Exploring the impact of socio-cognitive factors on adherence to asthma medication using traditional mixed methods and machine learning

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy, December 2021.

Vanja Ljevar, MSc

4305204

Supervised by

Dr James Goulding
Dr Alexa Spence

Signature ____________________

Date ______ / ______ / ______

Abstract

Asthma self-management and regular use of medication are crucial in preventing asthma attacks, which can be fatal. However, existing research usually does not take into consideration the impact that perceptions about asthma, of both patients and the general public have on adherence to treatment. Both methodological and empirical contributions are made, combining the approaches from data science and psychology. The thesis proposes new forms of triangulation of alternative research methods; introduces a novel process of identifying perceptions in text, as well as the use of Grouped Model Class Reliance, which is capable of assessing the most predictive feature group while accounting for non-linear relationships. Empirical contributions are reflected in identification of the leading perceptions, that include stigmatization and its underlying mechanisms, as well as the perceived sense of community.

Using a mixed method approach, this work combines traditional Psychological quantitative and qualitative methods and predictive algorithms/machine learning techniques. Using interviews, exploratory qualitative research identified patients' internal perceptions (about asthma) and external perceptions (what they considered were others' perceptions). Qualitative analysis of Twitter content formed the second part of this study, identifying several themes in perceptions expressed in tweets. Convergence analysis revealed mutual topics from tweets and interviews: self-pity, humour, disclosure, lack of understanding - topics that reflect stigmati-

zation; attachment to inhaler and perceived sense of community. However, the presence of negative humour and self-pity was much more prominent on Twitter than in interviews, signifying that some perceptions are more freely expressed on social media such as Twitter, than in the laboratory setting. Conversely, interviews provide more context for stigmatization, though examples. Having recognised the value of Twitter as a naturalistic setting for observing perception, this work created a novel procedure for analysing perceptions in 'big data' comprising of filtering, activation, evaluation and modality, that can be used in asthma non-related domains. The results indicate that perceptions related to stigma are the most prevalent negative perceptions about asthma held by both asthma patients and non-patients (following the Twitter analysis).

The next stage of research expanded on this initial conclusion by assessing the impact that negative perceptions have on adherence to medication by people with asthma. The third study measured the impact stigma-related factors have on adherence to asthma medication, concluding that denial was the strongest factor. Mediation analysis using coping mechanisms as mediators also highlighted the non-atomic nature of stigma, identifying different underlying mechanisms by which factors relating to stigmatisation of asthma impact patients' adherence to medication. However, this work also indicated there is potential information sharing and non-linear interactions occurring across factors. This led to the final study that mitigated the effects of non-linear relationships, using a first Grouped Model Class Reliance (Group-MCR) to compare and quantify the importance of several groups of factors in predicting adherence (including perceptions, demographics, lifestyle, coping, emotions, asthma and psychology traits). This final, fourth study was important in linking up the work of this thesis as it established that perceptions are not just important in predicting adherence - they are the strongest set of predictors of adherence when compared to other factors considered in the literature.

This thesis takes advantage of a mixed-method approach, highlighting the value of the exploratory nature of qualitative work that provided the context and enabled identification of relevant perceptions; the strength of traditional statistics in describing effects, which was evident in the mediation analysis that implied different stigma mechanisms; and the predictive power of machine learning when dealing with complex, non-linear relationships and large amounts of data. This work indicates that public health interventions should focus on patients' perceptions as an important component of treatment. In addition, the non-atomic and intrinsic nature of stigma identified within patients with asthma and the general public, underlines the importance of not only changing the negative perceptions of patients in the development of future interventions, but also engagement about asthma with the wider public, with the ultimate aim of reducing stigma.

List of publications

Ljevar, V., Goulding, J., Spence, A., & Smith, G. (2020, December). Perception detection using Twitter. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 4250-4256). IEEE.

Ljevar, V., Goulding, J., & Smith, G. (2020, December). Exploration of links between anxiety purchases, deprivation and personality traits. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3767-3774). IEEE.

Gaytan Camarillo, M., Ferguson, E., Ljevar, V., & Spence, A. (2021). Big changes start with small talk: Twitter and climate change in times of Coronavirus pandemic. Frontiers in Psychology, 12, 2308.

Ljevar, V., Goulding, J., Smith, G. & Spence, A. (2021, December). Using Model Class Reliance to measure group effect on adherence to asthma medication. In 2020 IEEE International Conference on Big Data (Big Data)

In review:

Ljevar, V., Nica-Avram, G., Harvey, J., & Goulding, J. (2021, September). Ill-fated interactions: modelling networked complaints on a food wastefighting platform

Presented at The Future of Food Symposium:

Ljevar, V., Nica-Avram, G., Harvey, J., Branco-Illodo, I., Gallage, S. and Goulding, J. Cocreating value from failed experiences in the sharing economy. Nottingham University Business School, University of Nottingham and Stirling Management School, University of Stirling.

In preparation:

Ljevar, V., Spence, A. & Goulding, J. Small talk, big data: Patients' and general public's perceptions about people with asthma.

Ljevar, V., Spence, A. & Goulding, J. Stigma mechanisms and their impact on adherence to asthma medication.

Ljevar, V., Goulding, J. & Skatova, A. Menstrual pain and period poverty disparities in England.

Acknowledgements

This PhD has been an incredible and insightful experience. I would like to thank:

My supervisors - brilliant minds that acted as teachers, resources, critics, contributors and inspiration. James, your lateral thinking has inspired me as an academic and as a business person. Thank you for recognising years ago that I would do work that I never ever dreamed I could. You have been a perspicacious mentor and your incredible energy and enthusiasm are contagious. Thank you for valuing my good ideas and blatantly rejecting rubbish ones. Lastly, thank you for making N/Lab more than a research group - a family. The journey has only started. Alexa, if there is one thing I would like to take from this journey it would be our discussions. I am perpetually wowed by your dept of thinking and dedication to push the envelope of novel research techniques, as well as your own. You are a charismatic leader who created space for myself and others to grow while providing the never-failing support. I will forever cherish your encouraging comments that I always reflect on when I need a reminder. I have learned so much from you and I hope our collaboration does not stop here.

Gavin, my fiance - I am incredibly lucky to have you in my life and simply said - there are no words to explain my gratitude for your support that was there every step of the way. You are my rock, my inspiration and my best cheerleader. I am looking forward to witnessing all the things you (and us together) will accomplish in the future.

My N/Lab group - Maddy, Georgie, Gregor, Rosa, Bertrand, John, Lizzie and Roberto: Thank you for everything, I consider you to be my second family. To my friends: Anja, Dragan, Victoria, Sarah, Natalie, Shazmin, Velvet, Mihajlo, Lucy, Cara - thank you for always having an extra kind word for me and a spare advice in your sleeve throughout this era, before and after it.

My parents, Vlado and Marina. Najveće hvala mom tati koji je uvijek vjerovao u mene i omogućio mojim idejama da rastu i van granica naše male države. Tata, ti si moj idol i hvala ti za inspiraciju, podršku i uvijek zanimljive razgovore. Mama,

hvala ti što si uvijek bila moja mirna luka, imala vremena da saslušaš i moju najmanju brigu i uvijek uputiš podršku i savjet. Ova doktorska teza je bila moguća zbog vas.

I dedicate this thesis to Dominic Reedman-Flint, my friend and companion on our CDT journey. You will stay forever in our memories and our hearts.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Theoretical Motivation

Asthma is a prevalent health problem, globally affecting around 300 million people of all ages, ethnic groups and geographic origins [43]. Substantial research has been examined on asthma incidence finding it has a significant geographical disparity and that traditionally assessed risk factors (e.g. air pollutants, environmental and in utero tobacco smoke, viral infections, indoor allergens etc.) do not fully explain these trends [205]. Despite the research efforts, we still have uncontrolled, unpredictable, and untreated asthma. Today, the main emphasis in asthma medication care has shifted from the relief of symptoms via symptom relief inhalers, to the prevention of asthma symptoms, using anti-inflammatory steroid inhalers, also called *preventers* [137]. This is mostly due to the fact that the over-reliance on use of beta antagonists from reliever inhalers is potentially harmful [50].

However, the effectiveness of medical regimens are dependent on patients' willingness to adhere to asthma medication [217]. In order to effectively participate in their asthma management, patients are required to understand what kind of medication they need, understand its purpose and follow through with their asthma treatment [229]. Moreover, there is a drive from the medical field to create more pro-active patients, since the regular use of medication would prevent many negative outcomes. However, despite being a highly treatable condition, asthma mortality rates remain unacceptably high, with *lack of adherence* to medications cited as *a major cause.* It is well documented in clinical practice that adherence to asthma treatment ranges from 38% to 50% [20, 21, 45].

Non-adherence to medication leads to uncontrolled asthma symptoms that could ultimately affect patients' quality of life and may ever result in deadly asthma attacks. In fact, it is stated that in the UK alone, an average of 3 people die from an asthma attack each day (Asthma UK). There is also a large proportion of healthcare spending due to non-adherence [300, 135, 10]. Understanding what makes people not adhere to their medication properly is one of the main tasks of this thesis. There are further methodological, and theoretical motivators for this research especially in relation to adherence and asthma perceptions, which is a complex topic to investigate. These are described in this section, along with the chosen research paradigm and methodological opportunities that correspond to research objectives.

- Non-adherence to medication is a complex problem. While various drivers of non-adherence have been considered in isolation, interactions between demographic, behavioural and situational factors have never been modelled in concert, which would uncover their interactions and highlight the most relevant group - that would direct where interventions' focus should be.

- Insufficient attention has been paid to negative perceptions such as stigma and perceived judgements of peers. However, if we aim to understand reasons for non-adherence, socio-cognitive factors, such as perceptions and beliefs should be taken into consideration, because patients have to deal with a whole set of situational, psychological and social factors that shape the way they perceive their condition [154]. When patients get diagnosed with a serious condition such as asthma, they create perceptions about the condition, treatment and themselves, based on their knowledge and experience [136]. As patients try to incorporate the identity of someone who has asthma, perceptions they create can change their experience, modify the meaning of it and increase or mitigate the stress related to asthma management [137]. How people frame their condition to themselves also relates to whether they subsequently engage in positive coping mechanisms, such as information seeking or maladaptive coping mechanisms, such as hiding or ignoring asthma, where both types of coping can impact their decision to adhere to their treatment [295].

- Previous research often does not take into consideration the context of patients' social surroundings that also affects perceptions and ultimately, adherence. While some studies focus on the impact of demographics, personality traits, or even the health polices, few studies have investigated the impact that patient-related socio-cognitive factors have on adherence. This includes perceptions other people have about asthma and asthma patients, which may inform patients' perceptions and behaviours. Patients' social surrounding is important not only in terms of the support they can provide, but also in terms of the potential negativity and stigmatization that would make them feel embarrassed and reluctant to use medication in public (or at all).

This PhD, therefore, emphasizes that perceptions that inform a patient's identity are 'complex, recursive, reflexive and constantly under construction' - as products of interchanges among people [319]. If we were, however, able to demonstrate the impact perceptions have on adherence and that both other people's and patients' view are relevant to one's treatment strategy, this would highlight the need for an upgraded paradigm in the medical field: patients should not be the sole culprits of non-adherence and subsequently, targets of interventions; society should accept some responsibility, as well.

1.2 Philosophical stance

This thesis sits at the intersection between psychology and data science while each have their own respective assumptions, especially with psychology mostly being interested in explanation, while data science is usually focused on prediction. The thesis also consists of two distinct parts: the first is focused on exploring which perceptions exist in the domain of adherence to medication regimes; how they can be understood and categorized. The second part examines the extent and relevance of the impact those perceptual categories, such as stigma, have on adherence to medication. Based on this research topic, the choice of the research paradigm is important, because it directs the intent, motivation and expectations of the research [189]. Perceptions are particularly fuzzy, complex, not easily captured or measured. This is why critical realism, as a research paradigm, is appropriate for this work. Critical realism enables capturing the complexities of the topic by untangling the underlying mechanisms and it also allows for measuring the impact on behaviour. This position of critical realism is also taken with the goal to drill down into the topic, both qualitatively and quantitatively [189]. Specifically, mixed methods are tightly assigned to this approach that requires iterations of theory using various methods, which has been recognised as a key element of critical realism, increasing the ability to symbiotically obtain new knowledge [12, 189]. This, therefore, highlights a mixed-methods approach embedded in this work, that aims to offer both methodological and empirical contributions.

1.3 Research paradigm

Research paradigms can be considered to be the philosophies of science and they consist of: ontology, epistemology, axiology, methodology and rigor [223]. In the field of research of asthma perceptions and asthma adherence, several paradigms were recognised, however, the most vivid separation exists between the extremes: positivist and interpretivist approach. Positivism is based on rationalistic, empiricistic philosophy and aims to observe and measure, in order to predict [189]. Positivists assume that a single, objective reality exists, regardless of individual perceptions [140]. This means that positivists tend to hypothesise and evaluate causal inferences with the focus on validity, explanatory power and parsimony [256]. This kind of work is devoted to quantified information and empirically observable phenomena. Even though it is not limited to specific methods, it most commonly aligns with a quantitative approach [256]. This research takes a positivist ideas to the extent, as it tests and measures relationships that exist between perceptions and adherence to medication, creates predictive models to detect perceptions and to quantify the prevalence of negative perceptions.

Conversely, interpretivism (or often called, constructivism), strongly rejects positivism [153]. This approach is based on the philosophy of hermeneutics, has the goal of understanding the human experience and relies on the participants' views and perceptions [189]. Interpretivists advocate that time and context-free generalisations are not possible and not even desirable, since the observer is not separate from observed entities [153]. Specifically, this paradigm posits that interpretative analysis of subjective meaning cannot and should not be tested across other cases, due to its unique tie to a particular system [161]. Contrary to positivism that tries to come close to the objective truth, interpretivism posits that multiple subjective truths exist. However, this is often criticised as there are no clear standards that determine which subjective interpretation is judged to be better than another [94]. Considering the differences between positivism and interpretivism and their fundamental assumptions, it is understandable why their goals are different as well: while positivists focus on prediction, generality and replicability, the interpretivists aim to obtain interpretation, specificity and self-validation [161]. In the context of this thesis, interpretivist view on reality is utilized when qualitative research was used to explore and identify which perceptions about asthma exist within asthma patients' group and on the publicly available tweets.

