient Journey Mapping"**	**"Journey Mapping"**	**"Patient Experience"**	**"Process Mapping"**	**"Patient Centered Approach"**	**"Patient Centric Approach"**
Embase	1438	20	52	11660	776	1466	166
ScienceDirect	2419	10	109	101463	2659	4919	4919
PubMed	578	8	27	6669	325	1079	106
CINAHL	449	9	25	4457	182	664	56

2.2 Inclusion and Exclusion Criteria

This study includes all articles published from April 3, 2003, to September 17, 2020. The articles comprised by this systematic review examine the patient experience and the use of patient journey mapping as a tool for capturing patient experience.

From the 5092 total articles found, 1309 duplicates were removed. The 3783 remaining records were title-screened. The keywords used for inclusion in this screening phase were: "patient journey", "interview(s)", "survey(s)", "mapping", "satisfaction", "questionnaire(s)", "patient experience", "semi-structured", "care pathway", "process mapping", "patient-centric", "patient-centered", "patient journey mapping", "patient-centered approach" and "patient interview". Conference poster abstracts, articles without access to the whole paper, book sections, foreign-language articles, articles not dealing with patient experience, studies measuring journey mapping exclusively from provider/clinician perspective, and animal experiment records were excluded from the search, resulting in a total of 3548 excluded records. A total number of 233 full-text articles were then assessed for eligibility. There were 138 records in which the abstract was irrelevant to the topic, thus excluded. The final number of studies included in the systematic review was 95 papers, as illustrated in figure 2.1.

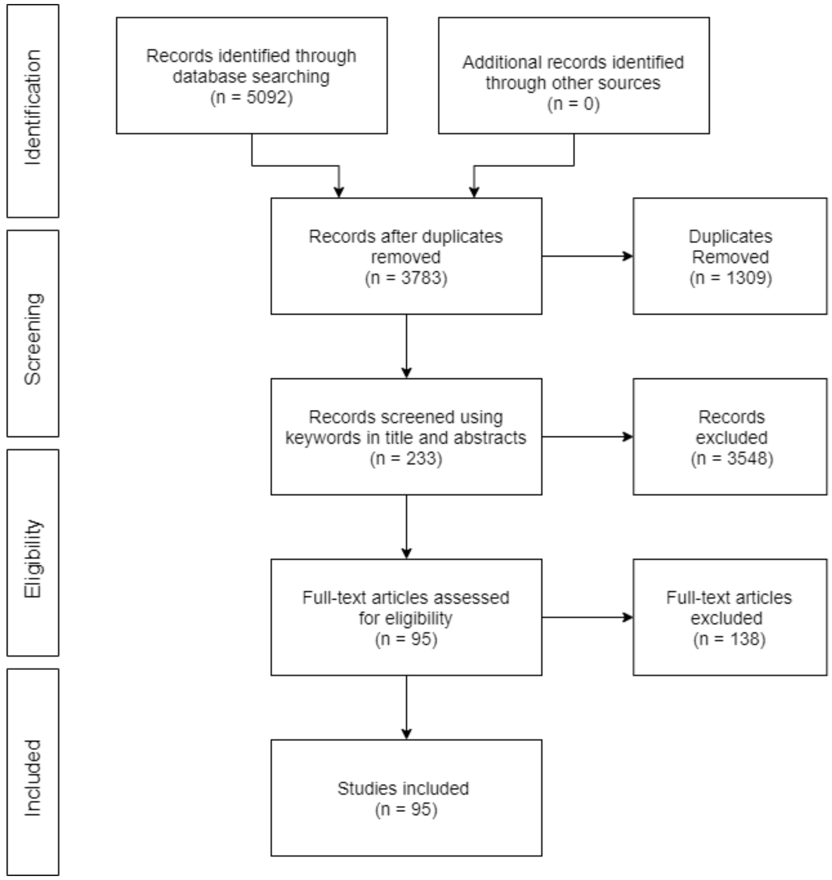


Figure 2.1: PRISMA guidelines for article selection

As shown in Figure 2.2, there has been an increasing interest in patient experience-related topics since 2003. For this systematic review, United Kingdom (28) has been the leading research country, followed by Australia (21), United States (18), Canada (6), Ireland (5), Netherlands (4), New Zealand (2), some other European countries (10) and Nigeria (1). In this paper, only 18.9% of the articles reviewed are from the United States.

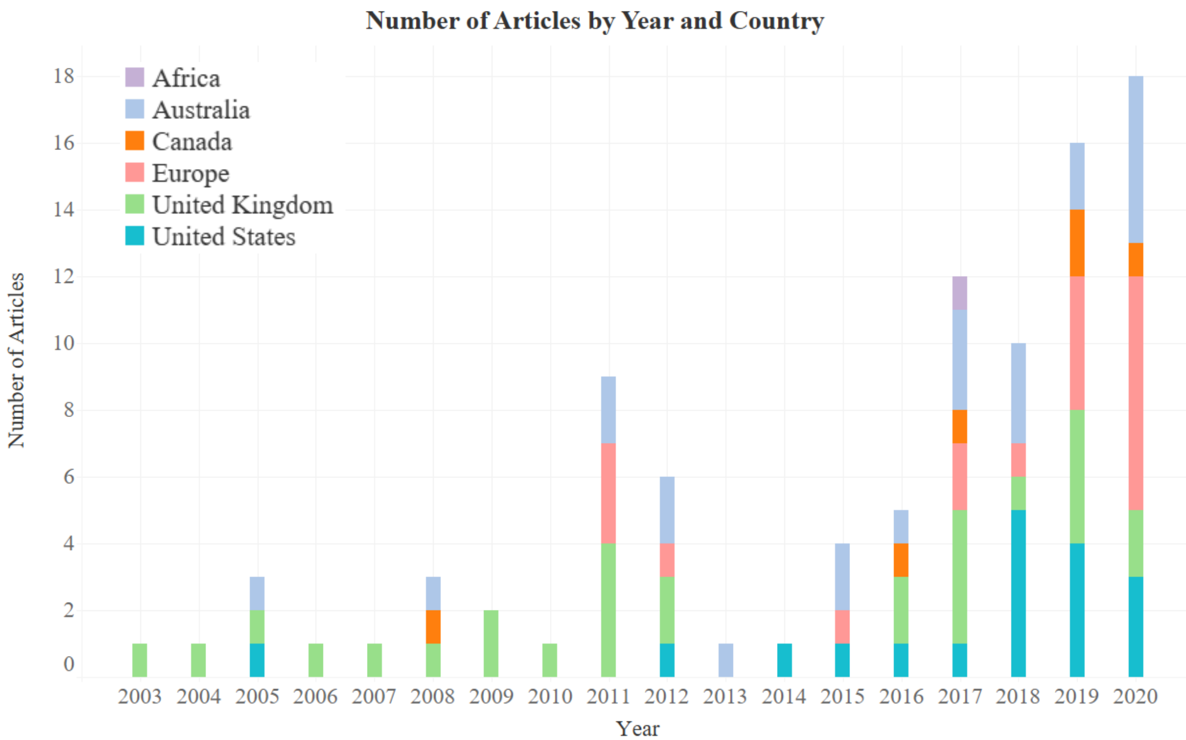


Figure 2.2: Number of synthesized articles by year and country (2013-2020).

2.3 Overview of Systematic Review

Although all 95 articles referred to patient experience, they adopted different approaches, data collection methods, analyses, administration techniques, technologies and focused on different illnesses and diseases. Figure 2.3 presents a concept map which starts with the six domains of health care quality proposed by the Institute of Medicine (IOM) in its publication "Crossing the Quality Chasm: A New Health System for the 21st Century" [2]. Then, it continues with a focus on the patient-centered care, which is divided into Patient-Reported Outcomes (PROs) and Patient-Reported Experiences (PREs). The expanding body of research into patient-reported experience / outcome measures (PREMs/PROMs) has generated questionnaires and insights for general practice [6] . These PROs and PREs are attempts to include patient perspectives in designing systems that truly meet their needs and have their corresponding branches to their measures (PROMs and PREMs). To get these measurements, a list of data collection methods is included along with the most common administration techniques and health implementations used. Data collected through these methods is evaluated to three approaches: qualitative, quantitative, and mixed methods (a combination of qualitative and quantitative). Currently, the measurement of patient experience leveraging commercial survey tools such as surveys from Press Ganey and other vendors has become a standard practice across most large health care organizations [4] .Other approaches are considered when performing different analyses and following certain methodologies such as: patient journey mapping (PJM), thematic analysis, content analysis, interpretative phenomenological analysis (IPA), Glaserian method of grounded theory and statistical analyses.

This review also encompassed the key elements for achieving a good patient experience, which are lived throughout the patient journey, thus the connection between these key elements and the patient journey mapping techniques (PJM). PJM evaluates the patient experience at different “hotspots” during the caregiving experience. These hotspots represent various points within the patient journey, such as registration, during transport, and admission to a hospital service, and include contact with staff, including physicians, nurses, dieticians, pharmacists, social workers, and environmental services. Having a visualization of the patient journey will allow patients to provide their comments, suggestions, and feedback throughout the different stages of their experience as patients [8]. PJM can also be divided into time-limited and open-ended depending on the illness or disease that is being treated. Additionally, PJM embraces different types of journeys according to different patient needs. Each patient journey is unique, and this methodology serves as a multidimensional framework that considers all aspects pertaining to the patients’ experience.

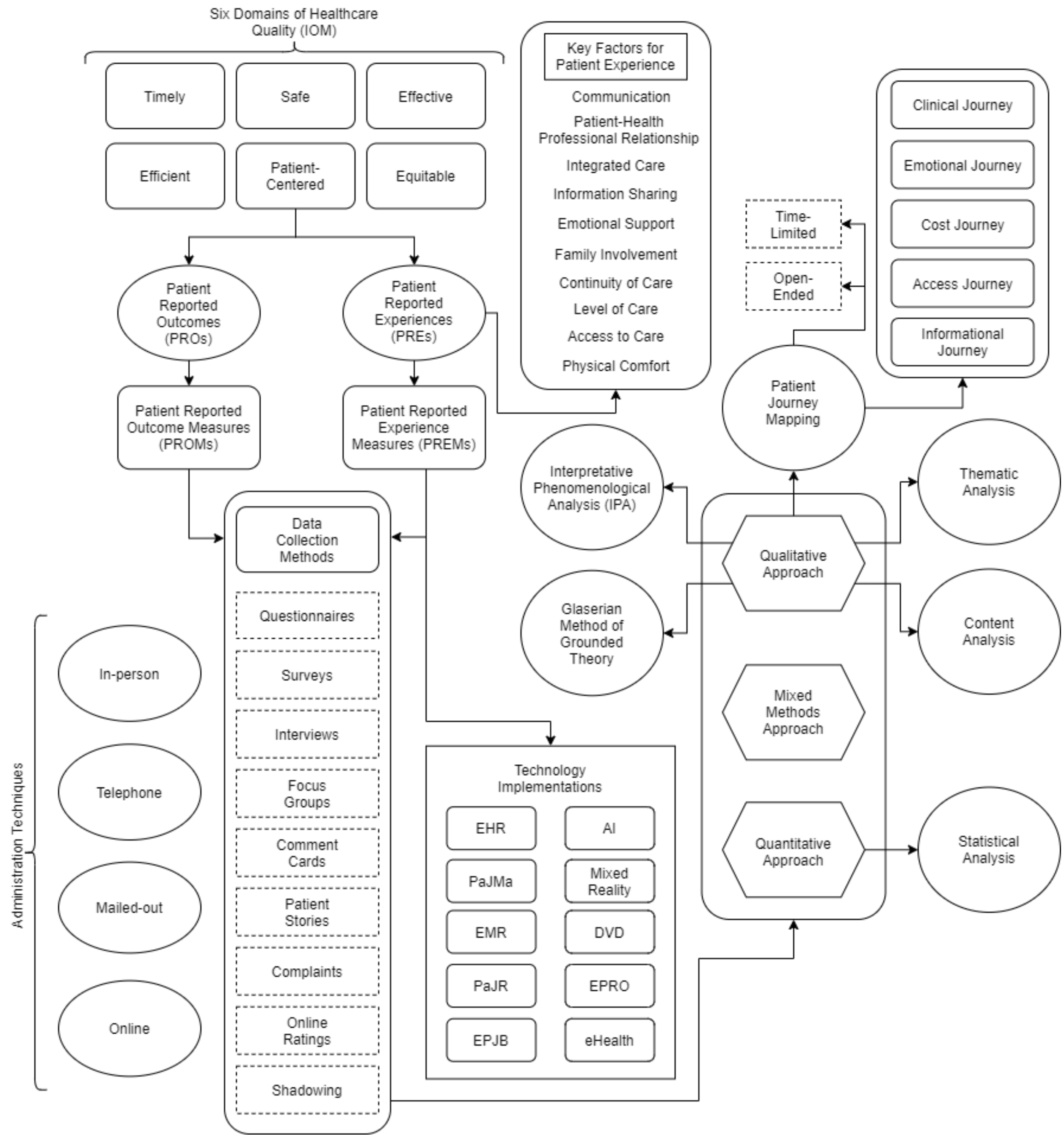


Figure 2.3: Concept map with identified research streams of the patient experience.

2.4 Research in Patient Experience

The quality assessment framework proposed by the IOM includes the six aims for health care systems: safe, effective, patient-centered, timely, efficient and equitable [2]. Doyle et al. adapted the six aims for healthcare system to constitute the three pillars for the health care quality:

effectiveness, safety and patient experience [9]. Patient Experience is an important health quality index. It has been demonstrated that patient experience and patient outcomes are correlated; better outcomes yield better experiences and vice versa [10]. Health care organizations generally emphasize on cognitive assessments and neglect patient emotions. However, it has been demonstrated that emotional aspects have an effect on patient satisfaction and clinical outcomes [11], [12]. Moreover, policymakers and regulatory bodies have increasingly recognized that the lack of patient-centered care results in unmet patient needs, higher costs and ultimately inefficient care [13].

Even though several studies have focused on the measurement of patient experience, only a few have reported the exact approaches used, limiting the information about instrument validity and reliability [14]. It is currently unclear whether the current tools for measuring patient experience really reflect what really is important to patients [15]. As some clinical effectiveness processes are relatively well established, other methods for improving patient experience or a patient-centered care are less ingrained and relatively new. The lack of a measurement standardization and patient experience instruments has resulted in the inability to nationally benchmark hospital performance, monitor effectiveness of interventions, establish hospital rankings and secure research funding [16], [17].

Patient satisfaction is a frequently used indicator for measuring patient experience, however there are some methodological deficiencies regarding the techniques used to address these experiences [12]. Satisfaction has erroneously been used interchangeably with patient perceptions and is defined as "fulfilling expectations, needs or desires" [18]. From this definition, Crow and colleagues [19] have identified two conclusions: (a) Satisfaction does not imply superior care, but only acceptable one; and (b) satisfaction is relative. Consequently, it is important to differentiate between the term patient satisfaction and patient experience. The latter uses questions that relate to actual hospital experiences, which aim to avoid value judgments and expectation effects [20]. As a multi-dimensional service, health care should be accounting for different attributes of patient experiences [19] such as the cognitive and emotional domains.

In addition, patient experience also has strong correlation with key financial indicators. For example, good patient experience is associated with lower medical malpractice risk [21]. Improving patient and family experiences also results in improving work processes and systems that enable clinicians and staff to provide better care which in turn results in higher employee satisfaction, reducing turnover [22]. Furthermore, patients keep or change providers based upon their experience with their providers, a study found that patients who reported poor relationships with their physicians were three times more likely to change providers than patients with high-quality relationships [23].

2.4.1 Key Factors for Patient Experience

Since the Institute of Medicine (IOM) presented patient-centered care as one of its six objectives for improving health care, many organizations have embraced patient-centeredness as a pillar to their strategy and mission [24]. Research by the Picker Institute and the Commonwealth Fund presented eight dimensions of patient-centered care [25]. Several authors have also proposed their own characteristics of patient-centered care (PCC), but they remain consistent with the Picker Institute's ones.

1. *Respect for the patients' values, preferences and expressed needs* [26]: involve patients in decision-making [27], recognizing them as individuals and treating them with dignity, respect, and sensitivity; plus, assessing if treatments provided are aligned with patients' goals and values [28].
2. *Coordination and integration of care* [26], [29]–[32]: facilitate the coordination of front-line patient care, clinical care, ancillary, and support services [33]. Patient-centered communication approaches and care coordination are associated with lower rates of patients seeking supplemental information from online sources among patients with lower levels of education [34].

3. *Information and education* [26], [29], [30], [35]–[38]: provide information about clinical status, progress and prognosis; the process of care and information to facilitate autonomy and health promotion. Research has found that patient information needs vary based on health contexts of chronic illness vs. acute medical situations [39].
4. *Physical comfort*: pain management, assistance with daily living needs, hospital environment. Some other factors that can also be considered as stressors under this category are hard work or mobility difficulties and the disappointment or the "it gets you down" feeling when patients look at their bodies after injuries [40], [41].
5. *Emotional support and alleviation of fear and anxiety* [38], [42]: caregivers need to pay special attention to anxiety regarding physical status, treatment [41], prognosis, illness and financial impact on patients and their families.
6. *Involvement of family and friends* [29], [30]: provide accommodations, involving support people in decision-making, support family members as caregivers, recognizing the needs of family and friends. One example is, when parents relinquish part of their protector role when their child is admitted to a hospital, as such, it is always encouraging that parents are working in partnership with professionals in the provision of clinical care and the development of services [43].
7. *Continuity and transition* [26], [30], [38]: provide patients with quality information about physical limitations, dietary needs, medications, and access to different types of support [43]; coordinate a plan with ongoing treatments and services after discharge [41].
8. *Access to care* [26], [33]: availability of transportation, easy access to the hospital, ease of scheduling appointments, clear instructions on how to get referrals, etc.

Literature has also highlighted the importance of information sharing and communication as key facilitators of patient experience [4], [24], [27], [29], [30], [33], [44], [45]. Information needs are completely subjective and individual-dependent. For an information need to arise in a person, there needs to be a stimulus (life major change, life-threatening condition, or illness) that is perceived as a challenge due to its unpredictability and uncertainty. In response to this event, individuals tend to adopt two major coping mechanisms: problem-focused and emotion-focused. The problem-focused coping includes this seek for information and direct action [46], [47]. A failure to meet this information need can result in detrimental effects as high-stress levels and difficulties in coping [42], [48]. With some conditions, the emotional and informational needs of family members might surpass the ones of patients themselves [49]. Other issues related to unavailable information, information given at a time when the patient is not ready to receive it, or receiving non-requested information also represent additional problems in this domain [36].

Another key element of patient experience is the patient-health professional relationship. Some physicians tend to reduce patients to their disease labels, with little focus or empathy on the person behind it [33], [50]. These circumstances have intensified negative encounters and the "walking a tightrope" sentiment for patients [36]. Research by Kreuzer et al. [51] illustrated that the most "luxurious" aspect of all their study interviewees was receiving "moments of care", a short-lived and prosocial interpersonal interaction experienced with medical staff characterized by an authentic presence, balanced power relationship and interpersonal synchrony. The way employees and staff interact with patients is critical not only for the experience but for outcomes as satisfaction and loyalty [32], [37], [41], [51]. Furthermore, Easton et al. have exhibited that poor communication contributed to the majority of medication errors [52]. The dynamic environment that characterizes patient care in hospitals requires extensive communication between staff and multidisciplinary teams (MDT) [53]. MDT meetings and communication are essential for discussing and addressing delays and bottlenecks in the care process [54]. Similarly, the collaboration between patients, carers and professionals by itself brings broad benefits and changes in the culture for individual services [55].

When measuring quality, the level of care associated with healthcare service delivery also needs to be evaluated. Caring is a highly valued human function that directly influences patient satisfaction. Caring is a component of patient-centeredness and involves interpersonal behaviors from all healthcare members, responsible for supporting meaningful communication with the patient/client and their family [56]. These behaviors can be referred to as personalization, participation and responsiveness when meeting a person's health and care needs [57]. Previously, the level of patient care was only associated with the nursing profession, which also has a high correlation with patient satisfaction. The greater the satisfaction with nursing care, the greater satisfaction overall. Patients want to be treated as individuals, rather than just any other patient [48]. This has been identified by the American Nurses Association as an outcome that should be measured and monitored [58].

Healthcare continuity is also essential for achieving high-quality patient care and experience [59]. Haggerty et al. [60] identified three types of continuity: (1) Informational continuity: documented information about the patient and his/her preferences, values and context; (2) Management continuity: a consistent and coherent approach to the management of the health condition that is responsive to patient changing needs; and, (3) Relational continuity: an ongoing relationship between patient and health provider. The operative and postoperative periods from surgeries have also proven the value of the outcome of the surgery [32] and continuity of communication [61]. Effective communication about medications after discharge is especially vital for people from non-English speaking backgrounds (NESB), who might also need an interpreter and translated educational materials [62]. Uncertainty is another component that plays a role throughout the continuum of care, especially for people with chronic illnesses, where it extends to broader life issues [63]. Assisting patients dealing with uncertainty associated with chronic health problems is essential for easing their patient journeys [64].