Purists from both sides of the spectrum see the two opposing paradigms as incompatible and impossible to mix. However, the mixed method approach eventually emerged and was suggested as a third paradigm [84]. Mixed methods are introduced not to resolve the debate between qualitative and quantitative approach, but to legitimate the value of using multiple approaches, by being inclusive and complementary, rather than limiting the choice to a single method [153]. For this reason, the mixed method approach was chosen in this thesis - in order to examine the research topic from a multi-layered perspective and obtain a more complete observation of perceptions and their impact on adherence. Even though mixed methods can be used with any paradigm, this approach is particularly supported by Post-Positivism and Pragmatism, the remaining two philosophical approaches that are, on the research continuum, positioned between positivism and interpretivism [189].

Pragmatism is not committed to a single philosophy [189]. In fact, pragmatism rejects traditional dualism, 'either' 'or' choices, and recognises that both natural and social worlds exist [153]. Pragmatically inclined researchers take the value-oriented approach to research and prefer action to philosophising [153]. Specifically, the focus of pragmatism is to provide the response to the question, without having the loyalty to any particular method. This means that there are no strict assumptions for the researcher apart from the agreement that scientific research occurs in many contexts, as we are historically and socially conditioned [59]. Pragmatists also imply that while reading the world we can never be sure if we are, indeed, reading the world or ourselves [59]. However, after reviewing interpretivism, positivism and pragmatism, *post-positivism* was determined as the best suited philosophy of science behind this research.

The ontology of Post-Positivism - called Critical Realism, asserts the existence of the external reality that can be imperfectly discovered [287]. Post-positivism is often seen as a middle way between positivism and interpretivism and it bridges the gap between these extremes by providing a more dynamic mode of enquiry [320]. Critical realism recognises that our world view is always mediated by our perceptual lenses, which supports the validity of the nature of perceptions as a research topic [203]. Critical realism is suitable for this work, as while it assumes that the world is driven by generalizable laws, it is also made of thinking, breathing human beings, events, objects and social phenomena that are context dependent [94]. These assumptions, therefore, support this research that aims to obtain the required balance between psychological interpretability (i.e. explanation) and data science-related analytical utility.

Similarly to the Copenhagen interpretation that states a quantum particle may be forced into a different observable state each time, perceptions are equally hard to observe. For example, people may be reluctant to express how they feel in front of a researcher due to the social desirability bias [61]. In order to increase the ability to examine perceptions, this work conducted a varied spectrum of methodological approaches, which has been recognised as a key element in creating an increased ability to synergistically obtain new knowledge [12, 189]. Critical realism is, therefore, best suited as a philosophy of science behind this research, since it accepts that different types of knowledge objects (including social and conceptual) have different characteristics and that multiple lenses can obtain more complete findings than a single approach [203]. While critical realism provided a basis for the philosophy of science in this thesis, the central theoretical framework of this work is related to health psychology and applied computer science, which is described in more detail in Section 1.7.

1.4 Methodological opportunity

It is relevant to discuss the methodological opportunities that arise as a result of adopting critical realism as a philosophy. In order to assess which perceptions exist in the world of asthma patients as well as non-patients, the exploratory part of this thesis was aligned with a qualitative approach. When it comes to eliciting the discovery of new dimensions of the phenomenon, qualitative research is particularly powerful, as it enables the researcher to seek shades of meaning in the investigation [292]. However, this work also measured the prevalence of topics within groups of perceptions that are discovered and, finally, assessed the impact perceptions have on adherence. Considering that these are, in nature, quantitative questions, quantitative research also has a major role in this thesis. After the exploratory research and establishing which perceptions exist, quantitative research was useful as it is deductive in nature: theories are tested by formulating hypotheses and applying analysis that accepts or rejects these hypotheses [307].

However, quantitative research faces some challenges. Considering the nature of perceptions, many people are not very open in expressing their true opinions and beliefs, often due to social desirability of responses and demand characteristics, since surveys are similar to a laboratory setting where people are 'put under a microscope' [146]. For this reason and for the purposes of this work, it was recognised that it can often be more useful to collect more naturalistic data (e.g. openly accessible social media data), since people express themselves there more openly. Another challenge of quantitative work is that it often only focuses on particular models, linear relationships and takes a theory driven approach, rather than a data driven approach. This means this approach seemingly ignores the rich and diverse span of context and experiences of social context. The current work in the field of adherence to asthma medication is primarily focused on describing the causes and drivers of one's behaviour, however, we should also be able to successfully predict adherence to medication. This is where data science, a field that represents an extension of the existing quantitative approach, is particularly effective, since machine learning, as a set of data science techniques, is (more than psychology) reliant on inductive approaches and focused on prediction, rather than explanation [167].

In fact, recently, a whole range of research opportunities has emerged with the development of technology, the velocity, variety, and volume of new big data sources [23]. The 'quantified self movement' or recording and observing daily fluctuations of many health-related aspects of patients' lives has particularly contributed to this [211]. This movement can be especially relevant to health research. Novel data sources are a good route toward capturing attributes that are relevant to an individual's thinking, mood and behaviour as changes in activities online are often related to changes 'offline' [81]. In fact, Quercia et al. (2012) examined the link between users' Facebook contacts and their personality traits and found that the nature of online interactions is not different to interactions in the real world [236], which adds to the validity of examination of online sources.

It has been argued that big data collection tools and machine learning techniques are particularly suited for pattern recognition and analysis of large and often unstructured data, relying usually on inductive methods. They often elevate traditional quantitative work as they offer an insight into people's real behaviour (outside of the lab context), and they have less issues with replicability or challenges that arise from small sample sizes. Social media and other 'novel data' sources are a powerful pioneering approach to research, bringing the potential to gather insights about different population perceptions, lifestyles [210, 224] and even psychology - Stillwell (2020) stated that users disclose information about themselves more than they even realise [283]. These characteristics of big data and machine learning techniques that utilize this kind of data are discussed in more detail in Section 2.3.

Methodological opportunity is also reflected in the requirements of the research topics, perceptions and adherence to medication. There is strong support in literature that claims adherence to asthma medication is influenced by a range of factors (other than perceptions), such as demographics, personality and psychological traits [18, 282, 19]. This means that asthma medication adherence is *more than a pharmacological treatment*, because patients have to deal with a whole set of situational, psychological and social factors that potentially shape their daily lives and the way they perceive their condition. As such, a holistic approach is likely to be suitable in this research. It would help to create a 'big picture' about adherence to asthma medication and the assessment of importance of various sets of perceptions in relation to adherence. In other words, there is a need to observe a patient not only through their demographic traits, but to also take into consideration the circumstances and their social cognition about asthma.

1.5 Research Gap

Previously used perceptions: Patients' perceptions have been the focus of both qualitative and quantitative research and that work was mostly based on a theoretical background, such as the Health Belief model [56]. Examples of perceptions that have been identified by the previous research, as having an impact on adherence include: acceptance of one's condition [171]; perceived consequences and potential outcomes; locus of control (what caused their condition); perceptions about the inhaler and inhaler technique [145]; social support [244]; the perception of risks of non-adherence and others [34, 127]. However, the known theoretical models only explain 22% to 36% of the variance in adherence to preventive inhalers, which means these models could be open to improvements [39].

Contradictory findings: Studies in this field also vary in methodological approaches and results include some contradictory findings with respect to whether factors predict adherence to medication [89]. For example, some studies found that stigma has no impact on asthma treatment adherence [277], whereas others claim stigma can have a significant negative effect [66]. Additionally, one study found that participants that were married, older and white had better adherence [241], whereas other study stated that demographics explain only a small amount of variance in adherence [137].

A lack of conceptual definition of perceptions in computer science: It is also important to note that previous work about perceptions has been conducted usually in the field of health psychology, or in a business context (e.g. estimating customer perceptions). However, while there are ways of detecting sentiment in big data (via sentiment analysis [103]) and even opinion (via subjectivity analysis [313]) there is no clear method of identifying perceptions and capturing them using quantitative tools for big data, which is a unique research gap that this work addresses in Chapter 4.

Interchangeable nature of perceptions: The particular type of perceptions that are mentioned in this work are socio-cognitive perceptions that take into account that patients are a part of the ever-changing socio-cultural environment. These perceptions have been investigated only sparsely in relation to their effect on adherence to asthma medication. This means that studies have largely ignored perceptions patients have about their illness and asthma in the context of the society and the impact other people might have on patients' decisions. This study, therefore, takes into account perceptions about people with asthma, both from their point of view and from the point of view of the general public, taking into account their potential interconnectivity.

More than just a medical condition: The research about asthma and asthma adherence has traditionally been focused on the physical aspects of asthma, while the effect of patients' social cognition and situational features are less investigated [45]. However, perceptions and adherence to medication are more than a clinical topic, because every patient has their own psychological and demographic characteristics, habits, and opinions that drive their behaviours [154]. This idea is in line with the previously stated need for a paradigm shift and having a more pro-active health care that takes into account the fact that patient experiences should be valued and understood [15].

The theoretical perspective on adherence: There is a body of literature about factors that predict adherence to asthma medication (see Section 2.1.3). However, most studies only examine individual factors that predict adherence without taking into account the holistic nature of human experience and that different predictors of behaviour interact and inform each other. For example, research was dedicated to estimating the impact demographic characteristics have on adherence [89]; associations between adherence and: characteristics of asthma including severity and length [29]; perceptions and beliefs [26] and other factors. Some of the quantitative studies even include several groups of factors. For example one study had questions about demographics and lifestyle behaviours and asthma severity [57]. However, studies usually do not approach the research of adherence from a combined perspective, using a holistic approach to examine how several groups of features interact and affect adherence. Studies most commonly take into account only individual features. Previously observed factors can be organised into distinct groups such as situational, demographics, behavioural and cognitive factors, but studies usually focus on investigating these groups individually, considering only one or few variables at a time [177]. As a result, there is little consensus as to which of these groups is the most relevant to adherence to asthma medication. This gap is addressed in Chapter 6.

Stigma mechanisms: Stigma, as a group of negative perceptions related to asthma adherence, remains largely unexplored. Several stigma-related factors have been identified in the literature, however, current research treats stigma as irreducible and does not differentiate between underlying stigma mechanisms when it comes to assessing the impact stigma has on patients and their behaviours. For example, there is little to no research about stigma-related features (and their impact on adherence) such as disparaging humour, or the negative impact of media in the context of asthma patients. Similarly, research does not delineate between the source and the target of stigmatization (hence disregarding the role of non-patients in interventions) [172]. This is important because understanding underlying mechanisms of stigma could lead to more effective communication strategies and a greater focus on the patient's perspective, unlike today's practice where the majority of designed strategies come from the perspective of health professionals [15].

Limited qualitative and quantitative findings: The public health field, in relation to asthma, could in general be considered as data-poor (due to small sample sizes, lack of geographically linked data, temporal lags between the studies and ethical requirements). Promising here are non-traditional data sources, particularly big data [23]. While such data has had noticeable success in the field of emergency room visits related to asthma, adherence to asthma medication has not been investigated in this manner [239]. This is particularly relevant given that the existing research is usually qualitative in nature and quantitative studies often examine only linear relationships. This means that some of the relevant insights could have been omitted and there is a need to re-assess not only the relevance of particular individual perceptions, but also their group relevance to adherence.

The need for mixed methods: A convergence of qualitative, quantitative approach and its novel extension - data science, can shed light on the topic of perceptions and adherence in a brand new way that has not been recognised enough in terms of methodological benefits. The existing research usually uses only a single lens approach. However, as it has been previously stated, adherence to asthma medication is a multidimensional problem and perceptions could be seen as a 'wicked topic', that needs to be examined from multiple perspectives.

1.6 Research Objectives

This thesis married the strength of several methodologies: qualitative research, traditional statistics and machine learning techniques to investigate the mechanisms of socio-cognitive factors and how they affect a particular behaviour - adherence to asthma medication. This PhD combined not only methodologies, but also disciplines, specifically, data science and psychology. This is reflected in research questions upon which this PhD investigation is based. The following sections describe the four studies that were conducted, that used qualitative interviews and descriptive analysis of tweets (Chapter 3); machine learning techniques: classification and topic modelling in detecting perceptions in tweets (Chapter 4); traditional statistics to investigate stigma mechanisms (Chapter 5) and Model Class Reliance to assess which group of features is the most predictive of non-adherence (Chapter 6).

1.6.1 STUDY 1a: Examining patients' perceptions about life with asthma through interviews and STUDY 1b: Examining perceptions about life with asthma based on content analysis of tweets

Study 1 (a and b) represent a convergence analysis that aims to identify categories of perceptions that exist about asthma and people who have asthma. Obtained results discriminate between internal patients' perceptions (how patients see asthma) and external perceptions (based on interactions with non-patients). Through the use of in-depth interviews with asthma patients and content analysis of social media data, this work also created the comparison of what people with asthma believe others think about them and what others *actually* think. There are some significant similarities and overlaps between common themes, mostly related to stigmatization.

1.6.2 STUDY 2a: Filtering perceptions from non-perceptions in tweets and STUDY 2b: Measuring perceptions from tweets based on activation, evaluation and modality

This part of work is concerned with creating a procedure of perception mining on Twitter, that is composed of four parts (filtering perceptions, activation, evaluation and modality). This is followed by grouping perceptions according to the level of activation (active or passive), evaluation (positive or negative) and then extracting emerging themes to understand the content within these established groups. These tasks are accomplished via Natural language processing and classification models, as classic data science methods. The primary goal of this study is to obtain insights about more naturalistic public perceptions as expressed on Twitter, that are not made in a controlled, laboratory environment. The resulting four groups of perceptions can, overall, be summarized with two leading themes: the presence of stigma and the perceived sense of community that exists online. The secondary goal is to also quantify the presence of each obtained topic. This method of extracting perceptions could potentially be applied to another case study that deals with a different research topic (therefore, this method is not exclusive for perceptions related to asthma).

1.6.3 STUDY 3: Stigma mechanisms and their impact on adherence to asthma medication

Since stigma is recognized as a set of dominant perceptions of asthma, this study builds on the previous work that examines effects of stigma on adherence [7, 277], with the goal of investigating stigma as a non-atomic concept. Specifically, this is achieved via dissection of individual and cumulative effects that stigma related factors have on adherence, through the lens of Multiple Linear Regression. The individual features taken into account are: discrimination, denial, the effect of media, disparaging humour, exogenous perceptions and internalised stigma. Results demonstrate that only denial has a statistically significant predictive value, even though the majority of features have significant correlations with adherence. The second part of this research is focused on the mediating effects that information seeking and ignoring asthma have on the relationships between adherence and individual stigma features. This work emphasises that these coping mechanisms differently mediate relationships between different individual stigma-related features - which indicates that they all have distinct underlying ways of affecting patients.

1.6.4 STUDY 4: Using Model Class Reliance to measure group effect on adherence to asthma medication

The final part of this thesis focuses on empirically assessing the role and impact of perceptions on adherence. Having defined non-adherence as a multidimensional problem in the study, this work advances techniques able to take into account potential interactions between the groups and unpick non-linear relationships that exist between groups of factors affecting adherence. This study introduces a novel approach, Group-Model Class Reliance (Group-MCR), a method that is capable of providing the importance of features, by taking into account that many prediction models may fit the data [106]. The use of this machine learning method enables creating a full circle in this work, since by progressively moving to the examination of the most complex type of relationships, this study was able to reiterate the relevance of perceptions in affecting adherence to medication. Results highlight that out of all previously mentioned groups (e.g. demographics, lifestyle, personality traits), it is perceptions and their nature that need to be in the focus for the future work and interventions.

1.7 Central Contribution

The contribution of this thesis can be described as trifold. Firstly, this thesis creates a direct contribution in the form of empirical implications for the field of health psychology (and specifically, asthma research), which are elaborated on in the following Section 1.7.2. Secondly, this work also contributes to the field of applied computer science (data science) by introducing two novel methods of analysis: extraction and analysis of perceptions from text and introduction of a method named Group-MRC. Both of these contributions are explained in more details in the Section 1.7.1, as well as Chapters 4 and Chapter 6. These contributions to both health psychology and data science fields are also exemplified by publication venues that accepted academic publications resulting from this work.

Thirdly, and most importantly, this work also demonstrated the utility of synergy created by using different methodological approaches, in researching complex social and psychological problems. This synergy can be applicable more broadly and on another range of similarly complex topics. Understanding reasons behind non-adherence to asthma medication through the role of perceptions is one of these complex tasks, and it is difficult to examine it because it includes a myriad of nuances and intertwined factors. As a result, there are many unanswered questions about what people really think about asthma patients and how these perceptions affect patients' views of themselves and their conditions. Using a single lens approach is often insufficient, as each approach has its own limitations (summarized in Table 7.1). Equipped with insights from deep conversations with patients, big data in the form of tweets and responses from a survey, this thesis created a synergy between different methodological approaches to respond to questions: what perceptions about asthma patients exist; how prevalent are they and what is their impact on adherence. Therefore, one of the main arguments of this thesis is that combining different methodologies: qualitative research, traditional statistics and machine learning techniques can produce information that may be missing if only a single perspective is adopted.

This thesis, therefore, proposes new forms of triangulation, as each study had a unique contribution and the resulting insights are greater that the sum of individual parts. Specifically, qualitative research enabled identification of perceptions and a better description of a wider socio-cultural context in which perceptions are created and re-defined. These findings would not be as complete if a quantitative approach was followed at the start, as many underlying nuances would be missed. Conversely, quantitative approach provided a quantified prevalence of identified perceptions and a possibility to measure and interpret the impact they actually have on adherence. Interpretation of stigma mechanisms was a unique contribution of traditional statistics that would not be possible using qualitative approach. This work also highlighted the benefits of using primary data, collected specifically for the purposes of this analysis, as opposed to using secondary data that usually does not enable connection between data sources, which also means that we often cannot test the impact of features on the outcome. Lastly, big data analytics, as a novel approach, demonstrated the power of big data and greater sample size in measuring the prevalence of perceptions in publicly available tweets, which has the competitive edge over examination of opinions using traditional tools of statistical approach, such as surveys, that obtain a much smaller sample size. Additionally, data science tools are capable of identifying and untangling more complex relationships, which provided much better understanding of the essential value of perceptions - something that both qualitative research and traditional statistics could not achieve on their own. In short, the synergy of these approaches is capable of capturing more complete information about topics that are notoriously hard to measure and assess.

In summary, this PhD leverages a transdisciplinary approach, combining psychology and data science in order to examine the impact socio-cognitive perceptions have on adherence. The thesis is also set against a background of the Horizon Digital Economy Centre for Doctoral Training (CDT; UKRI Grant), at the University of Nottingham. As such, a strong multidisciplinarity is a key characteristic, which for this thesis also means strong links with both application and industry stakeholders - in this case Glaxo Smith Kline. As a consequence, the studies in this thesis combine qualitative underpinnings to engage with socio-cultural context, move on to quantitative research (both via traditional statistical approaches and predictive algorithms/novel data science techniques), before considering implications for medical research. The following summarises some of the contributions of this work in more detail.

1.7.1 Methodological Implications

As mentioned in the previous Section (1.7), one of the main contributions of this thesis is the proposal of advanced mixed methods that can be used to examine other ambiguous, fuzzy and otherwise problematic topics. The main goal is to show that in this particular case, the triangulation of various methodologies (and even paradigms) can create more impactful insights, rather than relying on traditionally used single sets of lenses. Furthermore, some main lessons that have been learnt (see Chapter 7) through this process can be useful for other case studies that deal with similar topics.

Additionally, methodological contribution is also reflected in the proposal of novel methodologies:

- The first one is described in the paper [182] that details the process of detecting perceptions in text data, using linguistic features and word embedding approach. As the extension of filtering perceptions, Chapter 4 established four pillars of perception definition: in addition to filtering, there are activation, evaluation and modality. The goal of this methodology is to detect perceptions from large quantities of text and to quantify their prevalence in short text, such as tweets. Establishing that public data, such as Twitter, is a significant source of relevant public perceptions, this methodology can be applied to examination of perceptions related to another concept.
- Secondly, following the current deficiencies of explanatory approaches in methods from Computer Science, that give rise to issues related to multicollinearity and interaction effects [125, 276], the work in Chapter 6 introduced previously mentioned Group-MCR. This modelling approach, that is

able to take into account multiple models and indicate each group's utility, has also been summarised in a paper published at IEEE Big Data 2021 conference [181].

1.7.2 Empirical implications for the clinicians and the field of asthma research and health psychology

This thesis also contributes to the filed of asthma research in several ways:

- This work provides the interpretation in terms of which groups drive non-adherence, which could provide a better understanding of asthma patients and what drives their behaviour. Identifying this group to be the group of perceptions (see Chapter 6), provides implications for the future interventions that could develop markers of non-adherence and allow patients and clinicians to be more proactive in controlling asthma [177, 29]. Discovering the reason why a particular group of features is the most relevant could more specifically direct future communication techniques. Needless to say, focusing on the wrong type of interventions leads to no changes in treatment success [29]. Therefore, instead on focusing on reaching patients in rural areas (if demographics were deemed as the most important) or reaching patients with a particular personality trait (in case personality was the most relevant factor of non-adherence), this work emphasises that efforts should be focused on reduction of stigma and changing not only patients, but also public perceptions. These implications are explained in more detail in Section 7.8.
- Uncovering the damaging effects of stigma and that they are reflected not only in discrimination acts, but also in disparaging humour and most importantly denial (see Chapters 4 and 5), could enable potential untreated

patients to be helped by specifically tailored asthma campaigns, thus, hopefully increasing the diagnosis level. Additionally, these insights could inform a more efficient communication between patients and their health professionals. Lastly, Chapter 5 implies that social media could be both the source of stigma, but also the source of support, which means that the future communication could be developed for carefully chosen social media platforms.

- This work was based on hypothesises that asthma is more than a medical condition and that patients' socio-cognitive factors such as perceptions are not only informing patients' treatment - they are a crucial aspect of it. The results of this work emphasised the importance of not only observing patients as integrated in their social environment, but also that patients could be treated as more proactive participants in their treatment if we take into account their perspective and their views. This could lead to better patient outcomes and lowering both personal and stakeholders' costs [206].

1.8 Overview of the thesis

This section describes in more detail the structure of this thesis. This structure is also visually represented in Figure 1.1. The thesis starts with a literature review, which was conducted with a goal of assessing previous research that exists in terms of: methodological approaches and also in terms of the current advances in the field of asthma patients' perceptions of their condition and research on adherence to asthma medication. The overall work conducted in these studies has two parts: the first is related to the exploratory analysis of perceptions and the second part is dedicated to investigating how these perceptions impact adherence to medication. Studies 1 (a and b) and Study 2 (a and b) are dedicated to the first goal of exploratory analysis and Studies 3 and 4 are focused on the second goal of assessing their relationships to adherence. The Discussion chapter (Chapter 7) reflects on the 'lessons learned' about the methodologies applied and how they helped in answering the research questions, as well as the overall discussion of the empirical findings. This is followed by the Conclusions chapter (Chapter 8) that closes this work.

The overall thesis emphasises that mixed method approach can be extremely useful in investigating topics such as perceptions, however, there are some significant limitations that were acknowledged in terms of each method used. Most importantly, the order in which different methods are used is discussed as a potential weakness. Additionally, this work recognises the relevance of the basis set by traditional statistics, but also the potential of using 'big data' (e.g. social media data) for the medical field, as well as the benefits of machine learning approaches when it comes to investigation of complex relationships.

Practical conclusions are reflected in the idea that both patients' perceptions and other people's perceptions about asthma play a major role in the life of asthma patients and can significantly impact the way patients adhere to their medication treatment. Stigma, in particular, is recognised as one of the major sets of perceptions and due to its damaging nature, a special attention was dedicated to further elaboration of stigma mechanisms. This work emphasises the need to address negative public perceptions with a goal of providing patients with a more accepting social environment they deserve. Additionally, some beneficial perceptions were detected, in particular the perceived sense of community, which emphasised the importance of not only patients' social environment, but also of its digital component as well.

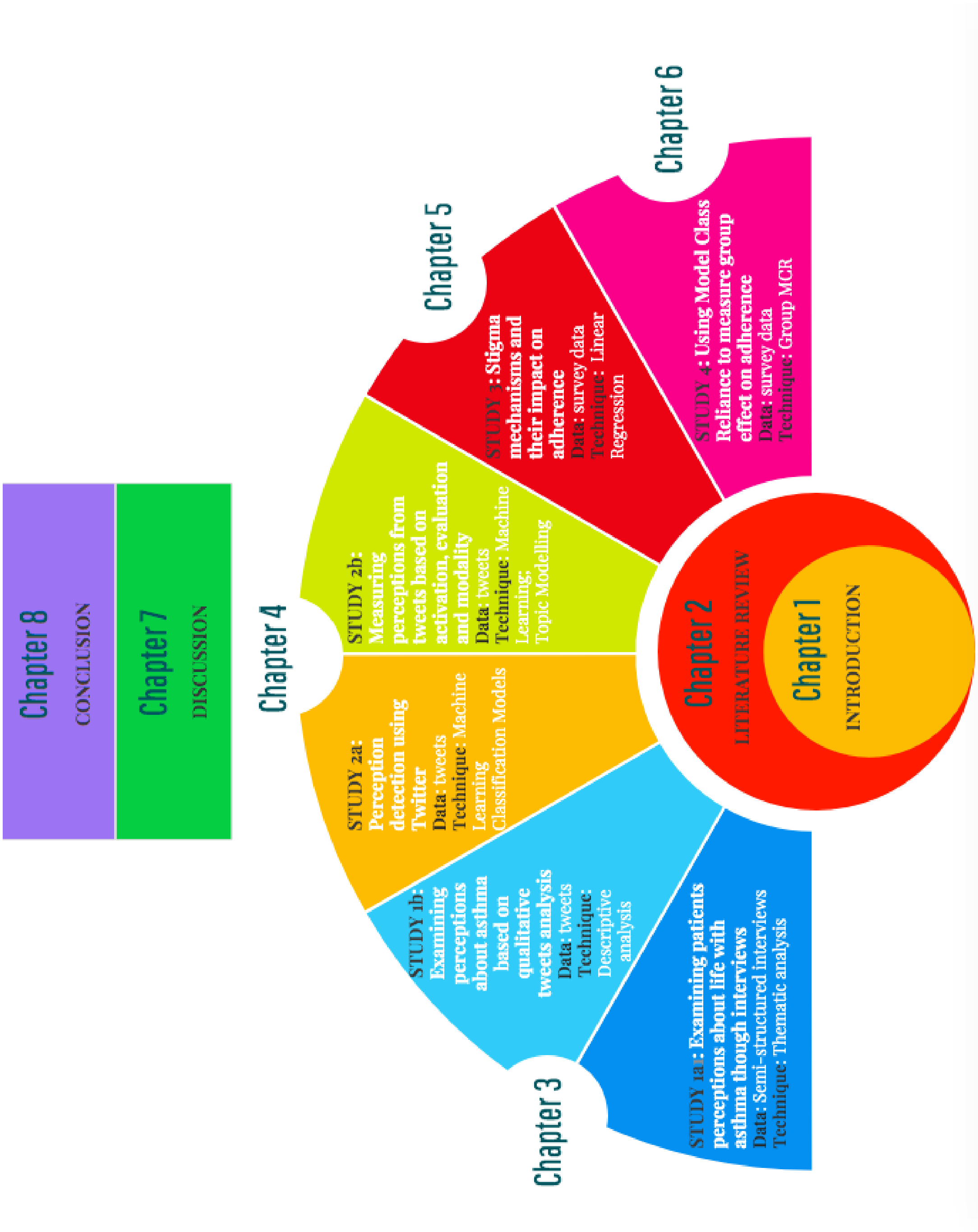

Figure 1.1: Structure of the thesis

Chapter 2

Literature Review

This chapter consists of three parts that support the theoretical and methodological aspects of this thesis. *First*, the general overview of the literature about the nature of perceptions and adherence to medication is examined, including the existing work about adherence to asthma medication, as well as perceptions, as socio-cognitive factors that affect adherence. Some previously used perception models are discussed, alongside their potential implications and importance to health. This chapter also highlights some relevant gaps in the literature that serve as a basis and motivation for the theoretical contributions targeted by this thesis. As well as providing a framework, this review of the literature around perceptions and adherence leads into the *second* part of this chapter, which reflects on methodological approaches that were previously used to examine the topic. Particular attention is paid to the results obtained in studies that were previously conducted in the field, in order to highlight how the research in this work can complement existing insights by introducing a novel mixed method approach.