Patient accompaniment is another factor that increases the patient ratings for visit satisfaction, dimensions of interpersonal rapport, information giving and care quality [65]. Evidence shows that the more active family members are in physician visits, the more highly satisfied patients are with their care provider [66]. Effective engagement of patients also implies allowing the patient's family to participate in caregiving and express preferences, which creates a respectful and empathic environment where individuals feel valued and cared for [9], [44], [67].

Research by Blair et al. demonstrates how by incorporating specialist teams and a family-friendly environment can enhance users' experiences, satisfaction and journeys [68].

Many studies also address how patients rate their quality of care. Nonetheless, it is important to differentiate between the ones using concepts and measures derived from the perspectives of patients versus the ones from staff, administrators, or physicians. Medical anthropologists involved in clinical research have detected variations in patient and health professional perspectives, which might be attributed to differences in education and socio-economical background [69]. These variations might represent important implications for the management of diseases and conditions [70]. Understanding these variations is crucial to achieving a better appreciation of patients' overviews and "reframe" practices for more patient-centered care [71].

Additionally, the literature suggests that good experiences are associated with clinical effectiveness, adherence to recommended practice and medication, lower lengths of stay (LOS) and decreased mortality rates in various settings [9]. Shorter LOS also continue having positive impacts on health systems by increasing patient access, flows, and reducing costs [72].

Sample size also ensures a range of indications, ages, genders, and ethnicities, enabling the exploration of patterns and relationships within and between group participants. For instance, a small sample size is the norm in IPA studies, since the analysis of large datasets may result in the loss of subtle inflections of meaning [73]. Also, the patient journey approach to sampling is sometimes based on a stakeholder perspective, which ensures that professionals "listen" to the data from patients [74].

2.4.2 Types of Patient-Reported Measures

The patient-reported experience measures (PREMs) and patient-reported outcome measures (PROMs) are recommended for measuring patient experience within the context of health care quality worldwide [75]. The use of patient-reported outcomes (PROs) is an attempt to include patient perspectives in designing systems that truly meet their needs. In PROs, patients are the ones reporting on their health statuses and quality of life [76]. PROs are measured by patient-reported outcome measures (PROMs) [77], which are standardized and validated questionnaires that measure patient's perceptions of their general health in relation to a specific disease and are generally delivered pre and post operatively [78]. Traditionally, the choices of PROMs are often based on professional judgment versus strong conceptual models creating issues with grouping and scoring items into domains [77]. They are generally focused on specific conditions and sometimes fail to capture the global impact of health care on the patient's life. This represents a deficiency in the current use of measurement of outcome measures, by ignoring patients' experiences before and during their treatment, which are often associated with the most pain and suffering stages [79].

On the other hand, patient-reported experiences (PREs) describe how patients experience health care throughout the various stages of the care process and are also self-reported interpretations from patients and support people [80]. PREs are measured by patient-reported experience measures (PREMs) normally in the form of questionnaires as well. PREMs differentiate from satisfaction surveys by reporting objective patient experiences. They do not measure the outcomes of care, but the impact of care on the overall patient's experience. However, acquiring patient data can be challenging. For instance, in post-procedures, many patients might still be sedated or unable to provide coherent and reliable feedback [81].

The use of PREMs and PROMs can lead to the recognition and early detection of symptoms and time-sensitive diseases such as idiopathic pulmonary fibrosis [75]. In the case of PREMs, the NHS (National Health Service) National Quality Board produced a working definition of patient experience with eight indicators to guide measurements. These indicators are: respect for patient-centered values, preferences, and expressed needs; coordination and integration of care; information, communication, and education; physical comfort; emotional support; welcoming and involvement of family and friends; transition and continuity; and access to care [82].

2.4.3 Approaches for Measuring Patient Experience

This section summarizes approaches used in empirical studies for measuring patient experience.

Qualitative Methods offer an in-depth understanding of patient experiences. This approach allows researchers and evaluators to gain a deeper understanding from open-ended questions without the quantitative limitations. These questions allow patients to describe their experiences in their own words. In doing so, qualitative methods gain better insights of patient perceptions, behaviors, and the meaning they assign to certain experiences. As such, methods as semi-structured interviews, and focus groups represent powerful tools in capturing patient experience [83]. Management observation is an example of qualitative technique, it gives the manager the opportunity to obtain patient feedback and identify problems. Although, this technique requires specialized training and the possibility of influencing providers with the management presence, lack of statistical validity and reliability makes it prone to misinterpretation [84]. To ensure transparency and credibility, several articles referred to the Consolidated Criteria For Reporting Qualitative Research (COREQ) in the reporting aspects of their studies [33], [85], [86]. A qualitative data analysis tool that has also been used by many authors is NVIVO 10, 11 or 12, which is used to facilitate data management and analysis [35], [37], [87]–[91]. Another qualitative statistical software used was ATLASti [45].

Quantitative Methods such as structured questionnaires measuring PROs are among the most common forms of quantitative measures of patient experiences. These questionnaires are designed to produce numerical data which can be further analyzed and provide useful insights, patterns, associations, and trends to health care systems. These questionnaires or standardized surveys ask identical questions to respondents. This approach can be utilized with relatively large samples and provide ability for comparison [8]. Nonetheless, the lack of patient expression in his/her own words decreases the depth in which experiences should be analyzed [83]. This has led survey experts to shift their attention from ratings of satisfaction to reports of experiences.

Mixed Methods include both qualitative and quantitative methods to gain broader perspectives [83]. The strength of the mixed methods is lies in cross-validating qualitative and quantitative data to understand where the findings converge [92].

Even though these analyses are wide common for studies on patient experience, there is a general acceptance that the patient experience incorporates patients' journey as a whole and it is a clinically important concept to measure [83]. Also, by considering and understanding individual patients, hospitals can improve the quality of service provided [93]. By gaining insight into patients' experiences, an overview of how certain illnesses are managed and how they are affecting patients can be established. This provides background to develop pathways that can guide best practices [94].

2.4.4 Methods for Measuring Patient Experience

Some of the most common methods used to evaluate patient experience after an approach was selected and data acquisition performed were:

Patient Journey Mapping (PJM): This method combines several methods to best understand the patient's experience by dividing the management of a specific condition or process into a series of consecutive stages [95]. Central to the process of developing pathways of care to deliver services following patients across and within organizations for particular conditions [96], "journey maps" are used to reflect the health care service from the perspective of different personas. In the case of a patient, a specific condition or treatment is broken down into steps (activities, interactions, locations, etc.) and the sequence of these steps can be viewed as the patient pathway, patient journey, or process of care [95]. These patient journeys may be time-limited (an episodic or acute illness where the patient's journey begins with the onset and ends with its resolution) or open-ended (e.g., chronic illnesses) [96]. This technique considers each touchpoint in terms of the relative contribution towards the patient's outcome taken from his/her own perspective. The final result from a patient journey mapping (PJM) is a visual representation of the patient journey: a map showing a timeline indicating all relevant interactions between the patient and the health provider in combination with several different aspects of the experience [93], [95]. In contrast with care pathways, where all problems are solved from the professional's perspective and in adherence to the principles of clinical governance, the PJM approach brings patients' views into consideration [96].

Mapping a patient's journey is a highly complex and time-consuming process [93]. PJM focuses on patients' progress along the health care system and aims to improve patient safety and overall health care quality by high pointing patients' flows in the care process [97]. A very general patient journey can be defined in stages as pre-diagnosis, diagnosis, treatment, adherence, and management [98], but these steps are broad and vary for every disease and individual. The process of mapping patient journeys also allows interactive troubleshooting and encourages the use of better databases, which can be augmented to improve diagnosis and provide a framework for personalized care. Additionally, there is no standard approach on how to perform the steps in the mapping process, as a result, the adoption of this methodology is hindered [99]. From a lean perspective, everything that enhances patient experience is adding value and anything else is a waste [99], [100].

Even though internet searches for "patient journeys" are dominated by cancer journeys, PJM can be used for any procedure or disease [101]. PJM has been identified by the literature to increase communication and establishment of partnerships and also recommended to create care pathways [102]. However, it is also a time-consuming activity and readily dependent on patients' recall and ability to distinguish between health professionals and different stages [90]. This technique can also be modified and refined to add more dimensions as feelings, treatments [103], information needs, symptoms etc. [36]. By examining the patient journey through multiple lenses, a granular and insights-rich view of the patient journey can be acquired. This can be done by understanding the different journeys the patient goes through, such as the clinical, cost, attitudinal/emotional and informational journeys [104]. For instance, a study by McDonald et al. [105] about people with unresectable or metastatic gastrointestinal stromal tumors (GIST) patient journey added and emphasized the different emotions felt by patients starting from the diagnosis and identified stages of crisis, hope, adaptation, new normal and uncertainty. The uncertainty stage in this and many PJM cases is always latent because disease recurrence remains a profound threat, which often causes the patient to go back to crisis states again. Crisis and no crisis stages, each one offers opportunities for physicians and staff to provide support throughout the patients' journey.

The remaining part of this section summarizes other methods used in evaluating certain aspects of patient experience.

Content Analysis: Is a method where text is read by all authors and meaning units are abstracted and labeled by codes; these codes are then compared and grouped into categories and subcategories until consensus is reached. Finally, the latent content of the categories is abstracted into a theme [36], [106].

Glaserian Method of Grounded Theory: This method involves data collection, analysis and comparison; analyses involve the construction of codes and categories to finalize in the development of theory [107], [108]. Memos are considered essential in this method as they provide a repository of ideas for theory development [109].

Interpretative Phenomenological Analysis (IPA): Is a qualitative method that involves an idiographic focus concerned with providing insights into how an individual in a specific context makes sense of a specific phenomenon [110]. From the review, research by Flattery et al. [64] followed the most common IPA method or Colaizzi's seven-step process for analysis of phenomenological data, which involves (1) reading all transcripts, (2) extracting significant statements, (3) creating meanings by coding repetition, (4) aggregating issues into themes, (5) writing an exhaustive description, (6) identifying similar concepts and (7) asking participants for validation [111].

Statistical Analysis: Descriptive statistics (mean, standard deviation) and categoric (frequency) data analyses were commonly used to characterize the samples. Some articles have used statistical software packages such as the statistical package for social sciences (SPSS) [48], [90].

Thematic Analysis: is a method for analyzing qualitative data that involves searching across datasets to identify, analyze and report patterns. The most widely accepted framework to perform it consists in six steps: familiarizing with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes and producing the report [112], [113].

Integrated Patient Journey Mapping (IPJM)*:* it is based on the concept of PJM, but also accounts for the factors of performance improvement, regulatory constraints, and patient experience. It offers a structured means of exposing service reforms and facilitating collaboration between multidisciplinary teams and different stakeholders [114].

Subjective Well-Being (SWB): this model uses a SWB assessment during the pre-treatment, treatment and post-treatment stages [79].

Clinical Process Redesign**:** it is a method that focuses on the patient journey as the primary involvement locus. It uses process mapping to identify value-adding steps and involves redesign teams identifying non-value adding steps to improve patient journey flows. This process involves clinicians, managers, patients, carers and multidisciplinary teams [115].

Business Process Re-Engineering (BPRE)*:* is a business management approach focused on the analysis and redesign of workflows and processes with the objective of improving "customer experience", improving efficiency and thereby lowering costs [116], [117].

Experience Based on Design (EBD)*:* is another patient-focused design process which allows designers to create experiences rather than services [118].

Service Design*:* is a novel discipline focused on ideating, defining, and implementing services using a consumer-centric approach. The holistic and multidisciplinary approach of this technique enables teams to develop services that account for the functional, emotional, tangible and intangible aspects of services [119]. Service design approach examines conscious and unconscious needs of customers with a range of methods and tools, which provides a deep understanding of user experience over time and during service touchpoints [120].

Metabolic Phenotyping: Along the same lines of additions to the PJM technique, metabolic phenotyping of the patient journey has emerged as a novel idea intended to enhance the patient journey, building on clinical diagnostic criteria to improve the sensitivity of diagnoses [101].

Patient Pathway Models: The detailed modeling of "patient pathways" for specific patient groups is another approach compared to PJM. However, patient pathways focus only on addressing short-term treatment episodes. They are suitable for modeling curable conditions in large homogeneous patient groups [121]. In the case of chronic conditions or multiple chronic conditions (which account for over two-thirds of health care costs [122]) most models lack the level of detail in process descriptions and resource requirements needed to determine service costs and effectiveness.

2.4.4 Data Collection Methods

Collecting information from patients can better inform decision-making about service improvements and play a role in some accreditation programs [123]. Another sign of the attention on patient-centered care is the increasing number of surveys about patient perceptions that are being developed and collected through different qualitative studies (e.g., Picker surveys). Similarly, since 1994 the Agency for Healthcare Research and Quality (AHRQ) has developed the CAHPS surveys with the explicit goal of capturing information from consumers and publicly report it [12]. CAHPS is a popular and highly used survey that has been applied to different domains: hospitals (HCAHPS), home health (HHC CAHPS), clinician and group (CG CAHPS) and many more [83]. Categories for measurement in HCAHPS include care from nurses and doctors, the responsiveness of hospital staff, hospital environment, communication about medicines, experiences at the hospital, discharge information, overall hospital ratings, and patient information and demographics [124]. Surveys from Press Ganey and other vendors sometimes include additional items (e.g. admission room, meals, treatments, visitors and family, physicians, etc.) for a more comprehensive understanding of the patient experience [83]. When developing survey instruments to evaluate patient experience, an acceptable theoretical basis, patient input and piloting of surveys need to be done to assess whether those instruments are measuring what they are intended to [125].

Several possible sources of bias might exist on data collection methods depending on for example, on the wording of the questionnaire, the questions' scope for ambiguity, the respondents' mental exertion required to answer, question sequencing, and the respondents themselves [19]. Some studies gathered the data from patients, but some others did incorporate the viewpoints of professionals and staff.

The main data collection methods found in this systematic review were:

- *Questionnaires:* they typically have a stem and a scale that describes the aspect of care respondents are being asked to evaluate and rank [19]. Open-ended questions allow patients to freely express their responses and experiences in their own words. Closed-ended questions include multiple-choice responses, Likert scales, visual scales, or other questions where there is a fixed set of answers to choose from. Research suggests that response to closed-ended questions is usually positive, while for open-ended questions it tends to be more negative [14].
- *Surveys:* generally, in the form of self-completion ones, they are administered after different lapses of time from service delivery encounters. Cross-cultural factors (translations) and timing might affect results [19]. To gain numerical data for analysis and comparison, they sometimes use scales (e.g., Likert scale) [14]. There are surveys for specific services, conditions, specified in carers, broad communication, hospital care, primary care and more [14].
- *In-depth interviews:* generally administered in person or by telephone [14]. They are applied to small samples and are fairly effective at assessing quality, yet they require significant time, skills and resources [126].
- *Semi-structured interviews (SSIs)*: this type of interview includes some structured and some open-ended questions [19]. SSIs approximately last for 1 hour and are

audiotaped [87], [127]. Most interviews were transcribed verbatim prior to analysis [29], [36], [37], [40], [64], [89], [103], [128]. Some studies offered interviewees the opportunity to read their interview transcripts and add more details [87].

- *Focus Groups/ Group Interviews/ Panels:* they assess the complexity of patient satisfaction; enable interaction and engagement between participants; the mutual stimulation of views can generate solutions, uncover unanticipated issues, and raise more issues than an individual interview [14], [19].
- *Comment Cards:* they generally focus on negative aspects. They are cheap and impersonal [19].
- *Patient Stories/Narratives:* they are known for offering a way to make sense of strong emotions [129]. Written or videoed patient stories have been used to encourage discussion in team/board meetings. Patient diaries capture the record of life as lived, including symptoms, medication, experiences of stages of recovery, rehabilitation adherence, health care appointments, attitudes and emotions throughout their journey [87]. The stories and narratives of patients range from near-misses, lack of professionalism, information mishaps to sentinel events [130]. In the same lines of storytelling and narratives, the method of "yarning" has also been employed by researchers to collect information and establish a relationship with Indigenous participants, especially from the Nyoongah culture in Australia [131].
- *Complaints/Claims*: they contain the date, diagnostic, procedure, provider information, and other data, which creates an overview of services provided and reveals a data-driven understanding of how patients traverse the health care system [132]. Evidence shows that most complaints are made to protect other consumers and start an investigation [133]. Administrative claim data from insurance providers in the U.S. offers a uniquely detailed retrospective account of patient perspectives [134].

- *Online Ratings*: online ratings can provide valuable insights of the quality of health care services. Examining these websites together with other information systems may provide important indicators of the overall quality of care [14].
- *Shadowing:* studies have reported that it can assess lived experiences from a patient-centric perspective [135]. It may also have a valuable role in gaining insights into complex cross-hospital processes, particularly when dealing with vulnerable people who could be excluded from interview studies [29], [136].

Other techniques used were expert panels and public meetings, which are less generalizable and descriptive [14]. Ethnography is another method used for the same purposes and borrowed from the social sciences to reveal previously undetected properties of the patient experience [137]. It consists of detailed observation and dynamic interviews, then, ethnography documents the culture, perspective and practices of individuals and aids in understanding the context in which patients operate [138]. In addition to traditional methods, some novel ones such as photovoice and guided tours might also help in a better immersion and understanding of patient experience [83]. When selecting measurement techniques, it is important to first define what needs to be measured and recognize the different aspects of the experience, expectations, and satisfaction [139].

2.4.4.1 Administration Techniques

- *In-person or personal interviews/surveys* might not reflect accurate information due to loss of anonymity, timing, or fear on the impact of care. Responses tend to be positive and important information might be omitted. Additionally, a lack of structure/format in the questions and interviewer skills can result in inconsistent findings [16], [84].
- *Telephone surveys* have their special challenges such as timing, the length of interview, the establishment of a personal bond, sample representativeness, number of contacts that need to be made for each encounter, among others. There are also higher costs associated with

this technique due to the number of persons needed, survey inter-rater training and number of calls needed to make contact [8], [84].

- *Mail-out or self-administered surveys* need to be perfectly designed so they can reach the intended audience with no need for human interaction [8]. When valid and reliable surveys are sent out to motivated patients, valuable feedback and information is obtained [84]. The survey must be visually appealing and a postage-paid return envelope should be included as well [16].
- *Online surveys:* are increasing in popularity. They are different from web surveys, which ask similar questions to paper surveys; online rating sites focus on gaining numerical feedback, which is then collated to share a score with specific services and health professionals [140]. Other online tools seek to gain qualitative information from descriptive feedback [141].

Timing effects are another consideration when capturing patient experiences. Evaluation approaches are generally applied at the time of care or sometime after. For instance, the HCAHPS questionnaire is collected after 42 days the patient is discharged. Some other surveys often require patients to recall interactions from the previous year(s), which might introduce bias and inaccuracy [142]. Research on the effects of survey timing remains contradictory [19]. Bjertnaes found that patients report worse experiences for half of the patient-reported experience scales when time passed. Individual responses were also negatively impacted by the time increase [143]. One way of gaining "real-time" feedback is to use kiosks or electronic devices at the point of care [14].

2.4.5 Technological Implementations

The use of health information technology (HIT) as a means for improving health care services, decreasing waiting times, and improving the standards of patient care [144], [145] has also been proposed in the literature. Furthermore, connected health, defined as "where devices, services, or interventions are designed around the patient's needs, and health-related data is shared, in such a way that the patient can receive care in the most proactive and efficient manner possible" [146] has emerged as a promising area of research for addressing today's health system challenges. However, this interconnectedness also requires an understanding of workflows among different stakeholders to design health solutions to be fit within the healthcare ecosystem and sustainable in the long term[147], [148]. The implementation of visual models such as patient journey mapping (PJM), unified modeling language (UML) and others, enables stakeholders to comprehend workflow-related issues, make audits and propose improvement processes [95], [149].

New information systems and technologies have arisen to positively influence clinical practice and organizational culture [150]. Although HIT implementations have been demonstrated to improve patient experiences, "adapting new information systems to health care has proven difficult and rates of use have been limited" [151]. Most of these applications have been centered on administrative and financial transactions in health care settings and cost remains the most important barrier for its adoption in smaller settings [152].

Electronic Health Records (EHR): collect patients' electronically-stored health information in a digital format, they provide out-of-the-box solutions for the summary of information after patient encounters with health care providers. Epic (Epic Systems Corp) is one of the most widely used EHRs and deploys the after-visit summary for use in ambulatory care settings [4]. Patients are increasingly accessing their health data as a movement towards patient-centered care and patient engagement [153].

Patient Journey Modeling Architecture (PaJMa): is designed specifically for healthcare [154]. It enables a visual representation of the processes interactions, technologies and people and includes staff roles, processes, information flow, HIT, information technologies (IT), patient needs, and metrics [155], [156]. The PaJMa approach involves modeling sessions with system stakeholders. The modeling style is explained and then participants use post-it notes and butcher paper to describe the current system (there is a special focus on those processes where the patient is involved). Following the sessions, the butcher paper models are transferred into the PaJMa software and opportunities for improvement are highlighted [154].

Patient Journey Record System (PaJR): is an application of a complex adaptive chronic care model in which early detection of adverse changes elicits tailored care [157]. The principle behind this method is that trained persons telephone patients at high risk of hospitalization, they engage in semi-structured conversation about health concerns and well-being [158]. The PaJR uses machine learning to evaluate the answers to the questions and free text narratives provided by patients about health status and alerts if an intervention might be required [159], [160]. A clinical case manager reviews the alerts and if appropriate, transfers the information to health providers.