Ultimately, the *third* Section of the chapter, is focused on discussion of the nature of methodological approaches that have been applied in the field, to provide substance for the methodological structure of the thesis. This section discusses both qualitative and quantitative techniques previously used to assess perceptions and behaviours. There is also discussion of novel methodologies that can be considered as quantitative, but which lean heavily on 'machine learning' techniques that are more inductive in nature than traditional statistical methods. Special attention is given to an overview of similarities and differences between these various new methodologies in order to highlight how they can complement each other, and how each has unique potential to contribute to the research.

2.1 Part I: The Nature of Perceptions and Adherence

> "If you say you have asthma, you're put into this unhealthy-something-wrong-with-you box"
>
> *- Nicole (34)*

The topic of this research are patients' perceptions and their impact on adherence to asthma medication. This part of literature review, therefore, focuses on unpacking adherence, and the factors that previous research has deemed as significant in terms of their impact on adherence to asthma medication regimen. The following sections explain the mechanisms that affect perception creation and the relevance perceptions have for asthma patients' behaviour and their health. Additionally, this section elaborates on the difference between medical and sociological views on perceptions, which at the same time highlights the key research gap considered in this thesis. Lastly, this section is concluded with discussion about the nature of this topic of perceptions, which due to its vague and unobtainable nature can be considered to be a 'wicked problem' [141]. Consequently, the literature review concludes with the discussion as to why a single-problem-solving mechanism may not be enough when researching perceptions and their impacts.

2.1.1 To adhere or not to adhere

There has been a long standing focus in the field of medicine to create effective medication that would be able to combat chronic conditions, including asthma. However, despite these efforts, effectiveness and benefits of the medication have often not been fully realised due to a fact that many patients simply do not adhere to their medication regimen [47]. Specifically, some studies argue that the level of adherence to asthma medication does not go over 50%, which represents a significantly suboptimal non-adherence for the population [20]. In fact, previous work states that despite a large body of research, adherence rates remain nearly unchanged [298]. This is problematic for several reasons.

'No one should ever die of asthma' is a common phase within health care, based on the idea that if an asthma patient used their preventer regularly, they would most probably prevent asthma exacerbation, attacks and other negative consequences of asthma [39]. Apart from the most extreme consequences, non-adherence can also lead to other (more mild, but still significant) asthma symptoms that affect both one's quality of life and everyday activities. Considering that around 3.5 million people in the UK alone have asthma, it is also easy to understand that the cost that asthma brings is not only significant to the patients and their carers [299]. Due to the frequent, urgent hospitalisations that are a result of asthma attacks, significant levels of health care spending are allocated to asthma yearly - in the United States non-adherence is estimated to bring the cost of 100 billion dollars per year, while in the UK, approximately £230 million of medicines are brought back to pharmacies [135]. There is little doubt that increased adherence to asthma medication would benefit a host of stakeholders, patients and public services alike.

However, making a choice to regularly use medication is a complex decision-making task and one where many competing variables are important. This is especially the case when it comes to *intentional* non-adherence, with a patient making a conscious decision to reject the treatment [284]. Previous work conducted states that asthma adherence, just as asthma itself, seem to be related to numerous risk factors. Recent, patient-related research implies that patient cooperation is of the essence, and that if patients have an active role in their treatment, this can significantly improve adherence [127]. Hence, socio-cognitive factors, such as perceptions should be taken into consideration in unlocking the roots of non-adherence. This is due to the fact that patients have to deal with a set of situational, psychological and social factors that shape the way they perceive their condition and how they behave as a result of this [154]. Before exploring these social factors in more detail, the next section tries to triangulate what 'adherence' is and what known factors have been examined so far, that implinge upon it.

2.1.2 What does adherence mean?

The literature defines adherence in relation to asthma, as a domain of discrete behaviours that correspond to different parts of its treatment [55]. For the purposes of this research, we consider adherence to be 'the extent to which a patient's behaviour corresponds to proscribed recommendation from their doctor' [49]. It is interesting to note that adherence, as a term, has not always been used in the literature - it has only recently replaced the term 'compliance', that arguably conveys negative connotations about patients. This is aligned with the idea that patients have a passive role in their condition management and that they are simply meant to follow their 'doctors' orders'. In other words, not complying with doctor's orders has been interpreted as patients' unwillingness to follow instructions and deliberately sabotage their own health. Today, describing patients as adherent or not helps us conceptualise them as active problem solvers that use 'common sense' when it comes to following their treatment procedure [137]. This change of terms is also reflective of the general shift in research that has started to increasingly observe patients as active partakers in a treatment [137]. This new perspective about patients also introduced a range of new factors that potentially have a role in affecting adherence. However, a patient's behaviour cannot be considered as completely led by common sense - there is still a vast array of factors that play a significant role and some of them can even be outside of the patient's control. Next, we briefly explore some of the factors previously mentioned in the literature, starting from a more traditional list.

2.1.3 Factors affecting adherence

Some of the factors that have been associated with non-adherence are: the complexity of therapy; fear of side effects; methods of drug taking; dosage regimen; understanding of the illness and its complications; lack of symptoms; lack of training on inhaler use; illness perceptions; social support; drug availability; disappointment due to no improvement [217, 50]; geographic macro areas; level of education; patients' beliefs about their therapy; type of inhaler [39]; perceived and actual severity of disease; asthma duration; locus of control [234] and others. In summary, and according to previous work, adherence is determined by one of the following categories: 1) therapy-related; 2) clinician related; 3) condition-related; 4) healthcare-related and 5) patient-related factors [39, 10]. Based on the variety of factors addressed here, we can see that non-adherence is considered to be a *multidimensional problem*.

It is also worth mentioning that a key delineation in patients' non-adherence can be observed as *intentional*, when a patient makes a conscious decision to not take their medication, and *unintentional*, when factors outside of patients control take place (such as: forgetfulness or physical inability) [135, 97]. The difference between these two types of non-adherence is that unintentional non-adherence can be understood in terms of skills and abilities, whereas intentional can be interpreted in terms of motivations [135, 20]. Therefore, the reason why patients intentionally do not follow clinicians' recommendations becomes more apparent when we take into consideration patients' attitudes, circumstances, lifestyle and related factors. To illustrate, Clifford et al. (2008) examined differences between non-adherers and adherers using the Necessity/Concern Model [64]. Intentional non-adherers had lower perception of the necessity of their medication, as well as higher levels of concern about taking it, than unintentional non-adherers and adherers [64]. Given this evidence of a key role of perceptions, the next section expands on some of the main characteristics of this concept.

2.1.4 What are perceptions?

> Not too long ago they discovered a painting, allegedly made by Van Gogh. By many accounts, it wasn't a pretty painting, it wasn't Van Gogh's style and not something that would catch any critics' eye. But what was interesting about this - they really struggled to conclude whether it was 'a Van Gogh' or not. For the purposes of this story it doesn't even matter what the conclusion was. What matters though is this: during the investigation, the financial value of that painting was changing from one moment to another - from being worth several millions (if it really was his) to virtually nothing. But the painting itself was always the same. And this is the power of our perceptions. Perceptions make us change our actions from logical to irrational. And we rarely estimate the real value of something only based on its functional feature. Even if it's only a piece of paper with silly flowers drawn on it.
>
> \- *excerpt from the Revisionist History podcast by Malcolm Gladwell*

Before we start the discussion about the nature and challenges of examining perceptions, it is crucial to try and pin down what perceptions actually are - particularly in the context of this work. In the example of food choice development, one study explains that 'perceptions' not only relate to both sensors such as vision, taste or hearing, but also to mental concepts, which include motivation, learning and experience [291]. Therefore, it is relevant to make a distinction between perceptions, based on *sensory information* and perceptions as *mental impressions*. This work is completely dedicated to the latter interpretation of the term. In other words, this research refers to perceptions as socio-cognitive factors, such as opinions and beliefs that affect behaviour (where in the case of this work, that behaviour is, of course, adherence to asthma medication). However, regardless of what kind of perceptions research is concerned with, dealing with perception-based information is deemed as much more complex than dealing with measurement-based information [220]. In the field of information sciences, one study argues that perceptions are intrinsically fuzzy and granular when they refer to perceptions in terms of recognising time, distance, force, shape and others [220]. Any framework created to examine and measure perceptions is burdened by a requirement of the complicated nature of the concept. In terms of psychological definition of perceptions, a more detailed description of differences between perceptions, opinions and attitudes are explained in section 4.1.3.

2.1.5 What is the relevance of perceptions, as mental impressions, with regards to health?

In the context of medical health, perceptions have been recognised as an important component, that potentially leads to emotional and psychological states that affect behaviours, such as asthma management [34]. This argument comes from the understanding that once a person is diagnosed with a serious health condition, such as asthma, they make several changes in their life [192]. The initial change happens when a person creates a description of the medical condition in their mind - their understanding of what is happening and how the diagnosis will impact their life. The second type of changes which occur are adjustments that a person needs to make in terms of their behaviours and thinking processes. In this step patients decide whether the prescribed medication and treatment is suitable and acceptable [192]. All of these changes produce a significant set of perceptions that one creates and uses in order to make sense of their situations, cope with it or simply learn how to adjust to life in a society with people who do not have asthma.

Understanding what kind of perceptions patients create and how these impact their treatment, and overall well being, can provide useful implications for health professionals and patients. Firstly, creating insights about what people think of life with asthma can inform current self-management education approaches [49, 188]. Secondly, if we could understand a non-adherent person, based on what they feel, think and believe in, this could lead to development of new interventions to identify other people with uncontrolled asthma. Moreover, this can then facilitate more efficient communication strategies with patients. Understanding what stops people from using their medication can help medical professionals 'speak the language' of non-adherent patients and address their negative perceptions directly. Negative connotations about medical conditions still exist, and many of them remain unaddressed, meaning that even today, many patients are burdened not only by their medical health challenges, but also perceived social oppression that remains unresolved [264]. This is one of the main reasons why we need to uncover potential negative perceptions, given the first step in solving any problem is, of course, identification.

2.1.6 More than a pharmacological challenge

For a long time, there has been a strong laboratory medical perspective taken when analysing asthma and similar health conditions. This means that even when it comes to researching adherence, the main focus has tended to be on pharmacological markers such as blood drug level whilst the behavioural science perspective emphasises pill counts and other non-adherent behaviours [45]. Research has rarely focused on mediating effects of both patients' social cognition about their treatment and impact of their situational factors (e.g. location, socio-economic status). However, adherence has to be considered as more than a pharmacological issue, given that patients have to deal with a whole set of situational, psychological and social limitations that potentially shape their daily lives and the way they perceive their condition [154].

There are also many unexplored factors that drive adherence. One study that conducted a hierarchical linear regression revealed that socio-demographic and clinical factors explain smaller amounts of variance than illness perceptions and beliefs about a treatment [137]. This highlights the space for improvement, especially as the overall medical field aims to shift towards a more *proactive* health care where patients have an active role [15]. The motivation behind such proactive initiatives is that inclusion of patients in the decision making process, through understanding their points of view will likely result in better healthcare, better health outcomes and increased patient satisfaction and quality of life [127].

2.1.7 Perception models from the current literature

As stated previously, people with a long-term condition, such as asthma, tend to create a mental representation of the condition in their minds - and these ideas subsequently affect their behaviours [136]. Perhaps the most comprehensive models used to investigate social-cognition related to patient behaviour, are the *Self-Regulatory* and *Health Belief models* [44].

The Health Belief Model (HBM) is one of the most widely used frameworks for examining and predicting health behaviours [212]. It is also a valuable instrument when considering why someone does not follow their treatment regimens. The model is based on the idea that in order for adherence to happen, a patient's perceptions about the benefits of taking medication must outweigh the fears and threats associated with the medication, which is called the attitude [17]. In addition, this model also takes into account self-efficacy - one's belief that one can successfully execute a behaviour [49]. According to HMB, these two components rest on one's actual knowledge about the condition, treatment and medication, but can also be influenced through communication with clinicians. Attitudes and self-efficacy are important determinants of the level of adherence [17]. However, these beliefs do not operate in isolation - HBM is often used in combination with socio-demographic characteristics, such as age, gender and location, which serve as moderating factors.

The second theoretical framework which is often used to explain adherence is the Self-Regulatory Model (SRM) [50]. This model, developed in the 1960s by Leventhal, aims to explain how an individual responds to a health *threat*. It follows the idea that perceptions, contextual factors and behaviours co-exist in a continuous feedback loop where behaviours are fed back into the formation of the illness/medication representation and they affect the adoption of coping responses [44]. SRM is based upon categories of patients' beliefs about the cause of an illness, symptoms that are part of the condition, consequences of the illness for the patient's life, how the illness is controlled or cured, and how long the illness will last [226].

Based on HBM and SR models, several studies have examined patients' perceptions including: 'acceptance' of their condition, a phenomena that was found to reduce psychological distress [171]; perception about time-line; and patients' descriptions of asthma and symptoms that also had impact on adherence [174]. Other studies have investigated perceptions of inhalers themselves (and found that patients' see them as effective) [145]; there are also studies about the perceived and actual severity of asthma and perceptions about asthma control that identified barriers and drivers to self-management [34]. Perceived necessity of medication has been identified as one of the main predictors of adherence [127, 50]. Similarly, patients who perceive fewer negative consequences have been identified as having better adherence [39, 64]. Intuitively, when patients perceive their asthma as more severe they are more likely to adhere to prescribed medication [21]. Conversely, patients who believe they do not have asthma when symptoms are broadly absent, are less likely to follow their treatment [128].