Electronic Patient Journey Board (EPJB): patient journey boards are tools for coordinating patient care and flow. These boards display relevant patient information at central locations [161]. An electronic patient journey board is similar, but offers advantages associated with digitalization. EPJB are used as a tool to facilitate team communication across wards and improve the efficiency of key hospital processes. These electronic status boards often contain the same matrix form as the whiteboards used in health care settings, they are typically positioned on large wall-mounted visual display screens and can be manipulated via keyboard and mouse [53], [162].

Electronic Medical Record (EMR): it captures services provided by non-physician providers, patient sociodemographic characteristics, and detailed diagnosis information that can enable a more complete understanding of the complexities of primary care. When linked with other databases, it can become a powerful tool for following the patient's care journey [163]. Nonetheless, in countries like Canada, EMR data is not always structured and is often provider-centric, making it difficult to share among different healthcare settings [164].

Artificial Intelligence (AI): studies systems that demonstrate behaviors associated with human intelligence. It has developed concepts, methods, and techniques relevant to natural language processing (NLP) that are increasingly being recognized as predictive tools in medicine and health. The essential elements of machine learning (ML) are an automated approach to learning patterns from empirical data using training examples [157]. One of the benefits of NLP, in particular, is its ability to take unstructured data and provide/extract useful insights rapidly [165].

Mixed Reality: it can compile patient-specific data in one anatomic model that can be used to educate the patient. A mixed-reality intervention at discharge can provide an overview of the importance of medication adherence [166].

DVD Records*:* it can represent an important clinical treatment tool by becoming the medium through which parents, family, and patients themselves can learn and grow in knowledge positively. It can be used as a humanistic approach bringing the clinical facts and experiences with the humanity of all the people involved in a common pathway [44].

Electronic Patient-Reported Outcomes (EPRO): Interest is expanding in the current availability of sophisticated, user-friendly electronic platforms for patient self-reporting, such as tablet computers, automated telephone calls and web kiosks. These platforms can provide actionable links to clinical care such as summary reports in a patient's EMR and real-time email alerts to providers. However, negative feedback has been received regarding the electronic self-reporting desks due to the difficulty of some patients when using them [31].

Electronic Health (eHealth): the term eHealth was introduced by Eysenbach [167] as "an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the internet and related technologies".

- Hybrid Services (e.g., online hearing screening, online motivational engagement, face-to-face diagnostic hearing evaluation, hearing aid trial and fitting and audiological rehabilitation, counseling and ongoing coaching for a hearing health care hybrid service delivery model [168])
- Feedback: healthcare professionals have proposed eHealth solutions that can provide them with anonymized patient feedback with suggestions for improvements, which could be addressed at the appropriate touchpoint of the patient journey [86].

- Referrals: eHealth solutions to monitor missing information from referrals and detect missing mandatory fields to be completed prior to a digital referral from primary to secondary care [86].
- Counseling: health professionals have identified the need of an eHealth solution that can support patient counseling by providing general information, checklists and reminders about what is required [86].
- Eligibility: healthcare professionals suggest eHealth solutions that could provide suggestions based on health metrics and provide self-care support to decrease risks before, during or after surgery and improve eligibility [86].
- Patient Flow: eHealth solutions can collect information and PROMs before and after surgical procedures [86].
- Healthcare guarantee: eHealth solutions able to collect information as referral status, and provide average processing and waiting times, which should be provided to the patient to help them prepare and plan [86].
- Post-discharge care: an eHealth tool capable of encouraging patients to be involved in their rehabilitation by providing personalized targets, reminders for activities, and the ability to monitor patient compliance with given instructions [86].
- Communication: eHealth solution capable of unifying and developing documentation in general, plus facilitating the information transfer between teams and organizations. The eHealth solution could also include the option to send messages or alternatively some kind of bot that can be used as a communication support tool [86].

Technological applications and devices design, development, deployment, and management are key to create an ecosystem that puts the patient at the center. Some examples of technological devices that are being used are registration kiosks, directional signage, patient room devices, electronic sign-in, notification systems, waiting room technology and monitors, etc. Evidence shows an overall increase in patient-reported satisfaction of wait times and courtesy after the use of electronics [169].

2.5 Research in Text Analytics

Text analytics is the automated process of drawing meaning out of large volumes of unstructured text into quantitative data. In a customer experience context, text analytics examines written text and finds insights, trends, patterns and topics of interest. Text mining, text analysis and text analytics are terms often used interchangeably [170]. The structured data created by text mining can be integrated into databases, data warehouses or business intelligence dashboards and used for descriptive, prescriptive, or predictive analytics.

Text analytics employs a variety of methodologies to process text, one of the most important ones being Natural Language Processing (NLP). NLP refers to the branch of computer science concerned with giving computers the ability to understand text and spoken words in the same way a person can. NLP combines computational linguistics (rule-based modeling of human language) with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process and understand human language in its full meaning. Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting, some of them are speech recognition, speech tagging (use of a particular word according to context), word sense disambiguation, named entity recognition, co-reference resolution, sentiment analysis, natural language generation and more [171].

2.5.1 Sentiment Analysis

Sentiment analysis (SA) or Opinion Mining (OM) is a computational study of opinions, sentiments, emotions, and attitudes expressed in texts towards an entity [172]. The target of SA as shown in figure 2.5.1 is to find opinions, identify expressed sentiments and classify their polarity.

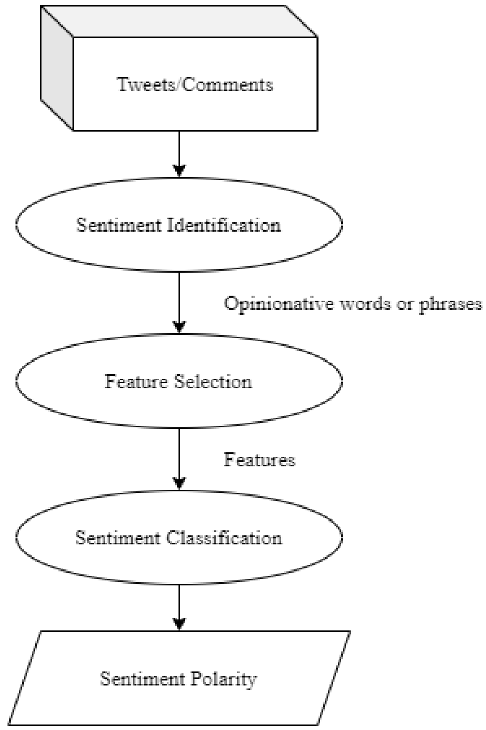


Figure 2.5.1: Sentiment Analysis Process on Tweets/Comments.

Sentiment Analysis can be considered a classification process. There are three classification levels in SA: document-level, sentence-level, and aspect level. Document-level consist of classifying an opinion document as expressing a positive or negative sentiment, it considers the whole document a unit (talking about only one topic). Sentence-level SA classifies sentiments expressed in each sentence. The first step performed in this case is identifying whether the sentence is objective or subjective. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents [173]. Aspect-level SA aims to classify the sentiment for the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity [172]. Data sets used for sentimental analysis vary from product reviews to stock markets [174] and political debates [175].

According to Medhat et al., there are multiple sentiment classification techniques as shown in figure 2.5.2. This paper will investigate a lexicon-based, corpus-based semantic approach for sentiment analysis. Additionally, this paper will examine the use of Latent Dirichlet Allocation (LDA) topic modeling to discover latent topics associated with comments from social media (Twitter and Reddit).

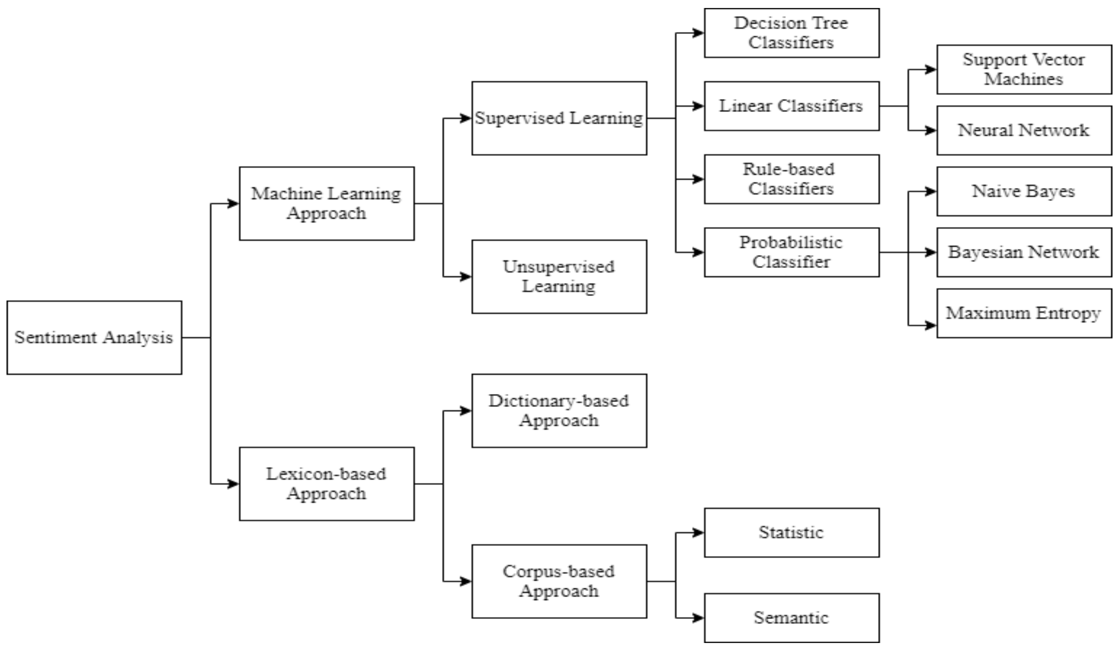


Figure 2.5.2: Sentiment Classification Techniques.

2.5.2 Topic Modeling

Topic modeling is a type of statistical modeling and unsupervised learning technique capable of scanning a set of documents detecting word and phrase patterns within them. It automatically analyzes text data to determine cluster words for a set of documents. Topic Modeling involves counting words and grouping similar word patterns to infer topics within unstructured data. By detecting patterns such as word frequency and distance between words, a topic model clusters feedback that is similar, and words and expressions that appear most often. Thus, it makes it easier to deduce what each set of texts are talking about [176].

2.5.2.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a generative probabilistic model, including a three-level structure with word, topic, and document [177]. It ignores syntactic information and assumes that all words in the document can be assigned a probability of belonging to a topic [176]. In LDA, a document is viewed as a distribution over topics, while a topic is a distribution over words. To generate a document, LDA firstly samples a document-specific multinomial distribution over topics from a Dirichlet distribution; then repeatedly samples the words in the document from the corresponding multinomial distribution. LDA assumes the topic proportions are randomly drawn from a Dirichlet distribution, which implies the independence between topics. [177].

A topic model can extract the latent topic structures by analyzing a large scale of statistical data. These structures are hierarchical and corpus-specific. In a good topic structure of LDA, every topic is understandable, meaningful, and exclusive to each other. However, finding the optimal K or number of topics is difficult and currently, there is no established method for it [177].

2.6 RESEARCH IN SOCIAL MEDIA

Social media's infiltration into the lives of internet users has been on the rise for the past years. The latest figures from Statista show that there are 3.78 billion social media users worldwide in 2021 [178]. Online media and networking sites are used to express and share public experiences in the form of product reviews, blogs, discussion forums, and more. Collectively, these media contain highly unstructured data combining text, videos, images, and animations that are useful in making the public aware of certain issues [179]. Social networking sites and blogs also offer a good source of information because users/people share their opinions freely about a wide variety of topics [172]. In the case of patient experiences, these facts about people freely expressing themselves on social media allow patients to describe their experiences in their own words. Storytelling and other trends to invite patients and family members to leave their reviews on online platforms like Twitter, and Facebook are also increasing in popularity [14], [180]. This facilitates qualitative methods to gain better insights into patient perceptions, behaviors, and the meaning they assign to experiences.

2.6.1 Twitter

Twitter is an American microblogging and social networking site where users post and interact with messages known as “tweets”. Tweets can be up to 280 characters long, including spaces, and can include URLs and hashtags [181]. As of the first quarter of 2021, Twitter had 199 million monetizable daily active users and exhibited a growth of 8% worldwide compared to previous years [182].

Twitter offers a developer portal or Twitter API where a set of self-serve tools are available to developers. In there, developers can create and manage their projects and applications, set up developer environments, and learn more about endpoints and features available [183]. Twitter API allows you to extract tweets and various tweet components as user name, timestamp, tweet text, hashtags, links, embedded media, replies, retweets, favorites, location, and more.

2.6.2 Reddit

Reddit is a network of communities based on people’s interests. Registered members can submit content to the website, which can be voted up or down by other members. This social platform attracts 430 million users each month, this represents a 30% increase in monthly active users compared to the last two years. This makes Reddit the third platform (after TikTok and Pinterest) with the highest increasing rate of monthly active users [182]. Reddit is broken up into more than a million communities known as “subreddits,” each of which covers a different topic. The name of a subreddit begins with “r/,” which is part of the URL that Reddit uses [184].

2.7 Study Contributions

After exploring the determinant elements of patient experience and identifying the lack of a standard method for measuring them according to what is important to patients, this study proposes the use of a NLP algorithm to evaluate patient experience following a breast cancer patient journey map by using social media data. The data imported from Twitter and Reddit will go beyond analysis of comments, taking a patient -centered approach that evaluates latent topics and sentiments from users. Findings from this study will practically inform evidence-based recommendations.

Chapter 3: Methodology

The methodology followed for this study consisted of extracting tweets and subreddits to use them as an input for the model. This model was created to evaluate patient experiences, latent topics and sentiments from Twitter and Reddit using Natural Language Processing (NLP) algorithms.

3.1 Research Questions

This research aims to answer the following questions:

- What are the current methods of capturing patient experience through PJM?
- How to efficiently capture patient experience using Machine Learning Algorithms/Natural Language Processing Algorithms?
- Can text analytics be used for measuring patient experience at various hotspots of a patient journey mapping for a breast cancer patient?

3.2 Study Design

From the literature review on patient experience, the need of involving family members and support people along with patients in their care was observed. By incorporating tweets and comments from Twitter and Reddit users, we are expanding the scope of this study to not only patients, but also support people, advocates, and people interested in this topic. Therefore, results obtained might not obtain the desired accuracy for mapping the breast cancer patient journey and experiences. However, results from this study can identify the latent topics and overall sentiments associated with breast cancer. Additionally, this study can provide useful insights for improving patients' and support people's experiences and give useful advice to them.

3.2.1 Cohort

Twitter and Reddit users posting/tweeting about breast cancer in the COVID-19 era.

Data extraction from both social media websites was performed using keyword search for the words "breast" and "cancer". Data scraping from Twitter was performed on June 15, 2021, June 28, 2021, and July 12, 2021, yielding a total of 3,151 tweets. The single data extraction performed for Reddit yielded 2,539 records from March 15, 2020, to March 15, 2021.

3.2.1.1 Relevance

Breast cancer has now overtaken lung cancer as the world's most commonly diagnosed cancer, according to statistics released by the International Agency for Research on Cancer (IARC) in December 2020 [185].

Breast cancer is the most common cancer in American women, except for skin cancers. About 1 in 8 U.S. women will develop invasive breast cancer over the course of their lifetime [186].

The COVID-19 pandemic has exacerbated the problems of late-stage diagnosis and lack of access to treatment, especially in low- and middle-income countries. In addition to having to cope with the disruption of services, people living with cancer are also at higher risk of severe COVID-19 illness and death [185].

For reference, a high-level breast cancer patient journey map is shown in Figure 3.2.1.1.

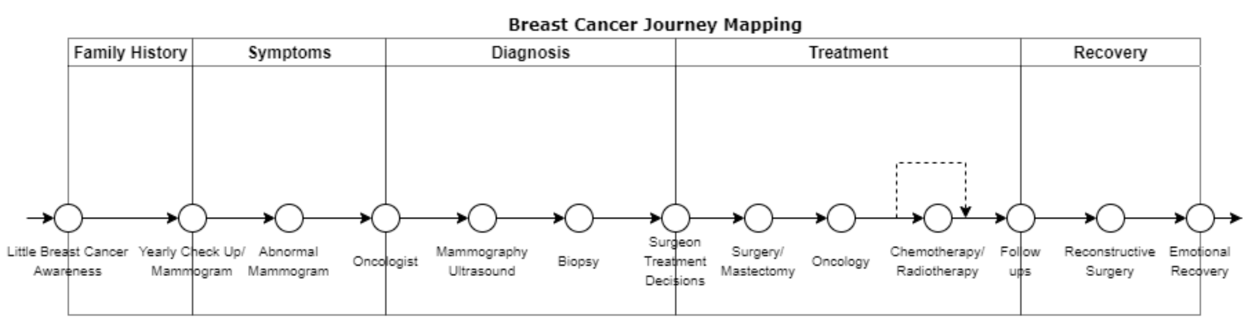


Figure 3.2.1.1: Breast cancer patient journey map.

3.2.2 Dataset Sources

Twitter: 3,151 records were obtained from three data extractions. The first web scraping was performed on June 15, 2021, and yielded 911 tweets. The second one was performed on June 28, 2021, and gave a total of 1,045 tweets. The third and final data extraction happened on July 12, 2021, and yielded a total of 1,195 tweets. In total, all three extractions gave a total of 3,151 tweets. Before being loaded to the NLP algorithm, these tweets went through various text-cleaning steps and the total number of records was reduced to 2,292 tweets containing information about the author, tweet text, time stamp, and geolocation.

Reddit: 2,539 subreddits were obtained from the data extraction. The web scraping performed for Reddit used a conversion of dates to Unix timestamp units, dates used for the extraction were from March 15, 2020, to March 15, 2021. Before being loaded to the final NLP code, the subreddits went through text-cleaning and pre-processing steps such as removing null values and removed comments. At the end of this process, the number of records was reduced to 2,028 subreddits containing author, body, and publish date.

The total amount of records obtained from initial extractions for both Twitter and Reddit was 5,690. Nonetheless, as mentioned before, to use the records for text analytics, a series of text-cleaning steps were performed. This resulted in a new total of 4,320 records. Table 3.2.2 illustrates the data extraction process and the number of records obtained.

Table 3.2.2: Data extraction process with the number of records.

Social Media Network	Number of records extracted		Total Number of Records Extracted	Number of records after text-cleaning steps
Twitter	First Extraction	911	3151	2292
	Second Extraction	1045		
	Third Extraction	1195		
Reddit	First Extraction	2539	2539	2028
		Total	**5690**	**4320**

3.2.3 Toolkit

This study was conducted using Python 3.7.4 and Jupyter notebooks, which is a browser-based interactive computing notebook environment.

Some Python libraries used to perform the analyses were:

- NumPy: one of the most used Python packages for scientific computing. It provides multidimensional arrays, as well as variations as masks and matrices that are used for various math operations [187].
- Pandas: an open-source Python package widely used for data analysis and machine learning. It is used for tasks as data cleansing, normalization, visualization, statistical analysis, data inspection, loading and saving data, and more [188].
- Matplotlib: is a comprehensive library for creating static, animated, and interactive visualizations in Python [189].
- Seaborn: is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics [190].
- Itertools: The module standardizes a core set of fast, memory-efficient tools that are useful by themselves or in combination. Together, they form an "iterator algebra" making it possible to construct specialized tools succinctly and efficiently in pure Python [191].
- Collections: This module implements specialized container datatypes providing alternatives to Python's general-purpose built-in containers, dict, list, set, and tuple [191].
- Re: This module provides regular expression matching operations [191].
- Warnings: Warning messages are typically issued in situations where it is useful to alert the user of some condition in a program [191]. For this code, the warnings library was used to ignore deprecation warnings.

- String: This library contains constants and classes for working with the text [191].
- NLTK: the most widely used framework for topic modeling and text analytics. It provides plenty of corpora and lexical resources to use for training models, plus different tools for processing text, including tokenization, stemming, tagging, parsing, and semantic reasoning [176].
- Sklearn/Scikit-learn: is a free machine learning algorithm in Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python's numerical and scientific libraries [192].
- Pickle: The pickle module implements binary protocols for serializing and de-serializing a Python object structure. *"Pickling"* is the process whereby a Python object hierarchy is converted into a byte stream, and *"unpickling"* is the inverse operation [191].
- Networkx: It is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks [193].
- Wordcloud: This Python's library allows the creation of word clouds, which is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.
- Gensim: a robust library that provides a suite of tools for implementing LDA and other topic modeling algorithms [176].
- PyLDAvis: It is designed to help users interpret the topics in a topic model that has been fit to a corpus of text data. The package extracts information from a fitted LDA topic model to inform an interactive web-based visualization [194].
- Pprint: The pprint module provides a capability to "pretty-print" arbitrary Python data structures in a form that can be used as input to the interpreter [191].
- Os: The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility module. This module provides a portable way of using operating system-dependent functionality [191].

- TextBlob: This is an open-source Python library for processing textual data. It performs different operations on textual data such as noun phrase extraction, sentiment analysis, classification, translation, and more [195].

3.3 Model Design

This model design uses Natural Language Processing (NLP) Algorithms to understand, interpret and manipulate human language or in this case, text extracted from Twitter and Reddit. The model design is illustrated as an IDEF3 diagram in Figure 3.3. The inputs used for the "Evaluating Patient Experience" model are the Twitter and Reddit datasets (with 2,292 and 2,038 records respectively). The mechanisms adopted for this study are the NLP algorithms using Python's Jupyter notebooks. The controls for this model are the NLP parameters applied throughout the code. Finally, the outputs given by this model are the results from the LDA and Sentiment Analyses. These outputs are the latent topics from the datasets and a data frame in the form of a CSV file including the polarity, subjectivity, and overall classification (positive, neutral, negative) of each comment/tweet.

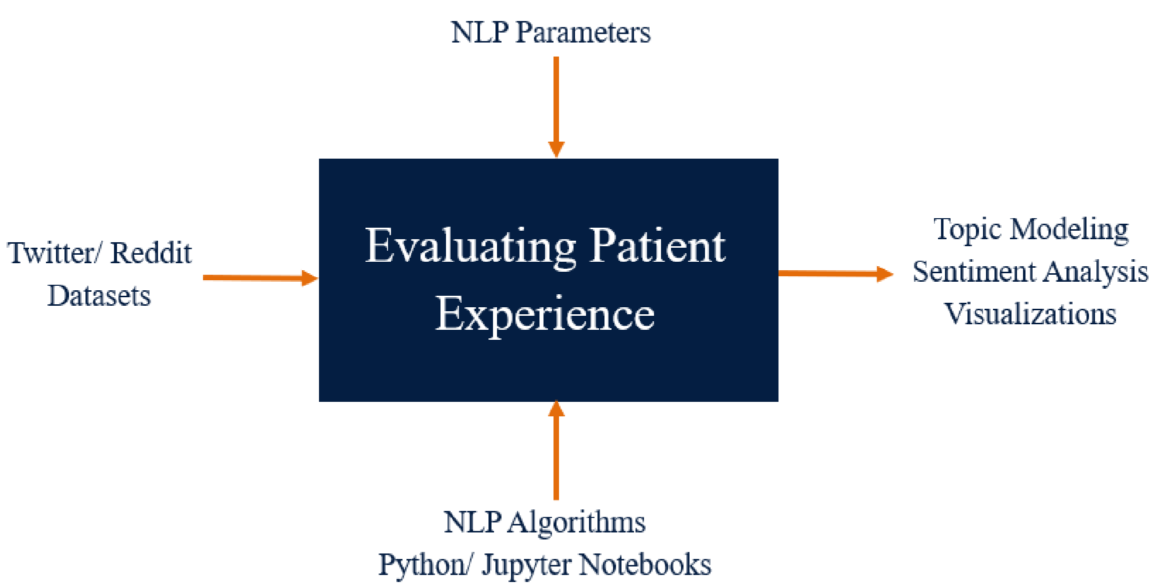


Figure 3.3: IDEF3 diagram of the model design.

3.4 PROPOSED MODEL

Figure 4.1.5 illustrates a flowchart with various steps of the proposed model starting from data extraction, cleaning, preprocessing, and topic modeling and sentimental analysis. Each of these steps is explained in detail in this section.

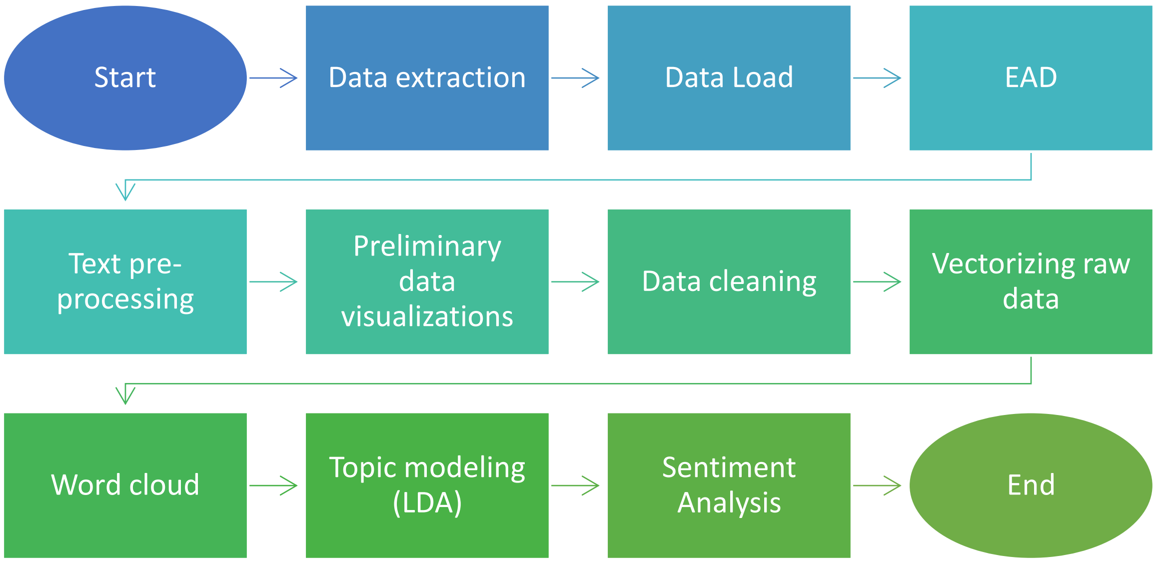


Figure 4.1.5: Flow chart of the proposed model.

I. Data Extraction (Twitter and Reddit)

This model for evaluating patient experience using Natural Language Processing algorithms started with the data extraction from Twitter and Reddit by keyword search for the words “breast” and “cancer”.

II. Data Load/Import

The extracted messages from Twitter and Reddit containing the keywords “breast” and “cancer” were imported to two separate Python’s Jupyter notebooks using the library “pandas”. Pandas was also used to drop missing or null values and removed content from both datasets.

III. Exploratory Data Analysis

Subsequently, an exploratory data analysis (EDA) was performed to analyze and investigate data sets and summarize their main characteristics. The first analysis performed was regarding text length.

IV. Text Pre-Processing

Some discrepancies were found in the number of characters in tweets and subreddits, and further investigation suggested that this was due to the use of mentions, and URLs. Therefore, the next step was the removal of mentions in the case of Twitter and URLs for both data sets with the use of Python's library "re".

V. Preliminary Data Visualizations

The last step of the EDA was to get visualizations about the 20 most common words used in both Twitter and Reddit.

VI. Data Cleaning

The first step in this stage consisted in removing punctuation using Python's library "string" and using functions and lambda expressions. Secondly, the tokenization step was also performed. Thirdly, stop words were removed by using NLTK's English stop words. All tweets containing a mixture of English and other languages were removed in this step. The last step in this phase was the lemmatization, which means that words were converted to their meaningful base form (words in the third person are changed to first and verbs in past and future tenses are changed into present).

VII. Vectorizing Raw Data

One of the major steps in text analysis is to convert tokenized texts in to vectors or sequences of numbers that classification algorithms can use. For this study, three different vectorizing techniques were employed and used for creating their corresponding sparse matrices (a matrix in which most entries are zero).

I. TF-IDF Vectorizer: TF-IDF stands for *term frequency-inverse document frequency*, and the tf-idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. This vectorizing technique creates a document-term matrix where the columns represent single unique terms (unigrams), but the cell represents a weighting meant to represent how important a word is to a document. Typically, the tf-idf weight is composed of two terms: the first computes the normalized Term Frequency (TF), or the number of times a word appears in a document, divided by the total number of words in that document; the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears. The combined formula used to obtain term weights is illustrated in figure 4.1.4. Weights were calculated with the TFIDF Vectorizer tool from Python's 'sklearn".

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

Figure 4.1.4: Formula used by TFIDF vectorizer to calculate term weight.

II. Count Vectorizer: creates a document-term matrix where the entry of each cell will be a count of the number of times that word occurred in that document. This was done with the Count Vectorizer tool from Python's 'sklearn".

III. N-grams Vectorizer: Creates a document-term matrix where counts still occupy the cell but instead of the columns representing single terms, they represent all combinations of adjacent words of length n in your text. The library used to create this vector was also Python's 'sklearn" with its feature Count Vectorizer.

VIII. Word Cloud

Word clouds, text clouds or tag clouds are a collection or cluster of words depicted in different sizes. The bigger and bolder the word appears, the higher frequency within a given text and the more important it is. They help to pull out the most pertinent parts of textual data, from blog posts to databases, and can also help business users compare two different pieces of text [196].

IX. Topic Modeling (LDA)

Topic Modeling is an unsupervised type of statistical modeling for discovering the abstract "topics" that occur in a collection of documents. Latent Dirichlet Allocation (LDA) is an example of topic modeling used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions. The goal of LDA is to determine the mixture of topics that a document contains. However, as stated earlier, finding the optimal K or number of topics pertaining to a document is difficult, varies within different techniques, and currently there is no established method for it [177].

To work on the LDA, the main Python's libraries used were: Gensim and pyLDAvis. The first steps for the LDA were to input the cleaned text and build a dictionary, a corpus, and a term frequency from it. Then, the number of topics and words in each topic was set to 10 and the LDA model was built. The different stages of topic modeling are listed below.

a. Data Preparation for LDA
b. LDA Model Build
c. LDA Model Results --- Output

X. Sentiment Analysis

Sentiment Analysis is a sub-field of NLP that measures the inclination of people's opinions within the unstructured text (after all pre-processing steps have been performed). As mentioned earlier, Sentiment Analysis can be performed using two approaches: Lexicon-based or Machine Learning-based.

This paper will investigate a lexicon-based, corpus-based semantic approach for sentiment analysis. This is a practical approach to analyzing text without training the models. The result of this approach is a set of rules based on which the text is labeled as positive/negative/neutral. These rules are also known as lexicons. The most widely used lexicon-based approaches are TextBlob, VADER, and SentiWordNet. For this sentiment analysis, Python's TextBlob library will be utilized.

The two most common measures that are used to analyze sentiments are:

- Polarity: how positive or negative an opinion is. Polarity ranges from -1 to 1, where 1 is the range of the most positive opinion, 0 is used to demonstrate neutrality and -1 is used to denote the most negative opinion.
- Subjectivity: how subjective the opinion is. Subjectivity ranges from 0 to 1, where 0 denotes an objective opinion and 1 is a subjective one.

Chapter 4: Results

This chapter discusses various results obtained from the text analysts of tweets from twitter and sub-reddit messages from the Redditt social media platforms.

4.1 Model Setup

The data extraction from Twitter (3,151 records) and Reddit (2,539 records) for messages related to breast cancer. Cleaning the text messages for missing and null values resulted in a reduction of 859 records from Twitter and 511 from Reddit.

Tweets have a limit of 280 characters, and this was also demonstrated by the analysis, which showed a mean of 193 characters. For Reddit, the self-post character limit is 40,000 and the data set obtained showed a mean of 818 characters. Figure 4.1. summarizes the top 20 most common words in both twitter and reddit messages.

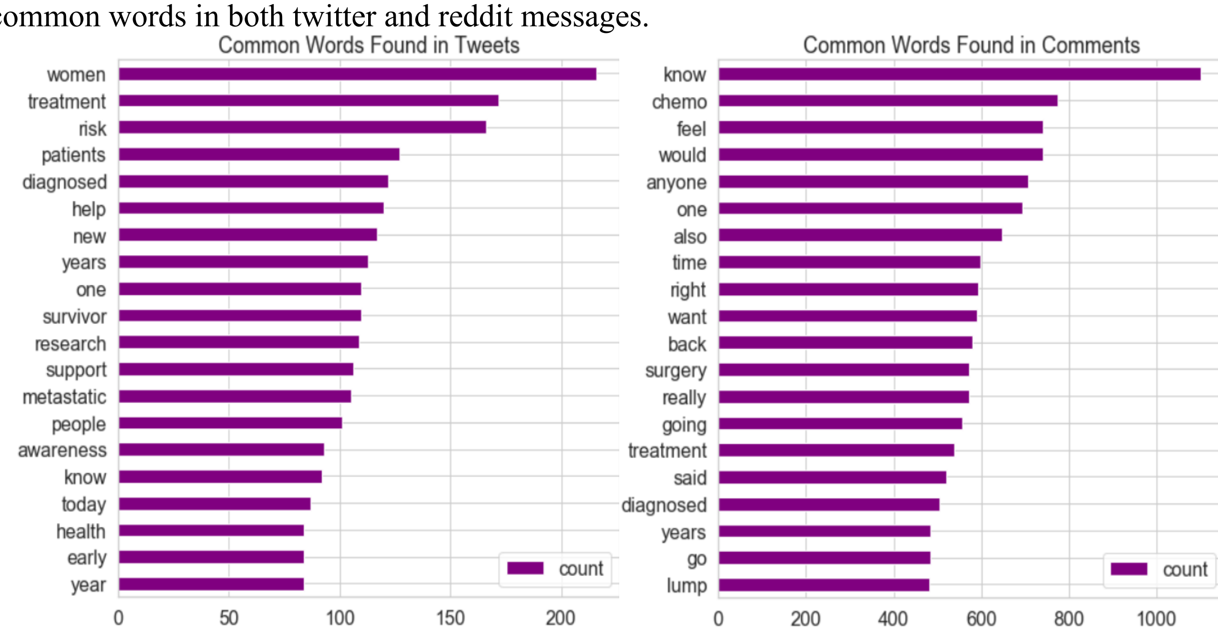


Figure 4.1: 20 most common words in Tweets (a) and Comments (b).

Stop words and collection words ("breast" and "cancer") were removed in this step only to visualize word frequencies. As illustrated in figure 4.1, common words vary between Twitter and Reddit. Twitter users mention more patient journey hotspots and words highly related to breast cancer, whereas comments from Reddit users refer more to everyday verbs. This can be due to the character limit in both social media websites (280 vs. 40,000 characters).

The next stage of the code was the data or text-cleaning. Data frames displaying the data cleaning using all these steps, i.e., removal of punctuation, tokenization, removal of stop words, and lemmatization, for the Twitter data set is shown in figure 4.1.2 and Reddit in figure 4.1.3.

	author	tweets	time	location	text length	clean tweets	text_nopunct	text_tokenized	text_nostop	text_lemmatized
0	ClownsAreEvil	@beaudyk Yes! My mom did pink at 86 in honour ...	6/15/2021 20:55	Toronto, Ontario	82	Yes My mom did pink at 86 in honour of her sis...	Yes My mom did pink at 86 in honour of her sis...	[yes, my, mom, did, pink, at, 86, in, honour, ...	[yes, mom, pink, 86, honour, sister, breast, c...	[yes, mom, pink, 86, honour, sister, breast, c...
1	tradertates	@Katniss_Amc and major respect for highlightin...	6/15/2021 20:52	Los Angeles, CA	180	and major respect for highlighting the double ...	and major respect for highlighting the double ...	[and, major, respect, for, highlighting, the, ...	[major, respect, highlighting, double, standar...	[major, respect, highlighting, double, standar...
2	Nikkixxx	Most recently Lainy's sister, Julie, has been ...	6/15/2021 20:52	St Helens	237	Most recently Lainys sister Julie has been dia...	Most recently Lainys sister Julie has been dia...	[most, recently, lainys, sister, julie, has, b...	[recently, lainys, sister, julie, diagnosed, b...	[recently, lainys, sister, julie, diagnosed, b...

Figure 4.1.2: Text-cleaning steps performed to Twitter data set

	Author	Body	Publish Date	text length	clean comments	text_nopunct	text_tokenized	text_nostop	text_lemmatized
2	HoudinisBox	Hello. I'm 39 years old and was told Friday n...	2020-03-16 09:36:50	622	Hello Im 39 years old and was told Friday nigh...	Hello Im 39 years old and was told Friday nigh...	[hello, im, 39, years, old, and, was, told, fr...	[39, years, old, told, friday, night, gynecolo...	[39, year, old, told, friday, night, gynecolog...
3	7BrdgesRd	In one week my sister went from needing a lump...	2020-03-17 07:50:22	517	In one week my sister went from needing a lump...	In one week my sister went from needing a lump...	[in, one, week, my, sister, went, from, needin...	[one, week, sister, went, needing, lumpectomy,...	[one, week, sister, went, needing, lumpectomy,...
10	Lemoa_	Hi, I'm really worried. I'm almost 20 and toda...	2020-03-18 04:23:28	508	Hi Im really worried Im almost 20 and today I ...	Hi Im really worried Im almost 20 and today I ...	[hi, im, really, worried, im, almost, 20, and,...	[really, worried, almost, 20, today, felt, kin...	[really, worried, almost, 20, today, felt, kin...

Figure 4.1.3: Text-cleaning steps performed to Reddit data set.

The output of this data frame is the tokenized form of comments and tweets. These tokenized contents are converted to vectors using TF-IDF Vectorizer, Count Vectorizer, and N-grams Vectorizer. After creating the sparse matrices with three different techniques, the next task of the code was to create word cloud visualizations from the lemmatized text. Figure 4.1.5 illustrates the word clouds obtained from both social media websites and includes 15 of the most common words without stop words or collection words.

Figure 4.1.5: Word Clouds generated from (a) Twitter and (b) Reddit.

From the two-word clouds generated, it can be noticed that again Twitter provides more concise and breast cancer-related words, whereas Reddit slightly deviates from the main topic with the inclusion of some other words like Friday, night, week, one, newly, and almost. The word "old" is considered of value here because in most of the comments when people were explaining their situation, they self-identified and included age information and previous conditions.

Next step was to determine the mixture of topics mentioned in these tweets and comments. This topic determination is performed using topic modeling. Figure 4.2 illustrates the 10 topics obtained with their word weights. Graphical representation of topics' word counts, and word weights are contained in the Appendix.

```
[(0,
  '0.044*"cancer" + 0.043*"breast" + 0.006*"get" + 0.005*"free" + '
  '0.005*"helped" + 0.005*"site" + 0.005*"clicked" + 0.005*"got" + '
  '0.004*"time" + 0.004*"need"'),
 (1,
  '0.055*"cancer" + 0.052*"breast" + 0.009*"year" + 0.005*"treatment" + '
  '0.005*"people" + 0.005*"patient" + 0.005*"risk" + 0.004*"breastcancer" + '
  '0.004*"amp" + 0.004*"help"'),
 (2,
  '0.036*"cancer" + 0.035*"breast" + 0.012*"risk" + 0.006*"life" + '
  '0.005*"breastcancer" + 0.005*"treatment" + 0.004*"medical" + '
  '0.004*"research" + 0.004*"know" + 0.004*"help"'),
 (3,
  '0.061*"cancer" + 0.052*"breast" + 0.008*"woman" + 0.007*"treatment" + '
  '0.006*"survivor" + 0.005*"patient" + 0.005*"breastcancer" + 0.004*"year" + '
  '0.004*"mother" + 0.004*"help"'),
 (4,
  '0.056*"breast" + 0.055*"cancer" + 0.009*"survivor" + 0.008*"treatment" + '
  '0.005*"woman" + 0.005*"patient" + 0.005*"risk" + 0.004*"diagnosed" + '
  '0.004*"year" + 0.004*"support"'),
 (5,
  '0.042*"cancer" + 0.041*"breast" + 0.009*"amp" + 0.006*"one" + 0.005*"make" '
  '+ 0.005*"helped" + 0.004*"metastatic" + 0.004*"free" + 0.004*"site" + '
  '0.004*"clicked"'),
 (6,
  '0.055*"cancer" + 0.049*"breast" + 0.005*"woman" + 0.004*"medical" + '
  '0.004*"day" + 0.004*"could" + 0.004*"may" + 0.004*"year" + 0.004*"covid" + '
  '0.003*"oncology"'),
 (7,
  '0.061*"cancer" + 0.052*"breast" + 0.009*"amp" + 0.009*"woman" + '
  '0.006*"research" + 0.005*"treatment" + 0.005*"patient" + 0.005*"one" + '
  '0.005*"year" + 0.004*"today"'),
 (8,
  '0.064*"breast" + 0.059*"cancer" + 0.008*"year" + 0.007*"get" + '
  '0.006*"woman" + 0.006*"amp" + 0.005*"men" + 0.005*"new" + 0.004*"diagnosed" '
  '+ 0.004*"time"'),
 (9,
  '0.075*"cancer" + 0.062*"breast" + 0.009*"breastcancer" + 0.009*"woman" + '
  '0.005*"screening" + 0.005*"get" + 0.004*"diagnosed" + 0.004*"dont" + '
  '0.003*"survivor" + 0.003*"metastatic"')]
```

```
[(0,
  '0.009*"cancer" + 0.008*"month" + 0.008*"breast" + 0.007*"know" + '
  '0.007*"one" + 0.007*"dont" + 0.006*"pain" + 0.005*"really" + 0.005*"week" + '
  '0.005*"feel"'),
 (1,
  '0.013*"breast" + 0.009*"cancer" + 0.007*"get" + 0.007*"treatment" + '
  '0.007*"year" + 0.007*"would" + 0.007*"feel" + 0.007*"chemo" + 0.006*"know" '
  '+ 0.006*"dont"'),
 (2,
  '0.010*"know" + 0.009*"cancer" + 0.009*"breast" + 0.008*"week" + '
  '0.007*"feel" + 0.007*"one" + 0.006*"dont" + 0.006*"chemo" + 0.006*"get" + '
  '0.006*"day"'),
 (3,
  '0.027*"cancer" + 0.027*"breast" + 0.007*"get" + 0.007*"year" + 0.006*"know" '
  '+ 0.006*"one" + 0.006*"dont" + 0.005*"week" + 0.005*"lump" + 0.005*"feel"'),
 (4,
  '0.011*"breast" + 0.009*"cancer" + 0.008*"get" + 0.007*"anyone" + '
  '0.007*"feel" + 0.006*"know" + 0.006*"week" + 0.006*"would" + 0.005*"month" '
  '+ 0.005*"doctor"'),
 (5,
  '0.023*"breast" + 0.015*"cancer" + 0.007*"right" + 0.007*"year" + '
  '0.006*"anyone" + 0.006*"get" + 0.005*"know" + 0.005*"feel" + 0.005*"biopsy" '
  '+ 0.005*"doctor"'),
 (6,
  '0.019*"breast" + 0.011*"cancer" + 0.009*"get" + 0.007*"really" + '
  '0.007*"year" + 0.007*"dont" + 0.006*"also" + 0.006*"feel" + 0.006*"know" + '
  '0.006*"one"'),
 (7,
  '0.017*"cancer" + 0.015*"breast" + 0.007*"chemo" + 0.007*"know" + '
  '0.007*"feel" + 0.006*"lump" + 0.006*"week" + 0.006*"anyone" + 0.006*"year" '
  '+ 0.005*"still"'),
 (8,
  '0.016*"breast" + 0.016*"cancer" + 0.010*"know" + 0.008*"get" + 0.007*"week" '
  '+ 0.006*"time" + 0.006*"going" + 0.006*"back" + 0.005*"go" + 0.005*"chemo"'),
 (9,
  '0.010*"dont" + 0.009*"cancer" + 0.008*"breast" + 0.008*"know" + 0.007*"get" '
  '+ 0.007*"chemo" + 0.006*"feel" + 0.006*"want" + 0.006*"week" + '
  '0.006*"thing"')]
```

Figure 4.2: Ten topics obtained from LDA Analysis for (a) Twitter and (b) Reddit.