Interestingly, studies detailed above have not taken into account *other* people's attitudes and beliefs. In other words, a patient has only been examined in isolation. However, it is important to consider the social side of patients' perceptions. To support this, Bolman et al. (2011), state that existing theoretical models that have been tested with regards to adherence are open to wide improvement - they only explain up to 36% of the variance in adherence to prophylactic (preventive) inhaler [39]. The following sections build on this idea and introduce the more social side of patients' perceptions.

2.1.8 Theories that assess social environment of patients

There are also some other theories that do take into account the interplay between individuals and their environment. For example, a study related to aging mentioned Social cognitive theory (SCT), which is focused on both personal agency and the importance of context, which includes barriers, expectations and facilitators of behaviour that originate from social relations and cultural forces [243]. Similarly, socio-ecological models identify several factors that may affect one's behaviour and their factors range from one's biology to community, geopolitical situation and policy making [243]. Therefore, these models of health behaviour address individual, peer, familial, relational, community, and societal factors that could affect one's health outcomes (such as in the domains of reproductive or mental health) [88]. However, this work in the domain of adherence to asthma medication is scarce and often ignores the aspect of *patients' perceptions*.

The socio-ecological model was used to identify the influence that individual, interpersonal, institutional, community, and policy levels have on adolescents who have asthma and recognised self-efficacy, outcome expectations, and perceptions about barriers as the main factors that affect students behaviours [247]. In terms of the interpersonal barriers, it was found that the lack of knowledge, and inadequate asthma management skills, as well as forgetfulness were the main barriers [37, 247]. Additional barriers were detected by Blaakman et al. (2014) that found that the distance to the nurse's office, the necessity of hall passes and morning school routines can also represent additional barriers to treatment adherence [37]. A study conducted by Zaeh et al. (2021) investigated the transitional changes young adolescents go through, in terms of their asthma treatment adherence and concluded that factors such as community and systems related challenges such as school level policies and the price of medication can contribute to an increasing adherence in adolescents [321]. However, while this work takes into account the social factors in the patient's environment, an aspect that is usually missing in this kind of work is related to other people's perceptions, such as existing and perceived stigma and other perceptions that arise as a result of patients' interactions with other people.

2.1.9 Redefining patient's identity

It is important to discuss the creation of perceptions about one-self that arise once a person is diagnosed with asthma. People assess how they perceive their condition (and treatment), but even more importantly, they create an image about 'an asthmatic' - and decide whether they can incorporate that new part of their identity with the existing sense of self. The key here is, this process does not happen in isolation - we can hypothesise that patients internalize what they know about asthma based on the public opinion, which they then use to make their own judgements. This section is dedicated to exploring this in more detail.

As a result of examining perceptions through a clinical lens, there have been many efforts to increase adherence by creating interventions simply based on education of patients. However, these interventions obtained a limited success [298]. A potential explanation for this is that patients themselves, may not be the only responsible culprits for a lack of adherence, nor even at the core of the problem. Even if we consider a patient's *intentional* decision to not use their inhaler regularly, this decision can still be a result of underlying fears of judgements, or even internalized stigmatization. A similar idea was recognised by the World Health Organization (WHO) who state that the patient cannot be considered the sole source of the problem of adherence [47]. It is possible that patients' social environment (including the online one) can strongly affect how patients feel and what they think [79]. This is another reason why the nature of the perceptions and adherence to medication, should be observed using a wider, social perspective.

Identity has been recognised as one of the dimensions of Leventhal's Self-Regulatory Model (SRM) [50]. However, this dimension has been described as the label a patient uses to describe asthma and symptoms they associate with asthma [73]. The concept of identity in this context is different to the notion of identity in social sciences, which is defined as a representation of a subjective concept of oneself [118]. This distinction is important as it implies that previous literature may not take into account the manner in which society contributes to the way an individual creates these perceptions. Perceptions about asthma, their meaning and reputation can be interpreted just as any other cultural capital - a set of beliefs and attitudes that inform patients' identity [309]. These perceptions are built on social norms, that just as for anything else, can change through time. This is an important consideration, as it opens up the examination of perceptions to a more socially constructed point of view. The relevance of this is that perceptions are now not only assigned to patients, but exist outside of their own realm of control. This is particularly relevant for negative perceptions - namely *stigmatization.*

A relevant characteristic of a social identity is that it represents an individual's self representation and can be created based on one's membership in a social group [3]. People like to signal their group memberships through many signs, whether through clothes, music preferences or some other form of assimilation.

However, the problem with asthma patients' arises when they are *obliged* to be a part of a group (of people who have asthma), yet cannot reconcile the identity of an asthmatic with the remaining characteristics of their social identity. In such cases, they can experience a loss of or a diminished sense of self [5]. Most frequently, this happens if patients perceive that being an asthmatic brings with it negative connotations they do not wish to be associated with, resulting in disassociation from the unwanted group. In the case of asthma, this dissociation may be reflected in low uptake of medication, not using an inhaler in public - and even denial of having asthma altogether. This issue lies at the heart of the reason why our perceptions of others could have a negative impact on our individuals' perceptions and even our own well being. However, the idea that a patient is a part of the ever-changing society and culture has rarely been taken into account in research about asthma treatment adherence. This is a further gap that this thesis aims to consider, while acknowledging that people and their opinions and beliefs are created through complex social processes existing in our culture, social practice and history [98]. Considering this, we next extend literature about perceptions by investigating what non-patients 'think' as well.

A social disability model can be a useful theory for exploration in this context. According to such a model, a health condition that is exposed to stigmatization can be observed as a *social phenomenon* that is not universal and permanent [264]. This is an important shift in our thinking about perceptions. Previously, even when the social dimension of perceptions that affect adherence is recognized, they are still considered at the level of individuals' perceptions only - related to their attitude toward their condition and medication [47]. There remains a lack of attention to social perceptions and social oppression, that may exist in the form of perceived socially-related barriers. Issues such as stigma are likely particularly important and are, therefore, examined in more detail in the following paragraphs.

2.1.10 The anatomy of stigma

The issue of stigmatization is one of the first prescient topics to be addressed once we adopt the notion that patients have a social identity that largely drive their decisions. Stigma is described in the literature as the gap between how society characterizes a person and the traits that a person actually possesses [7]. Stigma is often expressed through acts of discrimination, in which case it is called enacted stigma [278]. However, stigma can also be internalised, when a person who is an object of stigmatization, has a level of agreement with that social stigma itself [7]. Additionally, stigma can be *anticipated* or *perceived* [93, 278, 9]. Interestingly, stigma is frequently expressed in language and remains undetected. Research papers about asthma itself can sometimes contain stigmatized terms, with asthma being referred to as a 'disease', an 'illness' with people who have asthma frequently being called 'asthma-sufferers'. Lately, significant effort has been made to exclude the term 'asthmatics' and instead, use people first language, given that asthma, just as any other health condition, does not (and should not) define people [102]. Even though the use of language in academic papers is far from the biggest issue in the generation of stigma, it still serves as a good illustration of how stigma can be unnoticed and even seem harmless. However, even a simple use of language to describe someone as 'having a disease' can make a further ostracisation of people who are presented as the opposite of the normal standard - a healthy person that has nothing 'wrong with them'.

The relevance of stigma can be reflected through the idea that perceptions other people have about asthma shape the social standards of prejudices [108]. Negative impacts of marginalisation and judgments may be reflected in the way patients' see themselves [34, 127]. To illustrate, previous work has identified that asthma and inhaler are often characterised as *weakness* [82]. There is a body of research that indicates patients with asthma report feelings of shame, lower self-esteem and overall lower quality of life [7, 77, 15]. Research that investigated the impact of stigma on epileptic patients highlighted the negative effect stigma has on patients' self-efficacy and self-management [294, 7]. Indeed, stigmatization reflected in negative perceptions about asthma can be especially detrimental to one's health, as it is one of the main causes of delays in health seeking. Public stigma about mental illness, which is a more broadly investigated topic, is seen as a barrier for many people with psychiatric illness for reaching their personal goals and quality of life [71]. In the field of asthma research, some studies imply that patients' adherence to asthma medication and asthma self-management are influenced by perceptions that people have about asthma and its treatment [34, 127]. However, these are mainly qualitative studies, whereas quantitative studies are sparse and leave a space for further research due to some contradictory findings. Lastly and most importantly, stigma is currently observed as a unified concept, while its underlying aspects and mechanisms though which it affects patients remain largely unexplored.

2.1.11 Wicked topics

As previously mentioned, it can be particularly hard to research perceptions, due to their somewhat nebulous definition. Perceptions, as socio-cognitive factors, are related to *wicked problems*, a term used to describe ill-defined and ambiguous challenges that straddle across several disciplines [141]. This (almost comical) name for such problems was created because there is very little consensus about what a 'wicked problem' is, let alone how it might be solved [141]. In the case of perceptions, there are several reasons why we may struggle to even identify them, and the following sections describe some examples to illustrate this challenge.

In order to illustrate ways in which perceptions can be hard to observe, measure and quantify, it is useful to make a parallel with the radical concept from the field of physics - the Copenhagen Interpretation of quantum mechanics [280]. This perspective states that a quantum particle may be forced into a different state each time we try to observe it. Simply said, as soon as we attempt to measure a position of a particle, its momentum loses its certainty and we are left with no definitive information about particles movements. One of the reasons why this parallel is chosen is because in both cases, the only way a researcher can be objective in the measurement of perceptions is if the researcher is not involved at all [63]. There is a certain unification between the object and the subject that is explained as a phenomenon in Copenhagen Interpretation, since instruments employed predispose the scope and conditions of the observation that, based on this, introduce a level of determinism [63]. This is the case when examining perception as well.

Even though qualitative interviews are considered as one of the most powerful tools in capturing participants' experiences which are usually unapproachable by quantitative research, they are not devoid from the subjective input of the researcher [292]. Likewise, data science techniques also require a level of subjectivity from the researcher even when it comes to defining what is considered to be a 'perception' and what is not (for example when it comes to manual labelling of the training set in a classification task). The key issue here can be explained though pragmatic scales that are useful in linking cultural models - since the true meaning of what is written can be disconnected to what is read by the researcher [101]. Even quantitative techniques 'suffer' from biases created by researchers' subjectivity especially when it comes to their reductionist nature and making decisions about which perceptions should be researched in the first place.

There is another level of ambiguity that exists in relation to the challenges of researching perceptions. When it comes to assessing the impact and underlying mechanisms of negative perceptions, there is usually no distinction made between the object and the subject of stigmatization. In fact, stigma is observed as an atomic, overarching concept. Very little attention is given to the granulation of stigma into underlying stigma-related perceptions that may each have their own nature and different impacts on patients. Therefore, this issue is related to challenges in deciding how granular our perspective on perceptions should be and how this may impact future interventions.

Another significant characteristic of perceptions, that echoes the ambiguities inherent to the 'Copenhagen Interpretation', is related to how they are expressed or found. Not only can perceptions be contextual and unconscious (similar to the example of stigma found in the language in academic papers), they can also be intentionally undisclosed and inter-linked to perceptions of others. To illustrate, respondents may be reluctant to disclose how they perceive something when they are protecting their reputation, status or attempts to sway the outcome of the research, which is referred to as 'vested interests' [256]. The research of social desirability bias indicates that people sometimes deny socially undesirable actions because they wish to be perceived as more altruistic and seen in a positive way [61]. Similarly, in the case of negative perceptions, such as stigma-related perceptions, patients often choose to hide that they feel stigmatized or as an extreme - they completely dissociate from the identity of an 'asthmatic'. This makes examining their perceptions particularly hard, regardless of the approach. The potential cause of such phenomena could be that patients do not want to be exposed to negative perceptions, or they simply do not want to be pitied and be perceived as weak. Therefore, when it comes to investigating perceptions, participants can unknowingly hide information; they can lie or choose not to disclose information in the case of desirability bias. However, by far, the hardest obstacle in investigating perceptions is when they are completely internalized and unconscious, yet all the while still strongly driving one's behaviour (which is the case with completely internalized stigma).

In summary, perceptions are difficult to define, and it is likely impossible to do so exhaustively. Additionally, their research usually introduces a significant level of bias (that needs to be recognised as a limitation and carefully accounted for when possible). Even more importantly, regardless of whether it is the subjectivity or a researcher, or the lack of expressiveness of perceptions even when they are formally defined, perceptions can be extremely hard to capture (similar to the atom that keeps changing its form every time someone looks at it). Having elaborated on the nature and challenges of examining perceptions, the following section (Part II) reflects on methodological approaches emerging from the previous works. Finally, in Part III of this literature review, a discussion about accessible methodological approaches to be used in this thesis will be considered, aiming to turn this research problem into a challenge that is sometimes - a little less wicked.

2.2 Part II: Previous methods used in the field of asthma perceptions and adherence to medication

Previous sections introduced topics of perceptions and adherence to asthma medication. We can now reflect on methodology and techniques previously used to examine asthma perceptions and adherence to medication, as well as their results. To accomplish this, we examine relevant key studies focusing on their methodological approach, which highlight the varieties of research strategies available. For this methodological review, snowball approach was used to obtain scientific journals in English language. Papers taken into consideration were related to either of two core topics: perceptions about asthma and/or the impacts on adherence, in order to highlight potential opportunities in terms of methodological approach that this thesis aims to address.

2.2.1 Previously conducted Qualitative work related to perceptions about asthma

The majority of papers examining asthma perceptions or adherence performs either qualitative research or traditional statistics approach, while big data analytics and machine learning work is not as prevalent. Qualitative research is usually used to support an exploratory analysis. To illustrate, experiences about life with asthma have been examined in several studies via interview-based approaches, resulting in some valuable conclusions. Adams et al. (1997) identified two main groups of asthma patients: accepters and deniers, based on the level of denial they experience [5]; while a study conducted in Canada observed that experience of asthma can be described through phases ranging from diagnosis to acceptance [278]. A significant body of work was also dedicated to experiences of adolescents and examination of how they perceive life with asthma. Based on these qualitative studies it can be argued that adolescents often have a poor understanding of their condition [95] and do not want to be defined by asthma [154]. It is also worth mentioning that another set of qualitative research examined perceptions about asthma and asthma treatment [260, 21], as well as self-care strategies [184]. These findings reveal the power of qualitative research to provide a rich context to the topic at hand and provide a starting point for further work via exploratory research.