In LDA models, each document is composed of multiple topics, and typically, only one of the topics is dominant. Figures 4.3 (Twitter) and 4.4 (Reddit) exhibit these dominant topics for each sentence and show the weight of the topics and the keywords.

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	3.0	0.8875	cancer, breast, woman, treatment, survivor, pa...	[yes, mom, pink, honour, sister, breast, cancer]
1	1	4.0	0.9470	breast, cancer, survivor, treatment, woman, pa...	[major, respect, highlighting, double, standar...
2	2	4.0	0.9571	breast, cancer, survivor, treatment, woman, pa...	[recently, lainys, sister, julie, diagnosed, b...
3	3	2.0	0.9250	cancer, breast, risk, life, breastcancer, trea...	[proinflammatory, diet, may, increase, breast,...
4	4	0.0	0.9100	cancer, breast, get, free, helped, site, click...	[dont, want, risk, breast, cancer, wipe, reapp...
5	5	3.0	0.9640	cancer, breast, woman, treatment, survivor, pa...	[immunotherapy, great, treatment, option, type...
6	6	9.0	0.9500	cancer, breast, breastcancer, woman, screening...	[running, august, raise, fund, new, breast, ca...
7	7	4.0	0.9608	breast, cancer, survivor, treatment, woman, pa...	[diagnosed, breast, cancer, early, detection, ...
8	8	7.0	0.7477	cancer, breast, amp, woman, research, treatmen...	[donar, carro, ayuda, breast, cancer, tel, don...
9	9	8.0	0.9400	breast, cancer, year, get, woman, amp, men, ne...	[lady, store, fuck, getting, bunch, candyme, f...

Figure 4.3: Twitter's dominant topic with percentage contribution for different documents.

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	2.0	0.7250	know, cancer, breast, week, feel, one, dont, c...	[year, old, told, friday, night, gynecologist,...
1	1	1.0	0.3827	breast, cancer, get, treatment, year, would, f...	[one, week, sister, went, needing, lumpectomy,...
2	2	7.0	0.9820	cancer, breast, chemo, know, feel, lump, week,...	[really, worried, almost, today, felt, kind, h...
3	3	5.0	0.9845	breast, cancer, right, year, anyone, get, know...	[breast, cancer, moved, lung, stage, my, year,...
4	4	3.0	0.9866	cancer, breast, get, year, know, one, dont, we...	[partner, mother, year, old, daughter, diagnos...
5	5	5.0	0.9857	breast, cancer, right, year, anyone, get, know...	[ultrasound, done, last, week, finding, lump, ...
6	6	2.0	0.6581	know, cancer, breast, week, feel, one, dont, c...	[wanted, share, happened, last, hour, wanted, ...
7	7	2.0	0.9643	know, cancer, breast, week, feel, one, dont, c...	[everyoneampx, bin, december, started, feeling...
8	8	7.0	0.9859	cancer, breast, chemo, know, feel, lump, week,...	[hey, everyone, name, frankie, student, univer...
9	9	6.0	0.9883	breast, cancer, get, really, year, dont, also,...	[yo, female, prior, history, cancer, family, y...

Figure 4.4: Reddit's dominant topic with percentage contribution for different documents.

As illustrated in figure 4.2, words in each topic have different weights, and topics in each document have different percentages (figures 4.3 and 4.4). In general, topics touch on different elements of breast cancer. For instance, Twitter's first topic contains the words "cancer", "breast", "get", "free", "helped", "site", "clicked", "got", "time", and "need" which after performing a deep dive analysis in the dataset demonstrated to be a website where people clicked and helped in the breast cancer cause for free. Similarly, Reddit's first topic containing the words "cancer", "month", "breast", "know", "one", "dont", "pain", really", "week", "feel" was more related to people posting about detecting a breast pain and narrating how they felt, along with the time stamp of their process.

As observed throughout the different topics, some of them overlap in the words “breast” and “cancer”. Nonetheless, most topics include at least a word containing a hotspot or interaction of the breast cancer patient journey. Reddit’s topic number 2 includes “treatment” and “chemo”, topic 3 “chemo, topic 4 “lump”, topic 5 “doctor”, topic 6 “biopsy” and “doctor”, topic 7 “chemo” and “lump”, topic 8 “chemo”, topic 9 “chemo”, and topic 10 “chemo”. For Twitter, topic number 2 includes “treatment”, topic 3 “treatment”, “medical” and “research”, topic 4 “survivor”, “treatment”, “diagnosed” and “support”, topic 5 “treatment”, “diagnosed”, topic 6 “metastatic”, topic 7 “oncology”, “covid”, topic 8 “research”, “treatment”, topic 9 “diagnosed”, topic 10 “screening”, “diagnosed”, “survivor” and “metastatic”.

To perform this Sentiment Analysis, several functions were created with the use of TextBlob to obtain polarity and subjectivity values and to label the text lemmatized according to their polarity score. A dataset displaying the clean comments, lemmatized comments along with their polarity, subjectivity and classification was the output of this analysis, as illustrated in Figure 4.3.1 and 4.3.2.

clean tweets	text_lemmatized	Subjectivity	Polarity	Analysis
Yes My mom did pink at 86 in honour of her sis...	['yes', 'mom', 'pink', '86', 'honour', 'sister...	0.3000	-0.1000	Negative
and major respect for highlighting the double ...	['major', 'respect', 'highlighting', 'double',...	0.3800	0.2925	Positive
Most recently Lainys sister Julie has been dia...	['recently', 'lainys', 'sister', 'julie', 'dia...	0.3125	-0.0625	Negative
Proinflammatory diets may increase breast canc...	['proinflammatory', 'diet', 'may', 'increase',...	0.0000	0.0000	Neutral
For those who dont want to risk breast cancer ...	['dont', 'want', 'risk', 'breast', 'cancer', '...	1.0000	0.0000	Neutral

Figure 4.3.1: Output data frame from Twitter Sentiment Analysis.

clean comments	text_lemmatized	Subjectivity	Polarity	Analysis
Hello Im 39 years old and was told Friday nigh...	['39', 'year', 'old', 'told', 'friday', 'night...	0.223661	0.095617	Positive
In one week my sister went from needing a lump...	['one', 'week', 'sister', 'went', 'needing', '...	0.517857	0.128571	Positive
Hi Im really worried Im almost 20 and today I ...	['really', 'worried', 'almost', '20', 'today',...	0.388146	0.052670	Positive
Breast cancer moved to lungs stage 4My 74 year...	['breast', 'cancer', 'moved', 'lung', 'stage',...	0.349722	0.057500	Positive
My partner 30 and the mother of my 3 year old ...	['partner', '30', 'mother', '3', 'year', 'old'...	0.460548	0.111713	Positive

Figure 4.3.2: Output data frame from Reddit Sentiment Analysis.

A pie chart and other visualizations for comparison were created with the use of the matplotlib library. Figure 4.3.3 depicts the pie charts obtained from both datasets.

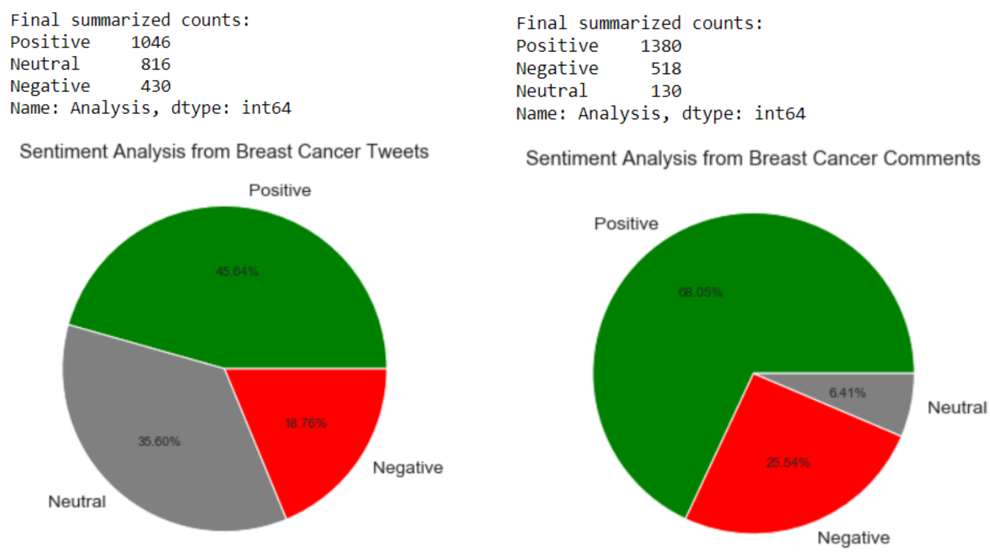


Figure 4.3.3: Pie charts with polarity classification from (a) Twitter and (b) Reddit.

As depicted by the pie charts above, the percentage of positive polarity tweets is 45.64% (n=1,046) compared to 68.05% (n=1,380) from subreddits. Neutral polarity tweets represent more than one-third of the records with a percentage of 35.60% (n=816) whereas neutral comments from Reddit represent less than one-tenth of the total subreddits with 6.41% (n=130). Finally, negative polarity tweets constitute only 18.76% (n=430) of the total tweets, versus a 25.54% (n=518) of negative comments from subreddits.

Further analysis was conducted to comprehend why Twitter and Reddit users would comment positively about breast cancer. Polarity and subjectivity distributions for tweets (blue) and comments (red) were plotted as part of these analyses, as shown in Figure 4.3.4.

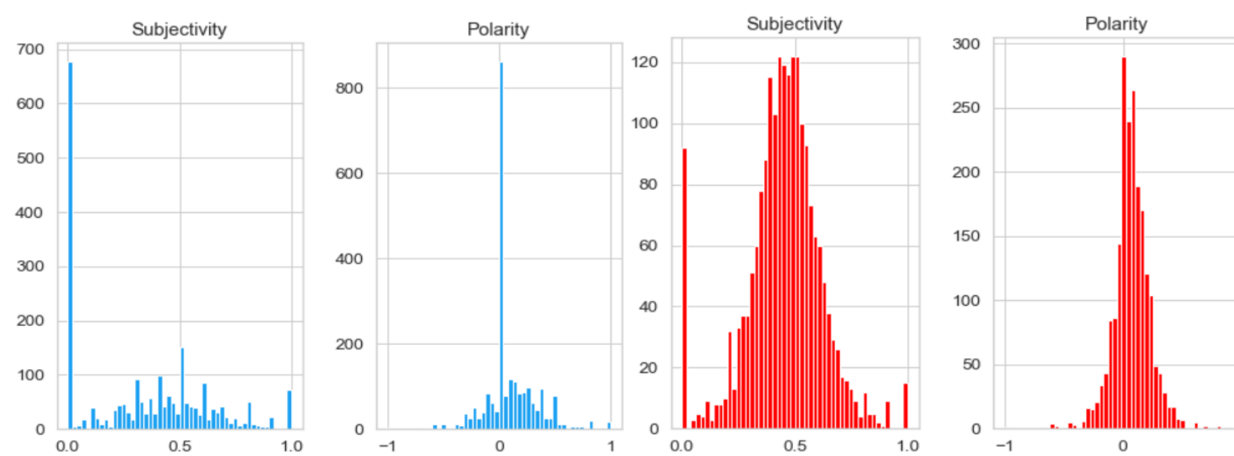


Figure 4.3.4: Subjectivity and Polarity plots for Twitter (Blue) and Reddit (Red).

As observed in these subjectivity plots, there is a higher number of objective comments coming from Twitter users (value of 0 with a frequency higher to 650 in the subjectivity chart) compared to Reddit ones (value of 0 has a frequency of approximately 90). Higher Twitter objectivity is also portrayed by the distribution of Twitter's subjectivity having a mean of 0.3407 (std = 0.289), compared to Reddit's mean of 0.4482 (std = 0.173)

Regarding the polarity bar charts, it is observed that Twitter's distribution is skewed to the right with a mean of 0.0914 (std = 0.241), whereas Reddit's distribution has a mean of 0.0694 (std = 0.158), denoting for a higher number of positive comments. Even though Twitter's distribution is slightly skewed to the right, the distribution resembles a normal one. Overall, we can observe that polarity has a higher density around 0 (neutrality), which implies that despite comments being classified as "positive" or "negative", there are ranges of positivity and negativity.

To illustrate the degrees of positivity, neutrality, and negativity, some tweets and comments from the sentiment analysis outputs are depicted in Tables 4.3.1 and 4.3.2.

Table 4.3.1: Examples of different degrees of polarity from Tweets.

Cleaned Tweets	Subjectivity	Polarity	Analysis
Shocking news just found out a friend has been diagnosed with breast cancer Reminder to feel those boobies	1	**-1**	Negative
Men and women can both benefit from moderate alcohol intake in terms of cardiovascular health but women might also suffer from a terrible side effect an increased chance of breast cancer If at all alcohol should be drunk in moderation by both genders	0.9	**-0.5**	Negative
Study ties vitamin D to health outcomes in breast cancer	0	**0**	Neutral
A Womans Diet Might Help Her Avoid Breast Cancer	0	**0**	Neutral
I would love this 5k to help with bills while I start my breast cancer journey Anything would help 5Gsfor5G Contest	0.6	**0.5**	Positive
Ohhh lmdao Its an awesome shirt in support of breast cancer	1	**1**	Positive
Come to MD Anderson in Houston TX Best Hospital in the world for Breast Cancer	0.3	**1**	Positive

Table 4.3.2: Examples of different degrees of polarity from Subreddits.

Cleaned Comments	Subjectivity	Polarity	Analysis
Is fatigue the worst part of chemo	1	**-1**	Negative
After a lumpectomy and radiation I started anastrozole to be taken for five years Ive been on it for about a month and its been awful I have headaches every day and terrible insomnia every night My doctor has advised me to stop taking it for seven days to see how I feel and then well discuss next steps Has anyone else had this type of reaction to this drug Thank you in advance for any assistance	0.666667	**-0.66667**	Negative
Hi All My mum had stage 1 breast cancer and needed surgery last october They removed the tumor and 1 node from her armpit Luckily she is now clear however she still has pain in her chest and arm area I think this is all nerve pain from the surgery and this was confirmer by her specialist however she was only prescribed paracetamol which does nothing for nerve pain While she tries to hide most of her pain I know this is greatly annoying her and causing discomfort Due to covid her follow up appointment was moved as shes a vulnerable person so she cant really get any help re medication for bow In the meantime is there something she could use to alleviate the pain such as a shoulder heat mat organic cream etc. Suggestions are highly appreciated	0.360714	**-0.13214**	Negative
Hello I am 23So for almost 34 years ago I have discovered a hard big lump right behind my nipples on my left boob And also it is only on my left breast and my left breast is larger compared to the right breast And since a year I get weird on and off pain in my shoulder back hand and on my breast itself left side When I lie down it gets better but it is on amp off everyday I am too scared to go for a checkup I have been stalling it for years and I have no courage to go because I feel it is Breast cancer amp also I am praying it is fibrocystic breast since they have kinda same symptoms too Please anyone experienced this let me know Please I am freaking out so bad	0.419983	**-0.00145**	Negative
Hi allI am trying to prep for an upcoming double mastectomy Whats your favorite compression bra postsurgery	0.5	**0.25**	Positive
Hello The title says it all I just want to know what should I know and educate myself to make sure I can support my mom the best as I can Do you also have any advices for me as her daughter Is there anything I can do to make her feel comfortable What are the steps to come She just been diagnosed today Help me please	0.662963	**0.633333**	Positive

As these tweets and subreddits demonstrate, values of polarity depend on the emotion conveyed by the Twitter or Reddit user and what is captured by the lexicon-based sentimental analysis algorithm.

Additionally, as observed in Tables 4.3.1 and 4.3.2, tweets and subreddit comments are not necessarily pertaining to a breast cancer patient's journey map. As mentioned earlier, the use of social media scraping opens the possibility of getting information from different users commenting about the issue or related issues, as the tweet about "awesome shirt in support of breast cancer".

4.2 Research Application

One of the main goals for this research was to link comments from Twitter and Reddit to specific hotspots. In this way, experts can provide evidence-based recommendations from social media at specific hotspots of the breast cancer patient journey map. Figure from section 3.2.1.1 was modified to include Twitter and Reddit comments from specific breast cancer patient journey hotspots. In figure 4.2, Twitter comments are depicted on the top part in blue, whereas Reddit comments are displayed in the bottom part of the figure in red.

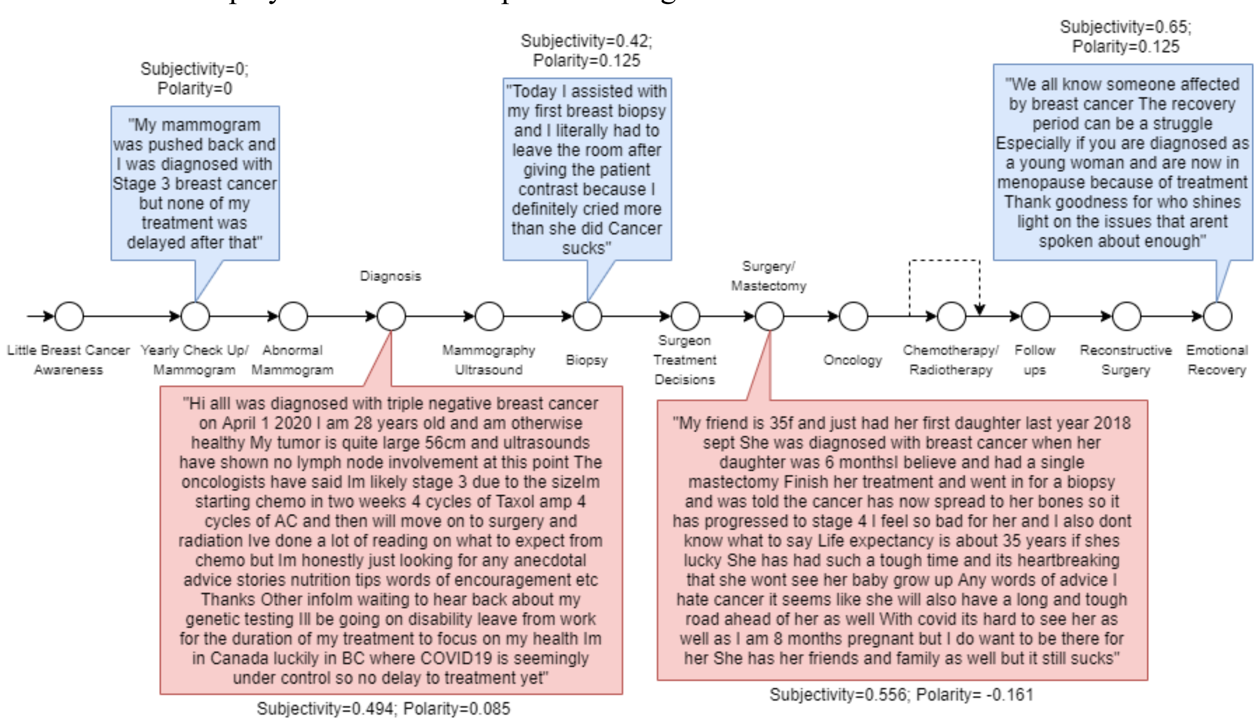


Figure 4.2: Comments from Twitter (blue) and Reddit (red) linked to the breast cancer PJM.