It is also relevant to acknowledge that this particular characteristic of qualitative work has led to uncovering stigma about asthma. Findings about stigma range from detecting the presence of stigma [82] to investigation about how stigma impacts daily life. A Malaysian study found that stigma was mentioned in terms of disclosure, discrimination and that patients' feelings were affected [6]. Another study adds to this list the prevalence of blame from non-patients, especially in the case when patients smoke [31]. Some notable mentions of exploratory analysis include studies that did content analysis of the newspapers [151] and films [62], adding to the existing literature the evidence about stigmatization in the media in the USA. Additionally, while many studies perform thematic analysis, some have different approaches: for example, one study conducted interviews and performed systematic text condensation to emphasise the experience of adolescents [154]. However, a significant characteristic of qualitative work is that is does not 'suffer' from contradictory findings, given that it is less focused on obtaining generalizable results.

2.2.2 Incongruent findings in the Quantitative field

Some quantitative studies in the field of asthma perceptions and adherence may reach different and even contradictory findings, which has been referred to as the 'replicability crisis'. For example, previous work has failed to identify clear and consistent relationships between adherence and socio-demographic variables, such as gender and age in adults [135]. On the other hand, a set of studies have implied these individual, demographic features are related to adherence. Three studies have indicated that patients with better adherence behaviour are older, male, married, white and have a higher social status [39, 241, 20]. In terms of perceptions related to treatment, one study, that used a telephone survey (n=150) and analysed data using bivariate linear regression, concluded that patients who have a greater concern about their medications were more likely to be less adherent [68]. A similar finding was obtained both in a study using multiple regression (n=64) to analyse survey data [50] and a further study consisting of far more extensive number of participants (n=622) [69]. Conversely, a different work (n=71) observed that patients' treatment beliefs *did not* seem to predict adherence [10]. Similarly, Sibbald et al. (1989) stated that attitudes, such as pessimism, were only weakly associated with behaviour (n=210) [267].

Negative perceptions and particularly stigma, have raised significant debate in the field, following some opposing results. A quantitative study by Vamos and Kolbe (1999), that examined psychological factors in severe asthma, found that asthma-related stigma was one of the 3 main attitudinal factors [293]. Stigma has also been found to be significantly associated with poor asthma control in another survey-based study by Andrews et al. (2013) [15]. Furthermore, subjects who feel embarrassment due to stigmatization were found to be less likely to use their medication in front of others or take it with them when they go out [66]. This stands in contrast to less recent work about self-care with 18 participants, which found that stigma was not established as a major theme in patients' experiences [277]. Whilst denial is often a prevalent topic related to stigma in the qualitative work, survey-based research found that denial and stigma are not significantly correlated [52] and that denial is not associated with adherence [70]. These examples of quantitative work exmplify that while quantiative work enables us to quantitatively assess the impact of perceptions on adherence (unlike qualitative work), it still has some disparities in results across studies.

The reasons as to why quantitative work has conflicting results in the field can be associated with a range of discussion points. First, problems with replicability can be there due to small sample sizes. While a smaller sample size is a common occurrence in qualitative work, some of the quantitative studies detailed above were based on less than 100 participants [10, 50, 17, 15, 293]. This is, of course, understandable as it is often not easy to access a large number of asthma patients. There is also a limited number of statistical techniques that were used to interrogate the data. Many studies were also only able to provide separate analyses about each independent feature and their relationship to adherence. A larger, model-based analysis, which could have vastly helped understand interactons between variables, have been absent likely due to their complexity (this particular challenge is one addressed in the final study of this thesis). In particular, a majority of techniques used in previous research could also be classified as ‘traditional statistical analyses’, such as Multiple Linear and Logistic Regression, Pearson correlations, t-tests, etc. The lack of big data methodology on design given the embeddedness of adherence in lifestyle, could also mean that contradictory results might reflect biases imposed by artificial laboratory settings.

2.2.3 Next steps

It is clear that the field of adherence to asthma medication and patients’ perceptions has previously been dominated either by qualitative research or analysis via traditional statistics, designed for appropriate setting. There are some studies in the field of asthma research that have used ‘big data’, however they are exclusively focused on predicting admission to the emergency rooms and unlocking clinical asthma patterns. For example, one study developed a syndromic surveillance to localization of illnesses by regions [224]. Similarly, Collier et al. (2011) used a more complex linguistic analysis, again, for health surveillance [67]. Lastly, some work in the field was dedicated to classification of tweets to detect those related to influenza [224]. Additionally, Twitter was used for sentiment polarity detection within asthma patients [186]. These are beginnings - yet, research about the impact *perceptions* have on adherence to medication, using ‘big data’, remains completely unexplored.

While there is evidently a significant amount of work undertaken, in the field of perceptions and adherence to medication, that contributed to novel insight and introduced some valuable tools, challenges remain. In particular, they are related to contradictory findings and a limited methodological set of tools that often do not allow examination of more complex non-linear relationships. In order to capture the richness of patients’ perceptions and their impacts on adherence, this PhD thesis, therefore, aimed to address these gaps in the methodological approach and introduce a fresh perspective on these topics by using a novel set of design and analytical tools. In light of this, this section investigates how new, potential methodological approaches differ to those previously adopted and how they can collaborate to provide new insights in the field.

2.3 Part III: Analytical approaches

Having reflected on the methodological approaches in the field of asthma perceptions and their impact on adherence to medication, this section is dedicated to discussion of opportunities that were identified in terms of methodological approaches. It is relevant to re-iterate that this thesis is multidisciplinary in nature, combining the fields of psychology and machine learning. The majority of the thesis is dedicated to quantitative research and, therefore, some of the main challenges and opportunities of novel quantitative approaches emerging over recent years and how they might apply to medical adherence are addressed here. Throughout, a contrast will be made between the classical approaches that underpin statistical analysis and the more inductive approach taken by data science, both of which have significant places in this thesis. Even though data science as a field has existed for many years, it became particularly popular due to the vast expansion of data sources, and sheer volume of data that has been created due to developments in technology. In fact, there remains some debate as to whether we even need a new term at all, given data science uses statistics at its basis [87] (this debate about similarities and differences between quantitative methods are acknowledged in Section 2.3.2). However, it is also relevant to elaborate on the long standing debate between quantitative and qualitative research, since a qualitative study represents the first part of this research work (see Section 2.3.1). Attention will also be given to often overlooked similarities between qualitative work and data science (or ‘big data analytics’) given their central contribution to this work (see

Section 2.3.3).

2.3.1 Qualitative vs Quantitative Research

For a long time, there has been a spirited debate between two extreme research paradigms: positivism and interpretivism [140]. Even though both paradigms integrate both qualitative and quantitative methods, qualitative research was most frequently aligned with interpretivism, whereas quantitative research is most commonly associated with positivism [284].

Qualitative research embodies the idea that people develop their identity with the special regard to the sociocultural context [25]. Having this in mind, it is especially useful to utilise the tools of qualitative research when investigating topics such as perceptions, since qualitative research enables a valuable context-sensitive interpretation of observed phenomena [28]. In addition, considering that qualitative research starts with the idea that everyone represents a world of their own, its research often starts with no ground truth. This is cited as the strength of the qualitative approach, that was also exercised in the context of this thesis: it enables elicitation of 'perceptions' that have not previously been mentioned in the research - and it sheds light on some interactions that quantitative studies might not be able to consider. Qualitative interviews remain being considered as one of the most powerful techniques in a researcher's armours, to capture valuable descriptions pertaining to participants' experiences and perceptions, which are often inaccessible to quantitative research [292, 27].

Quantitative research, nonetheless, also contains a whole range of methods that aim to investigate a social phenomenon. Previously, and most commonly, these methods have focused on statistics and the utilization of numerical data [307]. Sukamolson et al. (2007) state that quantitative research uses empirical methods in order to examine what happens in the real world, rather than what 'ought' to be happening [284]. This view puts forward the standpoint that quantitative research aims to maximise objectivity - and minimise the involvement of the researcher. Since these empirical statements are commonly expressed numerically, empirical analysis that evaluates these statements are, of course, grounded in mathematics. As such, statistics is most frequently applied in quantitative research. However, this also brings a slight misconception as it would be easy to assume that quantitative research is used to examine only phenomena that produce quantitative data [284]. This is not correct, since even though it may be difficult to conceptualise (and hence, measure) a particular phenomenon, it is often still *possible* - using various forms of extrapolation and ground truths. This is particularly relevant for this thesis. Perceptions can be considered as one of these potentially intangible topics, which are hard to 'translate' into a numerical form. Despite this challenge, there is a large value in measuring the prevalence of various perceptions, as well as their impact on adherence: these are the questions that are innately quantitative in nature if results are to be generalized and that qualitative research on its own may not be able to respond to.

There has been a dichotomous, (and potentially unfruitful) long-standing debate about whether qualitative or quantitative research is 'better', leading to many discussions and 'paradigm wars'. It perhaps is better to recognise the fundamental differences between these two approaches, to allow them to offer unique contributions when examining a particular topic. Other than previously mentioned differences regarding 'reality' (which is more related to their philosophical assumptions), some additional differences between qualitative and quantitative research include the reproducibility issue and differences between descriptive and inferential nature of research.

While qualitative research is more aligned with an interpretivist approach that states generalisation (and hence, replicability) is not possible (as there is no single reality), quantitative approach tends to follow a positivist idea that *there is* a single reality that can be observed [153]. Based on this, quantitative research should be able to be replicated and the same results should be obtained across different studies. Finally, while quantitative methods are better at looking at cause and effect (causality, as it is known), qualitative methods are more suited to unpicking the 'meaning' of particular events or circumstances. However, in some cases, it is not useful to only *explain* behaviour, but to be able to test that explanation and predict the future instances of the problem - which are more commonly tasks of quantitative approaches - both traditional statistics and data science [167]. The importance of these differences (the reproducibility and the nature of the research) are central to this thesis. Its aim is to make an empirical contribution that goes beyond the boundaries of explanation about which perceptions exist - the hope is that such research might impact policy making and design of future interventions for asthma patients. This is why both inference and explanation are equally important for the ultimate goals of this work.

2.3.2 Statistics vs Data Science: The Judge vs The Detective

In the last couple of decades we have witnessed exponential growth of databases, filled with data coming from diverse sources. Today we live in a society that continuously collects data based on various digital traces, about almost every single person. This data can be found in fields related to almost everything, from banking, businesses, to physics, medicine and accross the sciences [129]. This novelty of data growth was followed by new mediums, new approach to analysis and new problems [283]. This gave rise to data science, a discipline that combines statistics, machine learning techniques, pattern recognition and database technology [129]. Data science techniques can be considered to be quantitative in nature, similar to traditional statistics that are most frequently associated with quantitative approach, yet more inductive in nature. Traditionally used statistics represent one of the building blocks of data science. However, the introduction of data science, which often utilizes the large samples made available by technology development, enabled researchers to interrogate some problems that they were not previously been able to tackle. Before describing some of the focal points of discussions between classically trained statisticians and data scientists, it is useful to unpack these two fields in more detail.

Statistics have the roots that stretch back for at least a couple of centuries, even though the origins and the disciplinary boundaries of statistics are still a matter of discussion [130]. Some authors claim that the name of this discipline has Italian roots from the 16th century and carries the meaning of collection of information that are of interest to a salesman (statista) that deals with matters of the state (stato) [32]. This field encompasses a variety of techniques that can be seen as descriptive and inferential statistics [323]. In statistics, the 'theory' traditionally leads the investigation with data directly collected in order to respond to the particular question that motivated the research objective. Many modern disciplines rely heavily on statistics and hypothetico-deductive approach - and statistics provides the basic principles that are carried through to data science method. However, data science appeared as a novel discipline, with an intent to focus on similar problems that statistics had been addressing, but with new inductive tools and strategies [130].

Data science is a much younger field that has roots that go back to 1950's with Arthur Samuel coining the term 'machine learning', a crucial aspect of the discipline [131]. Even more than statistics, data science as a discipline has a notoriously vague domain and boundaries, with interpretations largely depending on the background and academic education of the person asked about them [109]. Nevertheless, it is possible to isolate some major tasks that data science is concerned about: exploratory data analysis, descriptive and predictive modelling and discovery of patterns and rules. However, unlike statistics, data science has a serendipitous element in the process of data analysis, as it does not always start with a clear hypothesis in mind and is more focused on modelling patterns and unearthing previously unknown relationships and structures [323].

The biggest difference between data science and traditional statistics, is the type of reasoning used. As mentioned, statistics has a symbiotic relationship with theory that provides its conceptual framework and based on previous knowledge, it aims to reject or accept that theory [323]. Most classically trained statisticians work with primary data, which is the data collected for the purposes of responding to predefined questions [129]. Statistics are *deductive* in nature. In contrast, data science is concerned with detecting and modelling relationships, without being as constrained to create assumptions about the nature of the relationships before the analysis is done. This makes data science *inductive* in nature, and more interested in detecting interesting, operationalizable patterns than respond to previously defined hypotheses [87].

One of the major criticisms aimed at classical statistics is that it adopts a reductionist approach that seemingly ignores the richness of human experiences and the world - since it only focuses on particular questions, based on predefined theories [161]. This is also related to a risk of confirmatory bias, the situation where researchers do not take into account anything that does not corroborate their theory of interest [323]. They are also expected to be unprovoked by data and create new knowledge without relying on new discoveries during the analysis. On the other hand, one of the most frequent criticisms aimed at data science is that search for interesting patterns can identify structures in data that emerge based only on pure chance. Additionally, while it is not possible to ask a computer to provide insights on its own, critics have also suggested that data science relies too heavily on computation. However, a purely inductive approach is in actuality very rare. Even when we use big data, a certain bias is still introduced. For example, in the step of 'feature engineering' (producing independent variables using techniques such as aggregation), a researcher can introduce a bias by simply choosing the method of feature creation or even by choosing which features to include in the predictive model. Additionally, computers cannot (yet) guide the whole research process. Patterns created in the data are not inherently meaningful until a researcher makes sense of them and assigns them value [161].

Within a mixed method approach, one might preferably consider a statistician to be like *a judge*, that verifies clearly stated hypotheses; whereas a data scientist is more akin to a *a detective*, that is open to unexpected turns and revelations [323]. Yet, data science and statistics overlap in many aspects, and the fact remains that statistical literacy is a crucial basis for a data scientist. Both philosophies have to be addressed as well, especially in the context of making a verdict of which of these two approaches is more suitable for the research in this thesis and if they should work together or not.

Type of data

The first difference between classical statistics and data science is related to the type of the data that each of these fields use in their respective analyses. While classically trained statisticians typically do not deal with data sets that have thousands or millions of data records, machine learning often uses such data. 'Big data' is now used to describe these large amounts of data that have varied and complex structures. Their use is normally followed by difficulties in storing, visualising and analysing when traditional techniques are used [257]. Big data can be in the form of text, images, even video and the analysis requires cleaning, integration and often, engineering of features that will be used in predictions [87]. In short, big data is different to traditionally used data sets: high in volume, variety (heterogeneity), velocity (the speed by which the data is generated), and low veracity (lower quality and reliability, with increased noise which often characterizes user-generated content) [262].

Big data often creates practical problems that statistics have previously not encountered. One of the most significant challenges is related to *storing the data*. While this would normally lead to minor obstacles, the issue of data storage can also be extremely important, as we witnessed during the COVID-19 pandemic in the UK. Namely, during 2020 a total of 15.841 COVID-19 cases in England were not correctly stored due to a Excel spreadsheet row limitation, which, a quasi-experimental evidence states, led to additional 125.000 infections and 1.500 deaths [105]. It is important to reflect on the idea that Excel can be interpreted as a go-to tool for many professionals who find themselves half way between statistics and data science. Unfortunately, there is often no good half-way solution and, as stated above, this can be very detrimental to the rigour of research. When dealing with large scale data, data should be processed and stored in the appropriate database, following a correct protocol and testing protocol should be in place [129].

While classical statistics works with solely numeric data, an array of opportunities are provided through 'big data' that does not only have the form of numerical data, but also images, sounds, text data and geographical data [129]. This opens the door to researching topics that previously suffered from potential biases introduced via classical data collection and processing. Perceptions, the focal point of this thesis, are a case in point. Previously, obtaining data on perceptions to inform adherence programs was extremely difficult in terms of collection and scale, despite some studies providing evidence that novel data sources are particularly useful in capturing behavioural attributes [81]. Twitter, as a source of text data (and a platform that was used in this thesis) stands out here, with only do less than 10% of Twitter users restrict access to their accounts, much of the previous work has recognised that Twitter can bring a better understanding of its users and impact on the society [196]. These disciplines range from computer science, health, communications, sociology, etc. So far Twitter was used to: classify tweet messages related to influenza [67]; to follow patterns in tweets to make various predictions [239, 38]; to monitor social dynamics [133]; to predict personality traits [235, 120, 317, 303] and others.

There are some additional reasons why big data sources can offer increased research potential. They are able to capture data unobtrusively, which means that data can be collected without the desirability bias and influences from the researcher. This is different to psychological studies that are conducted as laboratory experiments, or even following the rigorous structure of a qualitative interview. If big data is collected based on actions or behaviours of participants (e.g. transactional data or tweets), very often participants are not even aware that their data is a subject to analysis, and they do not consequently attempt to change or adapt their behaviour. Further, the collection of big data is often cheaper even though it enables more detailed analysis. For example dynamic tracking of customers behaviour enables analysis of trends over time and space, which depending on the topic, can be a much richer source of data. Perhaps most importantly, big data also does not struggle with selection bias-distortion, which can be one of the most significant obstacles for statistics: this is particularly a problem when a sample is based on opportunity, rather than representing an idealised random sample [129].

Even though big data is often secondary (not collected specifically for the purposes of the specific research), it still enables researchers, with careful ethical forethought, to avoid collecting data from people who are more open to be a part of a survey or are *more* convenient participants for other biasing reasons. Of course, a significant challenge that big data has, lies in the subsequent data cleaning process. Clean data is a key for analysis, however, in the case of big data, which is often secondary in nature, it is more possible that data stored in various formats is invalid in some way. The importance of data quality is often described using the acronym GIGO: Garbage In, Garbage Out. Simply said, just because data is there it does not necessarily means it will be useful. It still takes the expertise of a researcher to responsibly delineate between what data could and should be turned into information, and what cannot.

Simple models

As mentioned previously, data science usually deals with large data sets that are readily available for training models, which empowers the formulation of predictive algorithms. However, the focus on prediction also introduces an unsuspected challenge: emphasising the 'prediction accuracy' introduces a bias toward simpler theories, since the accuracy of sparser models tends to be more generalizable on other sets of data [87]. Specifically, even the use of Occam's razor has been criticised (unless the simplicity of the model is the goal in itself), given simplicity unto itself can often be demonstrably harmful and can fail as a heuristic [90]. However, simpler predictive models can be powerful when machine learning combines prediction and inference. This occurs in tasks when not only prediction carries significance, but also the *interpretation* of the predictive model. For example, churn prediction is a common and important task in business [139]. However, companies often want not only to identify customers that have the strongest probability of leaving them, but would also like to understand the main factors that drive this particular customer behaviour, so that they can be more proactive and keep their customers satisfied.

Even though there are more examples where simpler models can be extremely powerful, recent advances in machine learning have laid in the state of the art deep learning techniques. Such approaches construct complex models which are often described as a 'black box' due to the challenge of interpretation. Such models obtain a much greater accuracy, but often at the cost of interpretability. Some predictive models, especially deep neural networks, are becoming more and more sophisticated as they are not only able to detect patterns (that a human would never recognise), but they are also able to learn from them - which has been a source of both controversy and excitement. This potential of prediction, to solve very complex and sophisticated problems, reflects a further delineation between data science and statistics: on one hand, these complex models can handle both linear and non-linear relationships, a useful advancement to traditionally used statistical models that have typically focused on linear relationships or simple polynomial extensions. On the other hand, while statistics respond to questions that are asked of them, some predictive models in machine learning provide us no other option other than to trust their predictions, even as we are becoming more and more reliant on their potentially opaque solutions. As Michael Tyka says: 'Have we really understood anything? Not really — but the network has... The knowledge gets baked into the network, rather than into us' [54]. Nonetheless, despite the power of each technique, it is crucial to always reflect on the nature of the problem and the research goal. Sometimes, the best machine learning model is not necessarily the most complex one - particular models, even simple Linear Regression (which is used in both data science and statistics), can be best suited for the (type, size and nature of) data and/or particular task. Simply said, operalization and considering model usage becomes key.

Measurement techniques

> "If you torture the data long enough, it will confess to anything."
>
> \- *Ronald Coase*

As discussed, data science and traditional statistics have slightly different research goals and have, therefore, adopted different ways to measure the success of their analyses. Data science techniques that deal with machine learning to classify, tend to use measures such as ROC curve or confusion matrix to evaluate the success of predictive models, whereas if the task is regression (predicting continuous variables), then the model is evaluated using mean squared error (MSE), mean absolute error (MAE), or perhaps a coefficient of determination [131]. However, one of the most prevalent criticisms that data science faces (especially from statistics) is related to 'data fishing'. Data science, and especially machine learning, have been associated with the derogatory notion that exhaustive search for patterns in the data will reveal some patterns that are simply a result of random fluctuations [129, 323]. In psychology, this issue, is also called 'p-hacking' or 'data butchering' and is recognized as a serious issue in the field. In data science, this issue has also been met with a lot of attention. Since there is always a probability that some patterns detected in data are the result of a pure chance, prevention measures have been rigorously introduced. There are hold-out testing regimens such as cross-validation (in addition to other remedial strategies, such as optimizing a penalized goodness-of-fit function or imposing tougher pattern selection criteria) [129]. These strategies have been put in place within data science in order to create more trustworthy, robust and generalizable models. However, it is essential to acknowledge that any model can be wrong to some extent - they are a simplification of reality, and looking for a perfect model can be an impossible mission [323].

An issue that traditional statistics face, conversely, is one of overfitting. Unlike with predictive modelling, in statistics, models are fitted to a sample, often in order to 'explain' the variance occurring within that sample [167]. Overfitting then occurs when a model fits the sample well, but it is so finely tuned that it cannot be replicated on another sample dataset or wider population. This issue is very important for the field of statistics, because the sample sizes that are involved are often too small to create held-out sets. In order to combat such issues, statistics famously introduced the notion of significance (p value), to reflect the statistical ‘significance’ of any result. If the p-value is larger than the alpha level (usually either 0.01 or 0.05), then the difference observed is explained as a result of sampling variability [285]. Yet, psychologists that use classical statistics are more traditionally interested in effect sizes or magnitude of the difference *between* groups. Rather than ensuring some tiny effect did not occur by random chances of sample selection, other metrics that can be of interest to classical statisticians are, beta coefficients, correlation coefficients and F and t values. If the associations between variables (or models) do not have statistical significance, they are often not mentioned or used in the further analysis. This is in contrast to machine learning approaches because, due to a large sample size, effect sizes are almost always statistically significant, and the main focus is on the prediction accuracy and generalizability.

This difference between metrics used in classical statistics and machine learning is important, because it is often the reason why it is hard to replicate study results across different disciplines. However, an even more challenging issue related to replicability is associated with the view that reporting p-values in statistics may not be enough. P-value are considered to be highly dependent on sample size and sometimes statistically significant results may imply that simply a large enough sample was used [285]. Without a good enough safeguard of what a p-value is expected to be, and in combination with the lack of data to test effects on a held-out set, resulting statistics can sometimes succumb under the pressure of replicating results in new studies. The replicability crisis in the field of psychology can be used to explain why some studies offer dissimilar and even contradictory results related to the same research questions. Generally, thanks to a much larger sample size, data science does not face similar problems, but rigorous testing of generalizability becomes of increased importance as a consequence.

Conclusion about quantitative approaches

It can be argued that machine learning can be seen as powerful extension to traditional statistics when hypothesis-driven research does not scale well and particularly when we have an overpowering size of data. More importantly to the social sciences, the nuances and intricacies of human behaviour can likely be observed best in the *uncontrolled* circumstances when people freely leave their digital traces without being explicitly mindful of potential monitoring. Despite the success of big data analytics, however, it is an overclaim that such approaches will lead to a completely new paradigm in the research [161]. This is due to the fact that big data analytics still needs contextualization in order to create a 'big picture' about the phenomenon being examined, and this can only be done by adding information about theory, policy, history and other relevant parts of the story [76]. Statistics and data science still share a common ground, both are focused on deciphering data structure and there is even a significant overlap between the two disciplines especially when it comes to exploratory analysis [323].

Nevertheless, even once we establish that each quantitative approach has a unique potential, it might still be argued that data science possesses one characteristic that traditional statistics could learn from. Data science, as a discipline, promotes an attitude of openness, unlike perhaps, statistics that has conventionally adopted an attitude of over-caution. Statisticians, especially as teachers, often emphasise extreme cautiousness with Brown et al. (2009) believing that this leads students to risk aversion, fears from flaws in the analysis and fears about data fishing [46]. This is why data science, even as a younger sibling of statistics, might still inform some positive changes in statistics as a discipline itself, without forcing unnatural alignment in terms of respective measurements or research goals. Yet, if the two approaches remain separate - so be it. They each have their own purpose and after all, a variety of methodological approaches will remain necessary as long as we have a variety of tasks to be fully examined.

2.3.3 Qualitative research and big data

Having discussed qualitative and quantitative research - including both traditional statistics and data science approach (or big data analytics), it is useful to also discuss similarities and differences between qualitative research and big data analytics despite them seemingly being two extremes on the continuum. In fact, big data analysis and qualitative research have some interesting similarities which are discussed in this section.

Firstly, we can acknowledge the unequivocal differences between the big data analytics and qualitative research, that include the paradigm differences, selected techniques and sample sizes. In fact, it could be argued that the appearance of big data creates new problems in the research, problems that did not exist in relation to qualitative and quantitative research previously. For instance, the growth of data generated problems in terms of volume, variety, velocity, value, and complexity, results in companies lacking appropriate technical capabilities to handle and responsibly analyse big data [156]. Insights based on 'big data' also lose momentum faster due to fast moving changes in technology and changes in data itself. However, there is one more serious criticism of big data, that makes data science and qualitative work incommensurable. This is the idea that due to the shared amount of data available about individuals, we start to observe people as nothing more than numbers, since through big data analysis we cannot approach their emotional states or perceptions [178]. Some of these claims have been the result of a prevailing discourse about 'big data value', suggesting that its analysis can often just increase the divide between already existing opposing quantitative and qualitative approaches [80].

However, even though they seem unrelated, big data analysis and qualitative research do have some similarities. Thanks to a diverse set of digital footprints, social scientists have never had such a level of granularity and variability in observing human behaviour, which, depending on the research objective, may create even more individual-focused insights than qualitative approaches. For example, a significant work conducted by Kosinski et al. (2012) states that social media can be a source of highly sensitive information such as religious and political views, sexual orientation, intelligence, happiness and others - all of which may be uncovered based on features such as number or likes or other properties of users' Facebook profile [164]. Similarly, Park et al. (2015) highlight that the language (of which there is an unprecedented amount of on social media) is in itself a powerful predictor of ones' psychological traits [222]. It is also relevant to mention that such big data (e.g. social media data or transactional data that represents the content of someone's shopping baskets), do not face traditionally mentioned challenges of interviews or surveys, while still being a rich source of insights about one's psychological state or behaviour.

With the generation of a large amount of text data on publicly available data sources, such as Twitter, new types of big data analysis were developed (e.g. sentiment, aspect analysis or topic modelling). They enable us to not only 'access' a large number of individuals' thoughts and emotional states, but also to do so in a scalable manner. Big data can contain clues about one's movements or trends, but, often only through a qualitative investigation of these sources are we able to create features useful in generalising insights. For example, a qualitative content analysis may identify patterns in big data in the form of texts, such as idioms, that can later be used in big data analytics to detect patterns across a much larger sample size. Therefore, one of the greatest potentials of collaboration between qualitative approach and big data lies in using qualitative insights based on a smaller sample of big data, and then scaling them up to produce cross-contextual understandings

[80]. These are the types of analyses that initially focus on breadth rather than depth. For example, 'data ethnography' is a method that has been developed to emphasise the explanatory power inherent to transactional data [274]. It is a reflection on an idea that big data can provide a valuable insight about 'shades of meaning' similarly to qualitative work. Simply said, just as it is possible to obtain an insight about a person based on an uncomfortable laugh during a qualitative interview, it is also possible to better understand a person based on how many friends they have on Facebook [236].