As illustrated in Figure 4.2, some recommendations regarding delays with mammograms, emotional support for health professionals, support people and patients, and chemotherapy support groups can be made to alleviate what people are feeling and commenting about breast cancer. Something worth mentioning is that from all those comments, only one was classified with a negative polarity, meaning that negative emotions and experiences can be present even in positively classified comments.

Chapter 5: Discussion

This study included the analysis of tweets and subreddits with the keyword “breast cancer”. Analyzed records from both data sets showed an overall positive sentiment associated with breast cancer. There can be many reasons associated with this, first, the degree of polarity. Most comments were classified as ‘positive’ due to having a positive polarity, even though they were closer to 0 (neutrality) than to 1 (positivity). Secondly, the use of social media data sets by keywords allows any Twitter or Reddit user to post about the topic without being a patient or a patient’s support person. Therefore, users posting about breast cancer range from patients, support people, and health professionals to people organizing fundraisers, volunteering, buying clothes, and more. This reason makes it complicated for social media to provide reliable information that can be used to map specific hotspots and interactions. Nonetheless, as observed in the results from the LDA analysis, latent topics include breast cancer patient journey hotspots and interactions. However, the accuracy of these topic’s weights might not be suitable to use as a base for mapping a patient journey. Results obtained from the EDA also illustrate the variation of common words between Twitter and Reddit. Twitter users mention more breast cancer patient journey hotspots and have a higher degree of objectivity, whereas comments from Reddit users include more everyday verbs. The reason behind this can be the character limit from both social media websites (280 vs. 40,000 characters). Overall, this study using NLP algorithms to evaluate patient experience has found more patient hotspots and objective comments from Twitter. However, the literature suggests that the ability of patients to explain their experiences in their own words helps qualitative methods to gain better insights into patient perceptions, behaviors, and the meaning they assign to certain experiences. Based on this, it could be said that Reddit users can express themselves more freely in their comments because they do not need to limit themselves to 280 characters as Twitters users do. Furthermore, the use of AI to capture patient experience throughout the patient's journey during the healthcare continuum provides valuable insights to improve individualized, empathetic, and respectful care in clinical systems.

5.1 Social Media Limitations

There are many challenges and limitations associated with social media mining, including:

- Non-standard text/content: social media often involves a combination of text, URLs, hashtags, a mix of capitalized words, abbreviations, punctuation, and emojis.
- Text mining: pulling relevant information out of the unstructured text.
- Lack of consistent geolocation information: not all social media accounts are spatially located, or sometimes geolocation is provided in non-standard formats.
- Large data volumes: collecting tweets or comments can result in millions of records to sort through.
- API limitations: Twitter standard API only allows you to retrieve tweets from 6-9 days ago and is limited to scraping 18,000 tweets per API call and is divided in 15-minute windows for requests. Also, using Tweepy's Twitter API only allows you to return up to 3,200 of a user's most recent tweets [197].
- Lack of representation: social media is not a representation of an entire population. Furthermore, social media does not have the same amount of use and popularity among countries.
- Text length: each tweet is delimited to 280 characters, which might be challenging for some users to fully explain what they intend to.
- Topic accuracy: a method to determine topics is the use of hashtags, nonetheless these hashtags are determined by users and might not be accurate.
- Incorrect English: social media use is informal, and as such, people use incorrect English in the form of slang, abbreviations, lengthening of words etcetera, which makes it difficult to conduct text analyses and causes incorrect classification of sentiments or opinions.

- Data sparsity: since many tweets and comments have misspellings or words that are not recognized, those words will not be properly accounted for when analyses are performed.
- Negation: detecting sarcasm and other double negatives can be challenging for text analytic tools, especially sentiment analysis because if not properly accounted for, the result can be the opposite of the message's true polarity.
- Multilingual content: social media is used worldwide, therefore tweets or comments can be a mix of foreign languages. In this case, additional steps need to be added to the extraction process to account for only tweets in the same language or to clean tweets including more than one.

5.2 Study Limitations

Besides social media limitations, this study also found several ones. First, participants of this study were Twitter and Reddit users commenting about "breast cancer". Therefore, the scope of this study was not only related to patients, resulting in mixed findings that do not reflect the desired accuracy for mapping the breast cancer patient journey and experiences. Secondly, descriptive information about sample participants was not collected, limiting the sample descriptors to observable data. Thirdly, not all breast cancer patients or support people have access to social media and/or comment about this condition online. Also, many patients with breast cancer suffer from psychological and cognitive impairments, changes in body image and sexuality, fear of reoccurrences, economic stress, poor social support, role functioning constraints and family crises throughout their disease journey [198]. These difficulties are associated with their perceived health-related stigma, defined by Goffman as an attribute that links an individual to an undesirable stereotype or in this case, a disease [199]. Stigmatization includes negative emotions and attitudes towards the affected individual, as well as social avoidance [200]. Marlow and Wardle found that perceived health-related stigma is associated with certain socio-demographic factors [201]. Being male was one of these factors along with younger ages. For the most part, being male was correlated with higher cancer-related stigma, affecting areas as social responsibility, social avoidance, and financial discrimination. People with younger ages demonstrated higher scores for personal responsibility, awkwardness, social avoidance, and financial discrimination as well. Perceived health-related stigma represents another limitation for this study, given that patients who are feeling this stigma do not want to receive more negative attitudes and passive avoidances. This stigma also influences what people and social media users share about this topic. The lack of free expression due to character limit as a barrier to explain themselves as social media users is another limitation found for this study. Lastly, the COVID-19 pandemic eliminated the possibility to interview and collect data directly from patients, which translated into extracting data from social media.

Chapter 6: Conclusion

The analysis of tweets and subreddits from Twitter and Reddit users demonstrated that the scope of this study was not only patients, but also support people, breast cancer advocates, and people interested in the topic. Results obtained did not reflect the desired accuracy for mapping the breast cancer patient journey and experiences. However, results from this study were able to identify the latent topics and overall sentiments associated with breast cancer in general. Additionally, this work exhibited useful insights and advice for improving patients' and support people's experiences at breast cancer patient journey hotspots.

6.1 Recommendations for Future Work

Future work on this thesis can include loading more data records from both social media sites and perhaps incorporating some other social media networks. Also, code optimization can be done by changing unsupervised to supervised learning. For instance, changing the Topic Modeling to Topic Classification and the lexicon-based Sentimental Analysis to supervised Sentimental Analysis. The inclusion of more keywords in the data extraction process could also result in a higher number of latent topics and the possibility to link those keywords with specific hotspots. Another implementation could be the use of the timestamps from both subreddits and tweets to try to map a patient's journey using real timeframes. Above all, if the main objective of future work is that of automating and improving a way to measure patient experience with NLP Algorithms following the patients' journey maps, the data source should be obtained directly from patients via semi-structured or unstructured interviews for a more robust qualitative analysis and the ability to accurately map patient journeys and hotspots.

References

[1] Deloitte, "2021 Global Health Care Outlook," 2021. [Online]. Available: https://documents.deloitte.com/insights/Globalhealthcareoutlook.

[2] I. of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: The National Academies Press, 2001.

[3] The Beryl Institute, "Defining Patient Experience." .

[4] L. M. Philpot *et al.*, "Creation of a Patient-Centered Journey Map to Improve the Patient Experience: A Mixed Methods Approach," *Mayo Clin. Proc. Innov. Qual. Outcomes*, vol. 3, no. 4, pp. 466–475, 2019, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2542454819300852.

[5] N. Whitt, R. Harvey, G. McLeod, and S. Child, "How many health professionals does a patient see during an average hospital stay?," *N. Z. Med. J.*, vol. 120, no. 1253, p. U2517, May 2007.

[6] C. Williams, Kathryn and Sansoni, Janet and Darcy, M and Grootemaat, Pam and Thompson, "Patient-reported outcome measures," *Lit. Rev. Sydney Aust. Comm. Saf. Qual. Heal. Care*, 2016.

[7] S. C. Madathil, A. J. Lopes, and M. Alfred, "Patient Journey Mapping: A Literature Review," in *IIE Annual Conference. Proceedings*, 2020, pp. 937–942.

[8] L. D. Urden, "Patient Satisfaction Measurement: Current Issues and Implications," *Prof. Case Manag.*, vol. 7, no. 5, 2002, [Online]. Available: https://journals.lww.com/professionalcasemanagementjournal/Fulltext/2002/09000/Patient_Satisfaction_Measurement__Current_Issues.6.aspx.

[9] C. Doyle, L. Lennox, and D. Bell, "A systematic review of evidence on the links between

patient experience and clinical safety and effectiveness,” *BMJ Open*, vol. 3, no. 1, 2013, doi: 10.1136/bmjopen-2012-001570.

[10] N. Black, M. Varaganum, and A. Hutchings, “Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery.,” *BMJ Qual. Saf.*, vol. 23, no. 7, pp. 534–542, Jul. 2014, doi: 10.1136/bmjqs-2013-002707.

[11] L. Dubé and K. Menon, “Managing emotions. Accenting the positive might not produce the highest satisfaction payoff.,” *Mark. Health Serv.*, vol. 18, no. 3, pp. 34–42, 1998.

[12] S. Sofaer and K. Firminger, “Patient perceptions of the quality of health services.,” *Annu. Rev. Public Health*, vol. 26, pp. 513–559, 2005, doi: 10.1146/annurev.publhealth.25.050503.153958.

[13] C. Clancy and F. S. Collins, “Patient-Centered Outcomes Research Institute: the intersection of science and health care.,” *Sci. Transl. Med.*, vol. 2, no. 37, p. 37cm18, Jun. 2010, doi: 10.1126/scitranslmed.3001235.

[14] The Health Foundation, “Measuring patient experience,” United Kingdom, 2013. Accessed: Sep. 06, 2021. [Online]. Available: http://www.health.org.uk/publications/measuring-patient-experience.

[15] N. S. Bardach, A. Lyndon, R. Asteria-Peñaloza, L. E. Goldman, G. A. Lin, and R. A. Dudley, “From the closest observers of patient care: a thematic analysis of online narrative reviews of hospitals,” *BMJ Qual. & Saf.*, vol. 25, no. 11, pp. 889 LP – 897, Nov. 2016, doi: 10.1136/bmjqs-2015-004515.

[16] A. Bartu, “CIient satisfaction: Why bother?,” *J. Subst. Misuse*, vol. 1, no. 1, pp. 20–26, 1996, doi: 10.3109/14659899609094715.

[17] M. Beattie, D. J. Murphy, I. Atherton, and W. Lauder, “Instruments to measure patient

experience of healthcare quality in hospitals: a systematic review," *Syst. Rev.*, vol. 4, no. 1, p. 97, 2015, doi: 10.1186/s13643-015-0089-0.

[18] J. Sitzia and N. Wood, "Patient satisfaction: a review of issues and concepts.," *Soc. Sci. Med.*, vol. 45, no. 12, pp. 1829–1843, Dec. 1997, doi: 10.1016/s0277-9536(97)00128-7.

[19] R. Crow *et al.*, "The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.," *Health Technol. Assess.*, vol. 6, no. 32, pp. 1–244, 2002, doi: 10.3310/hta6320.

[20] C. Salisbury, M. Wallace, and A. A. Montgomery, "Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling," *BMJ*, vol. 341, p. c5004, Oct. 2010, doi: 10.1136/bmj.c5004.

[21] W. Levinson, D. L. Roter, J. P. Mullooly, V. T. Dull, and R. M. Frankel, "Physician-Patient Communication: The Relationship With Malpractice Claims Among Primary Care Physicians and Surgeons," *JAMA*, vol. 277, no. 7, pp. 553–559, Feb. 1997, doi: 10.1001/jama.1997.03540310051034.

[22] N. Rave, M. Geyer, B. Reeder, J. Ernst, L. Goldberg, and C. Barnard, "Radical Systems Change: Innovative Strategies to Improve Patient Satisfaction," *J. Ambul. Care Manage.*, vol. 26, no. 2, 2003, [Online]. Available: https://journals.lww.com/ambulatorycaremanagement/Fulltext/2003/04000/Radical_Systems_Change__Innovative_Strategies_to.8.aspx.

[23] D. G. Safran, J. E. Montgomery, H. Chang, J. Murphy, and W. H. Rogers, "Switching doctors: predictors of voluntary disenrollment from a primary physician's practice," *J. Fam. Pract.*, vol. 50, no. 2, p. 130, 2001.

[24] C. Rathert, M. D. Wyrwich, and S. A. Boren, "Patient-centered care and outcomes: a

systematic review of the literature.," *Med. Care Res. Rev.*, vol. 70, no. 4, pp. 351–379, Aug. 2013, doi: 10.1177/1077558712465774.

[25] D. T. Gerteis M, Edgman-Levitan S, Daley J, "Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care," 1993.

[26] K. Davis, S. C. Schoenbaum, and A.-M. Audet, "A 2020 vision of patient-centered primary care," *J. Gen. Intern. Med.*, vol. 20, no. 10, pp. 953–957, Oct. 2005, doi: 10.1111/j.1525-1497.2005.0178.x.

[27] M. MA, "Mapping the Patient Journey Across the Continuum: Lessons Learned From One Patient's Experience.," *J. patient Exp.*, vol. 6, no. 2, pp. 103–107, 2019, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31218254/.

[28] S. A, S. S, B. B, L. P, N. J, and C.-W. R, "Improving outcomes from high-risk surgery: a multimethod evaluation of a patient-centred advanced care planning intervention.," *BMJ Open*, vol. 7, no. 2, p. e014906, 2017, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28242771/.

[29] G. R, M. C, V. D, and T. D, "Exploring the hospital patient journey: What does the patient experience?," *PLoS One*, vol. 14, no. 12, p. e0224899, 2019, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31805061/.

[30] Z. L *et al.*, "How to measure experiences of healthcare quality in Denmark among patients with heart disease? The development and psychometric evaluation of a patient-reported instrument.," *BMJ Open*, vol. 7, no. 10, p. e016234, 2017, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29084787/.

[31] P. Väre, E. Nikiphorou, P. Hannonen, and T. Sokka, "Delivering a one-stop, integrated, and patient-centered service for patients with rheumatic diseases," *SAGE Open Med.*, vol.

4, 2016, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L614975191&from=export.

[32] P. Armanasco, D. Williamson, and B. Yates, "Integration of podiatric surgery within an orthopaedic department: An audit of patient satisfaction with labour force implications," *Foot*, vol. 22, no. 3, pp. 200–204, 2012, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0958259212000466.

[33] T. Q *et al.*, "Comparing patient and healthcare worker experiences during a dengue outbreak in Singapore: understanding the patient journey and the introduction of a point-of-care test (POCT) toward better care delivery.," *BMC Infect. Dis.*, vol. 17, no. 1, p. 503, 2017, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28724363/.

[34] N. Senft and J. Everson, "eHealth Engagement as a Response to Negative Healthcare Experiences: Cross-Sectional Survey Analysis.," *J. Med. Internet Res.*, vol. 20, no. 12, p. e11034, Dec. 2018, doi: 10.2196/11034.

[35] E. Swallmeh, V. Byers, and A. Arisha, "Informing quality in emergency care: understanding patient experiences," *Int J Heal. Care Qual Assur*, vol. 31, no. 7, pp. 704–717, 2018, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L624592092&from=export.

[36] S. K, F. O, K. Å, and E. M, "Not a Straight Line-Patients' Experiences of Prostate Cancer and Their Journey Through the Healthcare System.," *Cancer Nurs.*, vol. 42, no. 1, pp. E36–E43, 2019, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29334523/.

[37] I. Mahé *et al.*, "Patients experience of living with cancer associated thrombosis in France

(Le PELICAN)," *Thromb. Res.*, vol. 194, pp. 66–71, 2020, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0049384820302668.

[38] K. S. Deacon, "Re-building life after ICU: A qualitative study of the patients' perspective," *Rehabil. Dur. Post Crit. Illn.*, vol. 28, no. 2, pp. 114–122, 2012, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0964339711001273.

[39] M. A. Clarke *et al.*, "Toward a patient-centered ambulatory after-visit summary: Identifying primary care patients' information needs," *Informatics Heal. Soc. Care*, vol. 43, no. 3, pp. 248–263, Jul. 2018, doi: 10.1080/17538157.2017.1297305.

[40] M. Williams, "The patient's experience in a plaster cast," *Int. J. Orthop. Trauma Nurs.*, vol. 14, no. 3, pp. 132–141, 2010, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1878124110000444.

[41] N. Kinnear *et al.*, "Patient Satisfaction in Emergency General Surgery: A Prospective Cross-Sectional Study," *World J Surg*, vol. 44, no. 9, pp. 2950–2958, 2020, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L631753284&from=export.

[42] S. N. Rogers, N. Clifford, and D. Lowe, "Patient and carer unmet needs: a survey of the British association of head and neck oncology nurses," *Br. J. Oral Maxillofac. Surg.*, vol. 49, no. 5, pp. 343–348, 2011, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0266435610002081.

[43] V. McGurk *et al.*, "Treating Patients Well: Improving the patient experience," *J. Neonatal Nurs.*, vol. 13, no. 6, pp. 225–230, 2007, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1355184107001585.

[44] D. Collier, “A humanistic approach to paediatric radiation therapy – One family’s journey,” *Radiography*, vol. 19, no. 4, pp. 306–310, 2013, doi: https://doi.org/10.1016/j.radi.2013.07.009.

[45] S. S and C. I, “A Cancer Patient Journey: Complete Review During Acute Treatment Phase.,” *Heal. equity*, vol. 3, no. 1, pp. 403–408, 2019, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31406954/.

[46] R. S. Lazarus and S. Folkman, *Stress, appraisal, and coping*. Springer publishing company, 1984.

[47] F. Timmins, “Exploring the concept of ‘information need,’” *Int. J. Nurs. Pract.*, vol. 12, no. 6, pp. 375–381, Dec. 2006, doi: https://doi.org/10.1111/j.1440-172X.2006.00597.x.

[48] F. Heaney and S. Hahessy, “Patient satisfaction with an orthopaedic pre-operative assessment clinic,” *Int. J. Orthop. Trauma Nurs.*, vol. 15, no. 2, pp. 82–91, 2011, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1878124110001280.

[49] C. Font, A. Nelson, T. Garcia-Fernandez, H. Prout, P. Gee, and S. Noble, “Patients’ Experience of Living with Cancer-associated thrombosis in Spain (PELICANOS),” *Support. Care Cancer*, vol. 26, no. 9, pp. 3233–3239, 2018, doi: 10.1007/s00520-018-4174-5.

[50] S. Timmermans and R. Almeling, “Objectification, standardization, and commodification in health care: A conceptual readjustment,” *Soc. Sci. Med.*, vol. 69, no. 1, pp. 21–27, 2009, doi: https://doi.org/10.1016/j.socscimed.2009.04.020.

[51] M. Kreuzer, V. Cado, and K. Raïes, “Moments of care: How interpersonal interactions contribute to luxury experiences of healthcare consumers,” *J. Bus. Res.*, vol. 116, pp. 482–490, 2020, [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S0148296319307027.

[52] K. Easton, T. Morgan, and M. Williamson, "Medication safety in the community: a review of the literature," *Sydney Natl. Prescr. Serv.*, 2009.

[53] A. Tariq *et al.*, "Examining barriers to healthcare providers' adoption of a hospital-wide electronic patient journey board," *Int. J. Med. Inform.*, vol. 114, pp. 18–26, 2018, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505618302132.

[54] M. J. Kelly, T. D. R. Lloyd, D. Marshall, G. Garcea, C. D. Sutton, and M. Beach, "A snapshot of MDT working and patient mapping in the UK colorectal cancer centers in 2002," *Color. Dis.*, vol. 5, no. 6, pp. 577–581, 2003, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L37455624&from=export.

[55] D. Doherty, S. M. Benbow, J. Craig, and C. Smith, "Patients' and carers' journeys through older people's mental health services: Powerful tools for learning," *Dementia*, vol. 8, no. 4, pp. 501–513, Nov. 2009, doi: 10.1177/1471301209350286.

[56] E. J. S. Hovenga, C. Lowe, E. J. S. Hovenga, and C. Lowe, "Chapter 11 - Measuring health service quality," pp. 355–388, 2020, [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128169773000113.

[57] H. Strachan, "Person-centred caring: its conceptualisation and measurement through three instruments (personalisation, participation and responsiveness)." Glasgow Caledonian University, 2016.

[58] A. N. Association, "Nursing Care Report Card for Acute Care," Washington, DC, 1995.

[59] C. Tarrant, K. Windridge, R. Baker, G. Freeman, and M. Boulton, "'Falling through

gaps': primary care patients' accounts of breakdowns in experienced continuity of care," *Fam. Pract.*, vol. 32, no. 1, pp. 82–87, Feb. 2015, doi: 10.1093/fampra/cmu077.

[60] J. L. Haggerty, R. J. Reid, G. K. Freeman, B. H. Starfield, C. E. Adair, and R. McKendry, "Continuity of care: a multidisciplinary review," *BMJ*, vol. 327, no. 7425, pp. 1219 LP – 1221, Nov. 2003, doi: 10.1136/bmj.327.7425.1219.

[61] L. Newton and C. Sulman, "Use of text messaging to improve patient experience and communication with pediatric tonsillectomy patients," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 113, pp. 213–217, 2018, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165587618303720.

[62] J. Blennerhassett and J. Hilbers, "Medicine management in older people from non-english speaking backgrounds," *J. Pharm. Pr. Res.*, vol. 41, no. 1, pp. 33–36, 2011, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L362160761&from=export.

[63] V. Nanton, A. Docherty, C. Meystre, and J. Dale, "Finding a pathway: Information and uncertainty along the prostate cancer patient journey," *Br. J. Health Psychol.*, vol. 14, no. 3, pp. 437–458, Sep. 2009, doi: https://doi.org/10.1348/135910708X342890.

[64] M. P. Flattery, J. M. Pinson, L. Savage, and J. Salyer, "Living with pulmonary artery hypertension: Patients' experiences," *Hear. Lung*, vol. 34, no. 2, pp. 99–107, 2005, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0147956304001360.

[65] J. L. Wolff, M. L. Clayman, P. Rabins, M. A. Cook, and D. L. Roter, "An exploration of patient and family engagement in routine primary care visits," *Heal. Expect.*, vol. 18, no.

2, pp. 188–198, Apr. 2015, doi: https://doi.org/10.1111/hex.12019.

[66] R. L. Street and H. S. Gordon, "Companion participation in cancer consultations," *Psychooncology.*, vol. 17, no. 3, pp. 244–251, Mar. 2008, doi: https://doi.org/10.1002/pon.1225.

[67] D. D, B. SM, C. J, and S. C, "Patients' and carers' journeys through older people's mental health services: powerful tools for learning.," *Dement.*, vol. 8, no. 4, pp. 501–513, 2009, [Online]. Available: https://doi.org/10.1177/1471301209350286.

[68] M. Blair *et al.*, "Multi-method evaluation of a paediatric ambulatory care unit (PACU): impact on families and staff," *Arch. Dis. Child.*, vol. 93, no. 8, pp. 681 LP – 685, Aug. 2008, doi: 10.1136/adc.2006.115386.

[69] L. M. Hunt and N. H. Arar, "An Analytical Framework for Contrasting Patient and Provider Views of the Process of Chronic Disease Management," *Med. Anthropol. Q.*, vol. 15, no. 3, pp. 347–367, Sep. 2001, doi: https://doi.org/10.1525/maq.2001.15.3.347.

[70] M. Singer and M. R. Weeks, "Preventing AIDS in Communities of Color: Anthropology and Social Prevention," *Hum. Organ.*, vol. 55, no. 4, pp. 488–492, Jun. 1996, [Online]. Available: http://www.jstor.org/stable/44127868.

[71] T. Tjørnhøj-Thomsen, "Framing the clinical encounter for greater understanding, empathy, and success," *Hear. J.*, vol. 62, no. 8, 2009, [Online]. Available: https://journals.lww.com/thehearingjournal/Fulltext/2009/08000/Framing_the_clinical_encounter_for_greater.7.aspx.

[72] J. Cox, C. Cormack, M. Prendergast, H. Celestino, S. Willis, and M. Witteveen, "Patient and provider experience with a new model of care for primary hip and knee arthroplasties," *Int. J. Orthop. Trauma Nurs.*, vol. 20, pp. 13–27, 2016, [Online].

Available: http://www.sciencedirect.com/science/article/pii/S1878124115000313.

[73] J. M. Brocki and A. J. Wearden, “A critical evaluation of the use of interpretative phenomenological analysis (IPA) in health psychology,” *Psychol. Health*, vol. 21, no. 1, pp. 87–108, Feb. 2006, doi: 10.1080/14768320500230185.

[74] T. van der Velden, “J. Ovretveit, Evaluating health interventions, Open University Press: Buckingham, 1998, 321, pp., ISBN 0 335 19964 X (paperback), ISBN 0 335 19965 8 (hardcover); £19.99,” *Int. J. Health Plann. Manage.*, vol. 16, no. 4, pp. 370–372, Oct. 2001, doi: https://doi.org/10.1002/hpm.645.

[75] M. Kalluri, F. Luppi, and G. Ferrara, “What Patients With Idiopathic Pulmonary Fibrosis and Caregivers Want: Filling the Gaps With Patient Reported Outcomes and Experience Measures.,” *Am. J. Med.*, vol. 133, no. 3, pp. 281–289, 2020, [Online]. Available: https://doi.org/10.1016/j.amjmed.2019.08.032.

[76] U. F. and D. A. (FDA), “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims,” 2009.

[77] T. Weldring and S. M. S. Smith, “Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs),” *Heal. Serv. insights*, vol. 6, pp. 61–68, Aug. 2013, doi: 10.4137/HSI.S11093.

[78] D. of Health, “Using the Commissioning for Quality and Innovation (CQUIN) payment framework.” Department of Health London, 2008.

[79] H. Lee, I. Vlaev, D. King, E. Mayer, A. Darzi, and P. Dolan, “Subjective well-being and the measurement of quality in healthcare,” *Soc. Sci. Med.*, vol. 99, pp. 27–34, 2013, doi: https://doi.org/10.1016/j.socscimed.2013.09.027.

[80] C. Kingsley and S. Patel, “Patient-reported outcome measures and patient-reported

experience measures," *BJA Educ.*, vol. 17, p. 13W_144, 2017.

[81] A. L and E. L, "Mechanics of quality assurance - now and in the future.," *Best Pract. Res. Clin. Gastroenterol.*, vol. 25, no. 3, pp. 419–425, 2011, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21764009/.

[82] National Health Service, "NHS Patient Experience Framework," 2011. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf (accessed Jan. 07, 2021).

[83] SL LaVela and A Gallan, "Evaluation and measurement of patient experience," *Patient Exp. J.*, vol. 1, pp. 28--36, 2014.

[84] M. D. Fottler, R. C. Ford, and S. A. Bach, "Measuring patient satisfaction in healthcare organizations: qualitative and quantitative approaches.," *Best Pract. benchmarking Healthc. a Pract. J. Clin. Manag. Appl.*, vol. 2, no. 6, pp. 227–239, 1997.

[85] A. Tong, P. Sainsbury, and J. Craig, "Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups," *Int. J. Qual. Heal. Care*, vol. 19, no. 6, pp. 349–357, Dec. 2007, doi: 10.1093/intqhc/mzm042.

[86] J. MM, H. M, P. AP, and P. M, "Healthcare professionals' proposed eHealth needs in elective primary fast-track hip and knee arthroplasty journey: A qualitative interview study.," *J. Clin. Nurs.*, vol. 28, no. 23, pp. 4434–4446, 2019, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31408555/.

[87] A. Rushton *et al.*, "Patient journey following lumbar spinal fusion surgery (LSFS): protocol for a multicentre qualitative analysis of the patient rehabilitation experience (FuJourn)," *BMJ Open*, vol. 8, no. 1, p. e020710, Jan. 2018, doi: 10.1136/bmjopen-2017-020710.

[88] M. Ali *et al.*, “Qualitative Analysis of Factors Influencing Patient Persistence and Adherence to Prescribed Overactive Bladder Medication in UK Primary Care,” *Adv. Ther.*, vol. 36, no. 11, pp. 3110–3122, 2019, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L2003483962&from=export.

[89] M. Alkandari, K. Ryan, and A. Hollywood, “The experiences of people living with peripheral neuropathy in Kuwait-A process map of the patient journey,” *Pharm.*, vol. 7, no. 3, 2019, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L629393012&from=export.

[90] O. Fennelly *et al.*, “Advanced musculoskeletal physiotherapy practice: The patient journey and experience,” *Musculoskelet. Sci. Pract.*, vol. 45, p. 102077, 2020, doi: https://doi.org/10.1016/j.msksp.2019.102077.

[91] A. Lagad, K. Hodgkinson, and T. R. O. Newton-John, “Is ignorance bliss, or is knowledge power? When cancer healthcare professionals become cancer patients,” *Eur. J. Cancer Care (Engl).*, vol. 28, no. 4, p. N.PAG-N.PAG, Jul. 2019, doi: 10.1111/ecc.13066.

[92] J. W. Creswell, “Editorial: Mapping the Field of Mixed Methods Research,” *J. Mix. Methods Res.*, vol. 3, no. 2, pp. 95–108, 2009, doi: 10.1177/1558689808330883.

[93] D. I. Ben-Tovim, M. L. Dougherty, T. J. O’Connell, and K. M. McGrath, “Patient journeys: the process of clinical redesign.,” *Med. J. Aust.*, vol. 188, no. 6, pp. S14-17, 2008, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L351686311&from=export.

[94] P. Bibby, "The management of chronic neck pain – A retrospective survey of the patient journey using in-depth semi-structured interviews," *J. Orthop. Nurs.*, vol. 10, no. 1, pp. 25–32, Feb. 2006, doi: 10.1016/j.joon.2006.01.001.

[95] T. M. Trebble, N. Hansi, T. Hydes, M. A. Smith, and M. Baker, "Process mapping the patient journey: an introduction," *BMJ*, vol. 341, 2010, doi: 10.1136/bmj.c4078.

[96] S. Campbell, B. Watson, A. Gibson, G. Husband, and K. Bremner, "Comprehensive service and practice development: city hospitals Sunderland's experience of patient journeys," *Pract. Dev. Heal. Care*, vol. 3, no. 1, pp. 15–26, Mar. 2004, doi: 10.1002/pdh.118.

[97] NHS Modernisation Agency NHS, "Improving flow: Improvement Leaders' Guide: Process and Systems Thinking," 2005, [Online]. Available: http://www.institute.nhs.uk/building_capability/building_improvement_capability/improvement_leaders'_guides%3A_process_and_systems_thinking.html.

[98] K. S, H. KE, D. SA, and F. SR, "The rosacea patient journey: a novel approach to conceptualizing patient experiences.," *Cutis*, vol. 95, no. 1, pp. 37–43, 2015, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25671443/.

[99] M. Sijm-Eeken, J. Zheng, and L. Peute, "Towards a Lean Process for Patient Journey Mapping–A Case Study in a Large Academic Setting," in *Digital Personalized Health and Medicine*, IOS Press, 2020, pp. 1071–1075.

[100] M. Westwood, Neil and James-Moore, Mike and Cooke, *Going Lean in the NHS*. NHS Institute for Innovation and Improvement, 2007.

[101] E. Holmes, J. K. Nicholson, J. Li, and A. W. Darzi, "Chapter 3 - Phenotyping the Patient Journey," E. Holmes, J. K. Nicholson, A. W. Darzi, and J. C. B. T.-M. P. in P. and P. H.

Lindon, Eds. Boston: Academic Press, 2016, pp. 49–74.

[102] J. Koski *et al.*, “An Analysis of Journey Mapping to Create a Palliative Care Pathway in a Canadian First Nations Community: Implications for Service Integration and Policy Development,” *Palliat. Care Res. Treat.*, vol. 10, p. 1178224217719441, Jan. 2017, doi: 10.1177/1178224217719441.

[103] W. Y. Cheung *et al.*, “Patient-reported treatment-related symptom burden for patients with advanced melanoma in Canada,” *Support. Care Cancer*, vol. 27, no. 1, pp. 219–227, Jan. 2019, doi: 10.1007/s00520-018-4316-9.

[104] Decision Resources Group, “Integrated Patient Journey: One journey, many lenses,” 2019. https://decisionresourcesgroup.com/solutions/the-patient-journey/ (accessed Jan. 07, 2021).

[105] N. Macdonald, A. Shapiro, C. Bender, M. Paolantonio, and J. Coombs, “Experiences and perspectives on the GIST patient journey,” *Patient Prefer. Adherence*, vol. 6, pp. 253–262, 2012, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L370080831&from=export.

[106] U. H. Graneheim and B. Lundman, “Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness,” *Nurse Educ. Today*, vol. 24, no. 2, pp. 105–112, 2004, doi: https://doi.org/10.1016/j.nedt.2003.10.001.

[107] A. Mottram, “‘Like a trip to McDonalds’: A grounded theory study of patient experiences of day surgery,” *Int. J. Nurs. Stud.*, vol. 48, no. 2, pp. 165–174, 2011, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002074891000218X.

[108] A. L. Glaser, Barney G and Strauss, *Discovery of grounded theory: Strategies for*

qualitative research. Routledge, 2017.

[109] B. Glaser, *Theoretical Sensitivity*. Mill Valley, CA: Sociology Press, 1978.

[110] M. Smith, Jonathan A and Flowers, Paul and Larkin, “Interpretative phenomenological analysis: Theory, method and research,” *London Sage*, 2009.

[111] P. F. Colaizzi, “Psychological research as the phenomenologist views it,” in *Existential-Phenomenological Alternatives for Psychology*, R. S. Valle and M. King, Eds. Oxford University Press, 1978.

[112] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.

[113] M. E. Kiger and L. Varpio, “Thematic analysis of qualitative data: AMEE Guide No. 131,” *Med. Teach.*, vol. 42, no. 8, pp. 846–854, Aug. 2020, doi: 10.1080/0142159X.2020.1755030.

[114] S. McCarthy, P. O’Raghallaigh, S. Woodworth, Y. L. Lim, L. C. Kenny, and F. Adam, “An integrated patient journey mapping tool for embedding quality in healthcare service reform,” *J. Decis. Syst.*, vol. 25, no. sup1, pp. 354–368, Jun. 2016, doi: 10.1080/12460125.2016.1187394.

[115] D. I. Ben-Tovim, M. L. Dougherty, T. J. O’Connell, and K. M. McGrath, “Patient journeys: the process of clinical redesign,” *Med. J. Aust.*, vol. 188, no. S6, pp. S14–S17, 2008.

[116] M. Hammer, “Reengineering work: don’t automate, obliterate,” *Harv. Bus. Rev.*, vol. 68, no. 4, pp. 104–112, 1990.

[117] M. P. Grocott, “Pathway redesign: putting patients ahead of professionals,” *Clin. Med. (Northfield. Il).*, vol. 19, no. 6, pp. 468–472, Nov. 2019, doi: 10.7861/clinmed.2019-0292.

[118] P. Bate and G. Robert, “Experience-based design: from redesigning the system around the patient to co-designing services with the patient.,” *Qual. Saf. Health Care*, vol. 15, no. 5, pp. 307–310, Oct. 2006, doi: 10.1136/qshc.2005.016527.

[119] M. Erlhoff and T. Marshall, *Design dictionary: perspectives on design terminology*. Walter de Gruyter, 2007.

[120] S. M *et al.*, “Using a service design model to develop the ‘Passport to Safer Birth’ in Nigeria and Uganda.,” *Int. J. Gynaecol. Obstet.*, vol. 139, pp. 56–66, 2017, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29218714/.

[121] K. Vanhaecht, K. De Witte, M. Panella, and W. Sermeus, “Do pathways lead to better organized care processes?,” *J. Eval. Clin. Pract.*, vol. 15, no. 5, pp. 782–788, Oct. 2009, doi: https://doi.org/10.1111/j.1365-2753.2008.01068.x.

[122] M. E. Tinetti, T. R. Fried, and C. M. Boyd, “Designing Health Care for the Most Common Chronic Condition—Multimorbidity,” *JAMA*, vol. 307, no. 23, pp. 2493–2494, Jun. 2012, doi: 10.1001/jama.2012.5265.

[123] S. Auras and M. Geraedts, “Patient experience data in practice accreditation—an international comparison,” *Int. J. Qual. Heal. Care*, vol. 22, no. 2, pp. 132–139, Apr. 2010, doi: 10.1093/intqhc/mzq006.

[124] C. for M. and M. (CMS), “HCAHPS: Patients’ Perspectives of Care Survey,” 2020. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS.

[125] J. S. Swan, P. V Pandharipande, and G. M. Salazar, “Developing a Patient-Centered Radiology Process Model,” *J. Am. Coll. Radiol.*, vol. 13, no. 5, pp. 510–516, 2016, [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S1546144015012806.

[126] O. E. Firbank, “Connecting the voices of users, caregivers and providers on service quality,” *Int. J. Health Care Qual. Assur.*, vol. 25, no. 5, pp. 403–420, Jan. 2012, doi: 10.1108/09526861211235900.

[127] R. W. Trickett, E. Mudge, P. Price, and I. Pallister, “A qualitative approach to recovery after open tibial fracture: The road to a novel, patient-derived recovery scale,” *Injury*, vol. 43, no. 7, pp. 1071–1078, 2012, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020138312000472.

[128] L. Sanders, M. Donovan-Hall, A. Borthwick, and C. J. Bowen, “Experiences of mobility for people living with rheumatoid arthritis who are receiving biologic drug therapy: implications for podiatry services,” *J Foot Ankle Res*, vol. 10, p. 14, 2017, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L616536181&from=export.

[129] B. Greenhalgh, T.; Hurwitz, “Why study narrative?,” *Br. Med. J.*, no. 318, pp. 48--51, 1999.

[130] L. M. Cox and L. S. Logio, “Patient safety stories: a project utilizing narratives in resident training.,” *Acad. Med.*, vol. 86, no. 11, pp. 1473–1478, Nov. 2011, doi: 10.1097/ACM.0b013e318230efaa.

[131] D. Bessarab and B. Ng’andu, “Yarning about yarning as a legitimate method in Indigenous research,” *Int. J. Crit. Indig. Stud.*, vol. 3, no. 1, pp. 37–50, 2010.

[132] D. Bertsimas *et al.*, “Algorithmic prediction of health-care costs,” *Oper. Res.*, vol. 56, no. 6, pp. 1382–1392, 2008.

[133] J. ALLSOP, “Two Sides to Every Story: Complainants’ and Doctors’ Perspectives in Disputes about Medical Care in a General Practice Setting,” *Law \& Policy*, vol. 16, no. 2, pp. 149–183, 1994, doi: https://doi.org/10.1111/j.1467-9930.1994.tb00121.x.

[134] M. Martin, R. Champion, L. Kinsman, and K. Masman, “Mapping patient flow in a regional Australian emergency department: A model driven approach,” *Int. Emerg. Nurs.*, vol. 19, no. 2, pp. 75–85, Apr. 2011, doi: 10.1016/j.ienj.2010.03.003.

[135] J. Shaw, S. Pemberton, C. Pratt, and L. Salter, “Shadowing: a central component of patient and family-centred care,” *Nurs. Manag.*, vol. 21, no. 3, p. 20, 2014.

[136] H. Van der Meide, C. Leget, and G. Olthuis, “Giving voice to vulnerable people: the value of shadowing for phenomenological healthcare research,” *Med. Heal. Care Philos.*, vol. 16, no. 4, pp. 731–737, 2013.

[137] S. Reeves, A. Kuper, and B. D. Hodges, “Qualitative research methodologies: ethnography,” *BMJ*, vol. 337, p. a1020, Aug. 2008, doi: 10.1136/bmj.a1020.

[138] M. Hammersley and P. Atkinson, *Ethnography: Principles in practice*. Routledge, 2019.

[139] D. F. Hamilton *et al.*, “What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement,” *BMJ Open*, vol. 3, no. 4, p. e002525, Jan. 2013, doi: 10.1136/bmjopen-2012-002525.

[140] P. K. Hodgkin, “Web based patient feedback,” *Br. Med. J. Publ. Gr.*, vol. 338, 2009.

[141] J. H. Wasson, T. A. MacKenzie, and M. Hall, “Patients use an internet technology to report when things go wrong,” *Qual. Saf. Heal. Care*, vol. 16, no. 3, pp. 213 LP – 215, Jun. 2007, doi: 10.1136/qshc.2006.019810.

[142] M. P. Manary, W. Boulding, R. Staelin, and S. W. Glickman, “The patient experience and health outcomes.,” *N. Engl. J. Med.*, vol. 368, no. 3, pp. 201–203, Jan. 2013, doi:

10.1056/NEJMp1211775.

[143] O. A. Bjertnaes, “The association between survey timing and patient-reported experiences with hospitals: results of a national postal survey,” *BMC Med. Res. Methodol.*, vol. 12, no. 1, p. 13, 2012, doi: 10.1186/1471-2288-12-13.

[144] J. Pickles, E. Hide, and L. Maher, “Experience based design: a practical method of working with patients to redesign services,” *Clin. Gov. An Int. J.*, vol. 13, pp. 51–58, 2008.

[145] M. B. Buntin, M. F. Burke, M. C. Hoaglin, and D. Blumenthal, “The benefits of health information technology: a review of the recent literature shows predominantly positive results.,” *Health Aff. (Millwood).*, vol. 30, no. 3, pp. 464–471, Mar. 2011, doi: 10.1377/hlthaff.2011.0178.

[146] B. M. Caulfield and S. C. Donnelly, “What is Connected Health and why will it change your practice?,” *QJM*, vol. 106, no. 8, pp. 703–707, Aug. 2013, doi: 10.1093/qjmed/hct114.

[147] S. C.-I. Chen, C. Liu, and R. Hu, “Fad or Trend? Rethinking the Sustainability of Connected Health,” *Sustainability*, vol. 12, no. 5, 2020, doi: 10.3390/su12051775.

[148] B. Kaplan and K. D. Harris-Salamone, “Health IT success and failure: recommendations from literature and an AMIA workshop.,” *J. Am. Med. Inform. Assoc.*, vol. 16, no. 3, pp. 291–299, 2009, doi: 10.1197/jamia.M2997.

[149] M. Martin, R. Champion, L. Kinsman, and K. Masman, “Mapping patient flow in a regional Australian emergency department: A model driven approach,” *Int. Emerg. Nurs.*, vol. 19, no. 2, pp. 75–85, Apr. 2011, doi: 10.1016/j.ienj.2010.03.003.

[150] J. I. Westbrook, E. W. Coiera, and J. Braithwaite, “Measuring the Impact of Online

Evidence Retrieval Systems using Critical Incidents & Journey Mapping," *Stud Heal. Technol Inf.*, vol. 116, pp. 533–538, 2005, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L610095341&from=export.

[151] B. Chaudhry *et al.*, "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care.," *Ann. Intern. Med.*, vol. 144, no. 10, pp. 742–752, May 2006, doi: 10.7326/0003-4819-144-10-200605160-00125.

[152] S. Audet, Anne-Marie and Doty, Michelle M and Peugh, Jordan and Shamasdin, Jamil and Zapert, Kinga and Schoenbaum, "Information technologies: when will they make it into physicians' black bags?," *Medscape Gen. Med.*, vol. 6, no. 4, 2004.

[153] S. K. Bell *et al.*, "A patient feedback reporting tool for OpenNotes: implications for patient-clinician safety and quality partnerships," *BMJ Qual. \& Saf.*, vol. 26, no. 4, pp. 312–322, 2017, doi: 10.1136/bmjqs-2016-006020.

[154] C. McGregor, J. Percival, J. Curry, D. Foster, E. Anstey, and D. Churchill, "A structured approach to requirements gathering creation using PaJMa models.," *Conf Proc IEEE Eng Med Biol Soc*, pp. 1506–1509, 2008, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L355916591&from=export.

[155] C. MacDougall, Candice and Percival, Jennifer and McGregor, "Integrating health information technology into clinical guidelines," *2009 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4646--4649, 2009.

[156] J. Percival and C. McGregor, "An Evaluation of Understandability of Patient Journey Models in Mental Health.," *JMIR Hum. factors*, vol. 3, no. 2, p. e20, Jul. 2016, doi:

10.2196/humanfactors.5640.

[157] C. M. Martin *et al.*, "Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR)," *J. Eval. Clin. Pract.*, vol. 18, no. 6, pp. 1226–1234, Dec. 2012, doi: 10.1111/j.1365-2753.2012.01880.x.

[158] M. E. Jordan *et al.*, "The role of conversation in health care interventions: enabling sensemaking and learning," *Implement. Sci.*, vol. 4, no. 1, p. 15, 2009, doi: 10.1186/1748-5908-4-15.

[159] J. Cheshire and S. Ziebland, "Narrative as a resource in accounts of the experience of illness," *Socioling. Narrat.*, vol. 6, 2005.

[160] C. M. Martin, J. P. Sturmberg, K. Stockman, N. Hinkley, and D. Campbell, "Anticipatory Care in Potentially Preventable Hospitalizations: Making Data Sense of Complex Health Journeys ," *Frontiers in Public Health* , vol. 6. p. 376, 2019, [Online]. Available: https://www.frontiersin.org/article/10.3389/fpubh.2018.00376.

[161] D. Rolls *et al.*, "Before-after evaluation of patient length of stay in a rehabilitation context following implementation of an electronic patient journey board," *Int. J. Med. Inform.*, vol. 134, p. 104042, 2020, doi: https://doi.org/10.1016/j.ijmedinf.2019.104042.

[162] A. M. Bisantz *et al.*, "Emergency Department Status Boards: A Case Study in Information Systems Transition," *J. Cogn. Eng. Decis. Mak.*, vol. 4, no. 1, pp. 39–68, Mar. 2010, doi: 10.1518/155534310X495582.

[163] J. Rayner, T. Khan, C. Chan, and C. Wu, "Illustrating the patient journey through the care continuum: Leveraging structured primary care electronic medical record (EMR) data in Ontario, Canada using chronic obstructive pulmonary disease as a case study," *Int. J. Med. Inform.*, vol. 140, p. 104159, Aug. 2020, doi: 10.1016/j.ijmedinf.2020.104159.

[164] A. Singer, A. L. Kroeker, S. Yakubovich, R. Duarte, B. Dufault, and A. Katz, “Data quality in electronic medical records in Manitoba,” *Can. Fam. Physician*, vol. 63, no. 5, pp. 382 LP – 389, May 2017, [Online]. Available: http://www.cfp.ca/content/63/5/382.abstract.

[165] M. N. Levine, G. Alexander, A. Sathiyapalan, A. Agrawal, and G. Pond, “Learning Health System for Breast Cancer: Pilot Project Experience,” *JCO Clin. Cancer Informatics*, no. 3, pp. 1–11, Aug. 2019, doi: 10.1200/CCI.19.00032.

[166] A. D. Hilt *et al.*, “Perspectives of Patients and Professionals on Information and Education After Myocardial Infarction With Insight for Mixed Reality Implementation: Cross-Sectional Interview Study,” *JMIR Hum Factors*, vol. 7, no. 2, p. e17147, 2020, doi: 10.2196/17147.

[167] G. Eysenbach, “What is e-health?,” *J Med Internet Res*, vol. 3, no. 2, p. e20, 2001, doi: 10.2196/jmir.3.2.e20.

[168] H. Ratanjee-Vanmali, W. Swanepoel, and A. Laplante-Lévesque, “Patient Uptake, Experience, and Satisfaction Using Web-Based and Face-to-Face Hearing Health Services: Process Evaluation Study,” *J. Med. Internet Res.*, vol. 22, no. 3, p. e15875, 2020, [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L631296721&from=export.

[169] V. Le, E. A. Wagar, R. A. Phipps, R. E. Del Guidice, H. Le, and L. P. Middleton, “Improving Patient Experience of Wait Times and Courtesy Through Electronic Sign-in and Notification in the Phlebotomy Clinic,” *Arch. Pathol. Lab. Med.*, vol. 144, no. 6, pp. 769–775, Nov. 2019, doi: 10.5858/arpa.2019-0139-OA.

[170] F. Pascual, "Text Analytics Basics: A Beginner's Guide," 2019. https://monkeylearn.com/blog/what-is-text-analytics/ (accessed Jan. 07, 2021).

[171] IBM Cloud Education, "Natural Language Processing (NLP)," 2020. https://www.ibm.com/cloud/learn/natural-language-processing (accessed Jan. 07, 2021).

[172] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: https://doi.org/10.1016/j.asej.2014.04.011.

[173] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[174] L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," *Knowledge-Based Syst.*, vol. 41, pp. 89–97, 2013, doi: https://doi.org/10.1016/j.knosys.2013.01.001.

[175] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decis. Support Syst.*, vol. 53, no. 4, pp. 680–688, 2012, doi: https://doi.org/10.1016/j.dss.2012.05.025.

[176] F. Pascual, "Topic Modeling: An Introduction," 2019. https://monkeylearn.com/blog/introduction-to-topic-modeling/ (accessed Jan. 07, 2021).

[177] J. Cao, T. Xia, J. Li, Y. Zhang, and S. Tang, "A density-based method for adaptive LDA model selection," *Neurocomputing*, vol. 72, no. 7, pp. 1775–1781, 2009, doi: https://doi.org/10.1016/j.neucom.2008.06.011.

[178] Statista, "Number of social network users worldwide from 2017 to 2025," 2021. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

(accessed Jan. 07, 2021).

[179] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015, doi: https://doi.org/10.1016/j.knosys.2015.06.015.

[180] D. King, D. Ramirez-Cano, F. Greaves, I. Vlaev, S. Beales, and A. Darzi, "Twitter and the health reforms in the English National Health Service.," *Health Policy*, vol. 110, no. 2–3, pp. 291–297, May 2013, doi: 10.1016/j.healthpol.2013.02.005.

[181] Twitter, "How to Tweet." https://help.twitter.com/en/using-twitter/how-to-tweet (accessed Feb. 07, 2021).

[182] Statista, "Growth of monthly active users of selected social media platforms worldwide from 2019 to 2021," 2021. https://www.statista.com/statistics/1219318/social-media-platforms-growth-of-mau-worldwide/ (accessed Jan. 07, 2021).

[183] Twitter, "Developer Portal." https://developer.twitter.com/en/docs/developer-portal/overview (accessed Jan. 07, 2021).

[184] J. Widman, "What is Reddit?," 2021. https://www.digitaltrends.com/web/what-is-reddit/ (accessed Jan. 07, 2021).

[185] World Health Organization, "Breast cancer now most common form of cancer: WHO taking action," 2021. https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action (accessed Feb. 07, 2021).

[186] American Cancer Society, "How Common Is Breast Cancer?," 2021. https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html (accessed Feb. 07, 2021).

[187] NumPy Community, "What is NumPy?," 2021.

https://numpy.org/doc/stable/user/whatisnumpy.html (accessed Feb. 07, 2021).

[188] “What Is Pandas In Python? Everything You Need To Know.” https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/ (accessed Feb. 07, 2021).

[189] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Comput. Sci. \& Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.

[190] M. Waskom, “seaborn: statistical data visualization,” 2021. https://seaborn.pydata.org/ (accessed Feb. 07, 2021).

[191] Python Software Foundation, “The Python Standard Library,” 2021. https://docs.python.org/3/library/itertools.html (accessed Feb. 07, 2021).

[192] scikit-learn community, “scikit-learn: Machine Learning in Python,” 2021. https://scikit-learn.org/stable/ (accessed Feb. 07, 2021).

[193] NetworkX Developers, “NetworkX: Network Analysis in Python,” 2021. https://networkx.org/ (accessed Jul. 07, 2021).

[194] Python Community, “pyLDAvis 3.3.1,” 2021. https://pypi.org/project/pyLDAvis/ (accessed Jul. 07, 2021).

[195] S. Jain, “Natural Language Processing for Beginners: Using TextBlob,” 2018. https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/ (accessed Feb. 07, 2021).

[196] Boost Labs, “Word Clouds & the Value of Simple Visualizations,” 2014. https://boostlabs.com/blog/what-are-word-clouds-value-simple-visualizations/ (accessed Jun. 07, 2021).

[197] M. Beck, “How to Scrape Tweets From Twitter,” 2020.

https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1 (accessed Jan. 07, 2021).

[198] D. Suwankhong and P. Liamputtong, "Breast cancer treatment: experiences of changes and social stigma among Thai women in southern Thailand," *Cancer Nurs.*, vol. 39, no. 3, pp. 213–220, 2016.

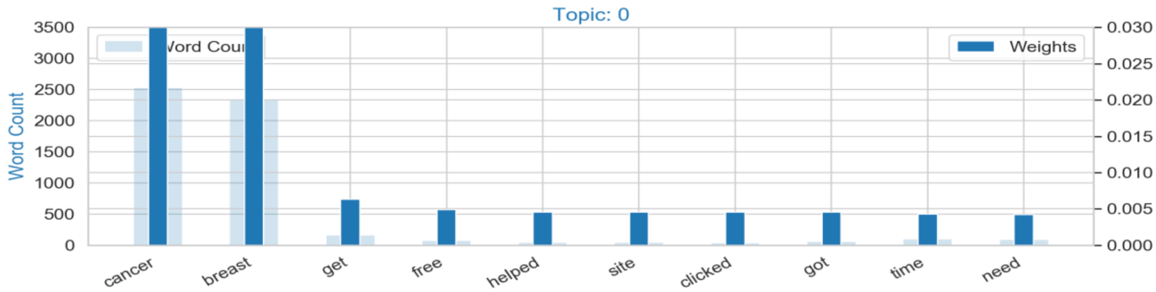
[199] E. Goffman, *Stigma: Notes on the management of spoiled identity*. Simon and Schuster, 2009.

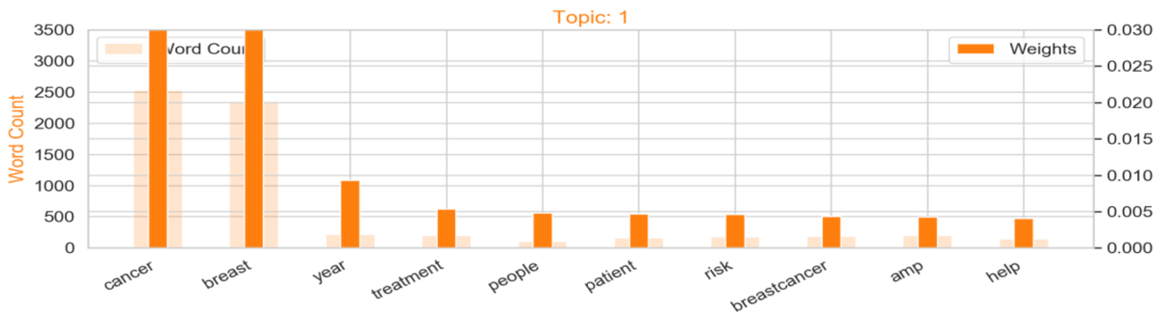
[200] C. E. Mosher and S. Danoff-Burg, "Death anxiety and cancer-related stigma: a terror management analysis," *Death Stud.*, vol. 31, no. 10, pp. 885–907, 2007.

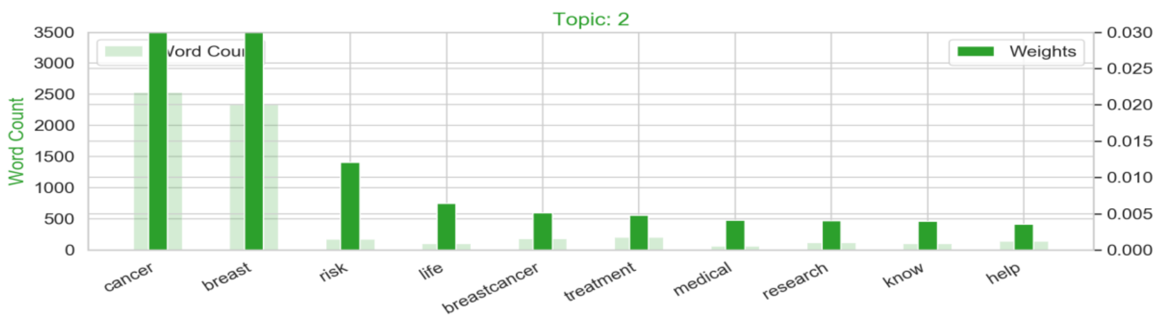
[201] L. A. V Marlow and J. Wardle, "Development of a scale to assess cancer stigma in the non-patient population," *BMC Cancer*, vol. 14, no. 1, p. 285, 2014, doi: 10.1186/1471-2407-14-285.

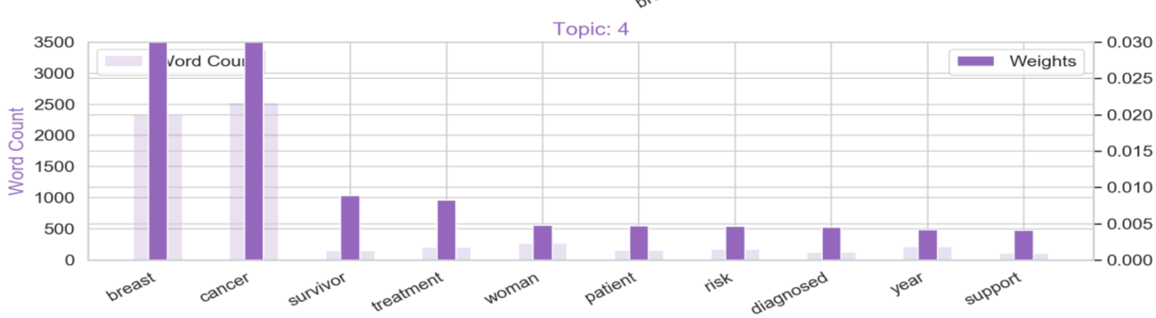
Appendix

Twitter’s 10 latent topics obtained from LDA.



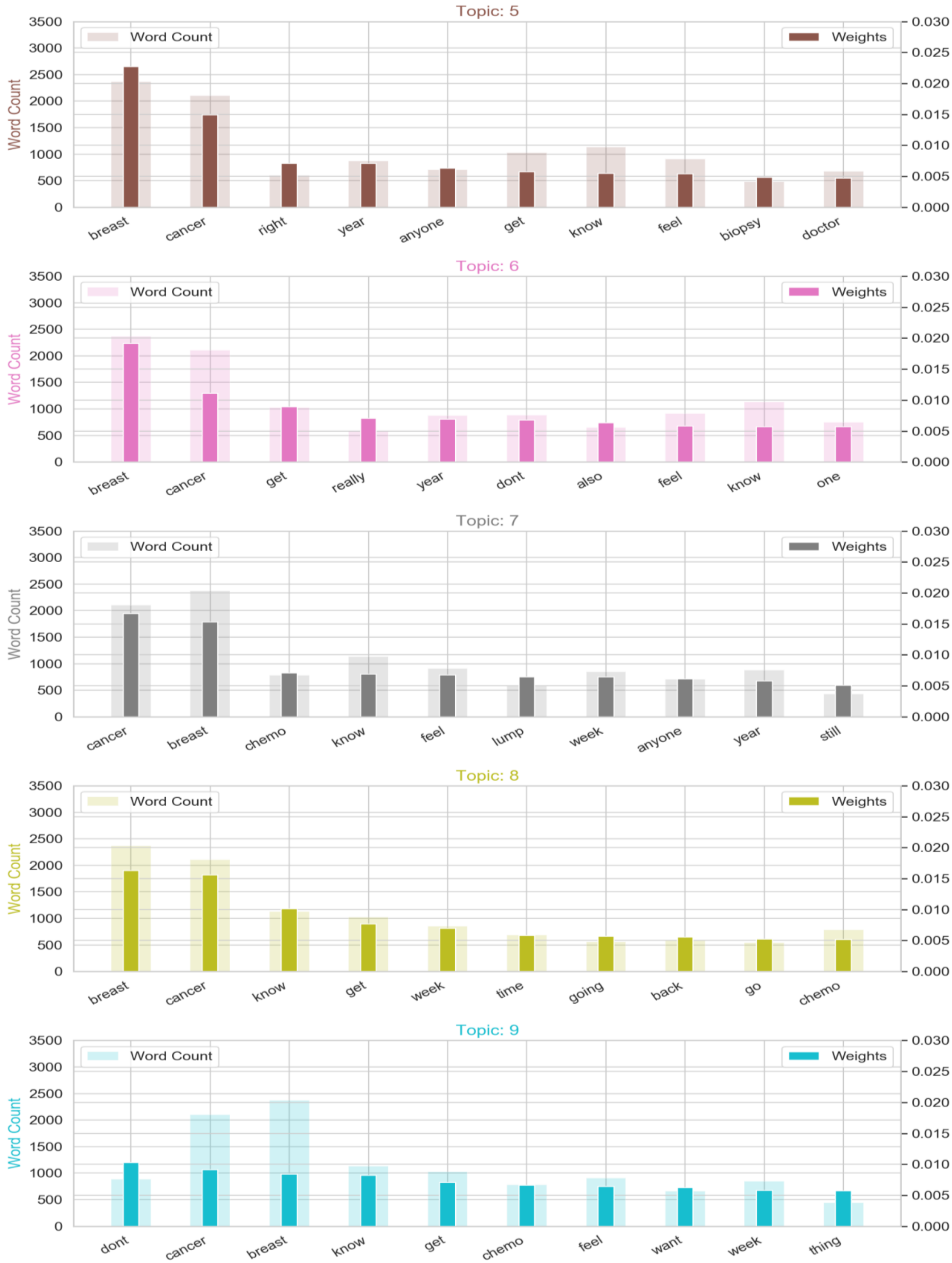







Topic: 5
Word Count
Weights
cancer
breast
amp
one
make
helped
metastatic
free
site
clicked
Topic: 6
Word Count
Weights
cancer
breast
woman
medical
day
could
may
year
covid
oncology
Topic: 7
Word Count
Weights
cancer
breast
amp
woman
research
treatment
patient
one
year
today
Topic: 8
Word Count
Weights
breast
cancer
year
get
woman
amp
men
new
diagnosed
time
Topic: 9
Word Count
Weights
cancer
breast
breastcancer
woman
screening
get
diagnosed
dont
survivor
metastatic

Reddit's 10 latent topics obtained from LDA.


Topic: 5
Word Count
Weights
breast
cancer
right
year
anyone
get
know
feel
biopsy
doctor
Topic: 6
Word Count
Weights
breast
cancer
get
really
year
dont
also
feel
know
one
Topic: 7
Word Count
Weights
cancer
breast
chemo
know
feel
lump
week
anyone
year
still
Topic: 8
Word Count
Weights
breast
cancer
know
get
week
time
going
back
go
chemo
Topic: 9
Word Count
Weights
dont
cancer
breast
know
get
chemo
feel
want
week
thing

Vita

Sofia V Ortega joined The University of Texas at El Paso (UTEP) in Fall 2015. She earned her Bachelor of Science in Industrial and Systems Engineering from UTEP in Spring 2020. She also earned her Master of Science in Manufacturing Engineering from UTEP in Summer 2021. Her research interests are user experiences, health care, sustainability, and environmental solutions.

From her freshman year, she has been part of several UTEP Edge activities as: on-campus employment for the Office of International Programs, faculty-led programs, study abroad experience in France, community engagement activities, research in sustainability and health care, student leadership in several student organizations and internships. In Summer 2019 she interned for Cummins Inc. as a Supplier Quality Improvement Engineer (SQIE) Intern and in Summer 2020 she interned at Microsoft Corporation as a Program Manager Intern. Sofia will return to Microsoft as a full-time employee in August 2021.

Contact Information: sofia87_b@hotmail.com