Lastly, a collaboration between big data and qualitative work can simply be on the level of combining insights in order to make connections. Through this kind of 'convergence analysis', increasing the diversity of samples may not claim generalisability, but may strengthen claims about social processes [80]. In summary, qualitative techniques and big data are not necessarily incommensurable. Combined approach may utilize the strengths of both and produce insights that would otherwise be unavailable or overlooked, something that is sought after, if not necessitated, in this analysis of perceptions and medication adherence.

2.4 Conclusion

> "Those who ignore Statistics are condemned to reinvent it."
>
> - *Bradley Efron*

After this involved discussion about methodological approaches, it is inevitable to raise the question: Which approach is better? The response to this question largely depends not only on the goal of the research, but also on the affiliations of the researchers, their philosophical stances, and the skills and backgrounds. This is why researchers from both sides of the spectrum continue to argue about the benefits of their approach - while statisticians claim that statistics have a more scientific and rigorous approach, other academics claim that predictions made by data science stand the test of time and replicability [156]. Similarly, while quantitative methods are looking at cause and effect, qualitative studies can be of better use when it comes to developing theories and hypotheses and explaining the deeper meeting of the particular event or phenomenon.

It is impossible to not notice how the world has become increasingly more quantifiable. This has brought with it a notion that novel approaches such as big data analytics may be not only more efficient, but might also produce more accurate insights, which in turn could mean that traditionally used qualitative and quantitative techniques become less relevant [80]. However, having discussed some of the major similarities and differences between methodological approaches, it can also be argued that this fear appears to be generally unfounded. Statistical literacy, as well as application of traditionally used qualitative techniques such as interviews and focus groups, not only have a valuable place in research, but also continuously inform new methodological developments. Moreover, it can also be argued that these older approaches to research can offers a strong basis to new hybrids of methodologies, which may increasingly develop due to new oportunities in terms of data sizes and structures. Each approach has unique characteristics, however, there remain many (sometimes unexpected) similarities between them. Therefore, in the spirit of mixed methods, it is beneficial not only to focus on how these approaches may feed into, complement and support each other via new triangulations, but also how their unique traits could contribute to a partnered knowledge discovery. Based on the nature of the topic of perceptions and adherence, which should be investigated through various lenses, gaps in previous literature and opportunities recognised in methodology, a combination of different approaches was deemed as the optimal choice in the context of this research.

Chapter 3

Small talk, Big data: Patients' and general public's perceptions about people with asthma

> "It started with a simple problem... A key with no lock. And I designed a system I thought fit the problem. I broke everything down in the smallest parts and tried to think of each person as a number... In a gigantic equation. But it wasn't working... Because people aren't like numbers. They're more like letters... And those letters want to become stories."
> - *excerpt from the film 'Extremely loud and incredibly close'*

As stated in the previous sections, the topics of asthma medication adherence and perceptions about asthma are complex topics that should be investigated though a triangulation of various approaches. This work, therefore, started with the study that aimed to leverage qualitative research in combination with existing large data sets ('big data'), to create a better understanding of perceptions of asthma. It can, therefore, be argued that the first study of this PhD is exploratory, as it aimed to respond to the following research questions: *what perceptions do people have about asthma; how are they created in the light of interactions patients have with other people; and in which ways are patients' perceptions and perceptions captured on Twitter similar and/or different.* There are two parts of this study and the first one was focused exclusively on experiences of interviewed asthma patients, whereas in the second part of this study (see Section 3.4) this work was expanded with the views from publicly available Twitter data.

Semi-structured interviews were conducted and analysed using thematic analysis as a first part of this work. In-depth interviews were used in order to enable more context – sensitive interpretation of perceptions of asthma [48]. However, traditional techniques such as interviews, face some challenges, such as desirability bias [146]. Additionally, with the increasing popularity of social media, today it is possible to harness interactive platforms in order to capture unprovoked opinions and beliefs about many different topics, including asthma. Therefore, a coding schema was developed to capture perceptions on Twitter and a content analysis was conducted on 3.000 tweets. In order to compare patients' and public's opinions, convergence analysis identified similar topics between interviews and social media data and where these were similar or different. The potential of this research rests on valuable implications for clinicians and patients, as this work can provide not only a better understanding of asthma, but also appoint us to untapped potential of social media.

3.1 Background

As previously mentioned, asthma adherence can be strongly impacted by patients' perceptions, given that perceptions can potentially result in emotional and psychological states that in turn affect behaviours [34]. Creating insights about what people think of life with asthma can inform current self-management education approaches [49]. For example, some studies show modifying patients' perceptions such as a fear of long-term effects, or a fear of social stigma can be effective in improving adherence to asthma medication [188]. However, very few studies focused on asthma patients' and non-patients' perceptions about asthma, the similarities between them and a resulting, potential influence on adherence.

There are many potential reasons for low adherence, however research indicates socio-cognitive factors including perceptions have a crucial role [49]. It is theorised that when a person is faced with a chronic health condition, they create an opinion about how the diagnosis will impact their life [137, 192], while they are facing practical and psychological limitations caused by their condition [83, 49]. People are then thought to make adjustments in terms of behaviours and thinking processes [83], which may influence their adherence to medication prescribed [254]. Understanding patients' perceptions may, therefore, help health practitioners understand how to promote adherence to their patients' treatments [194]. Additionally, knowing which negative perceptions to tackle may help in developing more focused interventions to approach and help high risk, non-adherent patients, and educate the general public [206].

Perhaps the most comprehensive models used to investigate perceptions related to patient behaviour are the previously mentioned Self-Regulatory (SR) [173, 44] and Health-Belief model (HBM) [117, 212]. As described in the literature review, Section 2.1.7, both models propose that health threat or symptoms result in creation of cognitive or emotional representation of a health condition [44]. However, these models do not address the importance of perceptions that exist outside of patients' control, such as non-patients perceptions of asthma, which may also not only inform patients' illness representation, but also affect patients' behaviours in front of other people and treatment adherence, in general.

The majority of perceptions that were previously researched, only relate to patients' personal views of themselves and their condition, but not necessarily in the context of their social surroundings. The role of the patients' social surrounding has been touched on in the context of adherence to asthma medication, using the Socio-ecological models, which is described in literature review, Section 2.1.8. As mentioned, this work insufficiently focuses on both patients and non-patients perceptions such as experienced and perceived stigma. However, these factors are important given that patients are inevitably intertwined with their social environment (from which they also learn) and change their behaviour depending on the context they are in [21]. As mentioned in Section 2.1.10, negative public portrayal of asthma and patients can be damaging when patients cannot reconcile the identity of 'an asthmatic' with the remaining characteristics of their social identity [5]. In this case, it is possible that patients experience stigma and as a result, deny having asthma and stop using medication [5]. This is why it is relevant to identify the nature of stigma in patients' environment and how pervasive it is, in order to combat this with information campaigns [79, 47].

3.1.1 Negative perceptions

For people with asthma, as for many other patients' with various health conditions, social environment is especially relevant. Firstly, perceived social support can significantly help during times of social and emotional difficulties [316]. Secondly, negative perceptions in the form of stigmatization, could create psychological barriers to adherence to asthma preventer medication [108]. Stigma has previously been mentioned as a damaging factor in Section 2.1.10. As previously mentioned, stigma can be framed as *self-perceived* (subjective experience of stigma), *enacted* (related to discrimination against patients) and *internalized* stigma (level of patients' agreement with social stigma) [7, 258]. Enacted stigma was found to be less prevalent than perceived and internalized stigma that may lead to feelings of isolation and deleterious consequences of low self esteem [286]. However, regardless of the type, the heart of stigma lies in social norms and how they are perceived and enacted [225]. In fact, Masuch et al. (2019) conducted a study about stigma associated with ADHD and claimed that patients face public antagonism to the condition and that it is the integration of these negative experiences in the definition of identity that leads to internalized stigma and self-devaluation [198]. Such comparison between public views and patients' beliefs in the case of asthma are scarce.

3.1.2 Coping mechanisms and Humour

Some perceptions of asthma found in Chapter 3 were related to patients' coping mechanisms. These mechanisms do not only represent an adjustment in terms of behavioural patterns and coping with the limitations that asthma imposes - they are also followed by a change of perspective and a creation of a belief system [192]. Coping styles are, therefore, another factor that may impacts adherence [192, 5]. In this sense, coping efforts could roughly be grouped into: active, problem-focused and adaptive; passive, emotion-focused, related to negation of the medical condition; denial, underestimation; or even exaggeration [217]. Coping strategies have also been described as appropriate and active, such as information seeking or positive reappraisal [192, 70] and negative, such as hiding, ignoring and worry about asthma [217].

Humour or denial, can be used as examples of coping strategies that are a result of perception creation between people [193]. Specifically, a particular form of humour, disparagement humour, is often associated with negative social consequences [108]. This is important for patients given that disparagement humour may create increased tolerance for discrimination [108]. This is also important for patients since this humour leads to a reduction of the perceived importance of serious health conditions [2]. However, the effects of disparaging humour on individuals' perceptions of asthma, as well as the themes this humour uses for its basis and its functions, in the context of asthma remain largely unexplored [169].

One of the potential reasons why patients' and public perceptions are not given a lot of attention to could be the difficulties that arise as a result of examining perceptions. There are, for example, research biases that may arise when using the more traditional research settings, given that this may stop people from expressing true opinions [146]. Additionally, obtaining data on public perceptions of asthma has previously been a difficult task due to expensive and timely sample collection. However, novel data sources, such as social media are a rich source of people's opinions and even insights about their personalities [235].

3.1.3 Novel data sources and opportunities

Previously, obtaining data on perceptions about asthma to inform adherence programs, was extremely difficult in terms of collection and scale. However, novel data sources, offer new opportunities. Twitter, for example, is a rich source of public opinions about health conditions [221, 23], which is described in Section 2.3.2. Twitter provides a large amount of data on public perceptions, where people are thought to express themselves more openly [221]. These information can be used to obtain a better understanding of a public opinion about a specific health condition [23]. However, the number of studies that use Twitter, in the context of examining topics related to asthma, are very scarce and to our knowledge, tweets have not previously been used to examine perceptions about asthma.

In conclusion, previous work has examined perceptions, but mostly perceptions patients have about their condition and treatment, and usually not in the context of perceptions of non-patients. Studies that examine the role of public perceptions about asthma are scarce, more traditional research approaches are usually used and work around stigmatization has some contradictory findings. To our knowledge, no studies compared patients and public perceptions about asthma for potential congruence. Examining this relationship is important especially in the context of negative perceptions as the nature of these perceptions could be used as a direction for future communication strategies. Based on all of this it can be argued there is a research gap in the field of adherence to asthma medication. This gap is important because, currently, much advice on asthma-management strategies are designed without taking into consideration patients' experiences (which may partly explain a low uptake of these measures) [15]. Therefore, a more detailed observation of patients' and public perceptions about asthma is needed, especially in the context of their social environment.

3.2 Current research

This study offers an overview of patients and public perceptions by combining interviews and content analysis of Twitter data in order to detected similarities and differences in perceptions of asthma that arise on these different sources. Therefore, the aim of this work was to explore what patients assume others think of them in the light of their condition and compare that with what others *actually* think. Qualitative interviews were conducted in Study 1a and only asthma patients participated. Twitter data was used in Study 1b and this enabled the identification of perceptions held by both patients and non-patients.

Qualitative research deemed appropriate for these goals since it was crucial to identify the nuances of subjective understanding that motivate various perception creation and elicit the discovery of new dimensions of a phenomenon [98]. This is especially significant as people develop their beliefs and attitudes with special regard to the socio-cultural context. The relevance of this is reflected in the hypothesis that public, negative connotations about asthma may influence how patients see themselves and their treatment [225]. Therefore, we discuss their similarities and discuss about potential effects patients and non-patients perceptions about asthma could have on patients' adherence to medication.

3.3 Study 1a: Examining patients perceptions about life with asthma through interviews

3.3.1 Design

Semi-structured interviews were conducted face to face. Semi-structured interviews were deemed appropriate because there were key topics that were of the particular interest [48]. The first part of the interview had a set of questions that helped in building the rapport with participants, where patients were asked about their work and general views about asthma. This was followed by questions about patients' personal experiences, feelings and beliefs related to asthma and inhaler use. These interviews were informal in nature and follow-up questions were aimed at investigating participants' opinions relevant to the topic, but they were also designed to enhance exploratory nature of the research and potentially generate new knowledge [309]. Additional details about the survey and the interview protocol are presented in the Appendix, Section C.

3.3.2 Participants

In total, 14 participants were interviewed. Five were male and nine were female. The age range of participants was between 19 and 54. The main criterion for participant recruitment was the medical diagnosis of asthma, at any point in their life. Five participants had severe asthma and nine had mild asthma. All participants' names were replaced with pseudonyms in the reported quotes.

3.3.3 Materials

Interview questions were semi-structured. Initial questions were related to age, occupation, the length of patients' asthma treatment and general questions about asthma and adherence to asthma medication (e.g. *'What do you do to keep your asthma in check?'*). The following section contains questions about general and personal experiences related to asthma. The study aimed to examine perceptions through examples from everyday life and work [5], so participants were asked questions such as: *'What kind of jobs would people with asthma not be able to do and why?'* and *'If you were a Chief Executive Officer (CEO) of a company, would you disclose to employees that you have asthma?'*. These questions were related to potential stigmatization and disparaging humour about asthma [192, 15]. Survey questions were designed to elicit participants' personal opinions (e.g. *'What is the first thing that comes to your mind when I say asthma?'*), but also their ideas about how others perceive them (e.g. *'How do you think others would react if they knew you have asthma?'*).

3.3.4 Procedure

Participants were recruited using personal contacts and a snowball technique. The study was also advertised to undergraduate Psychology students, with completion in exchange for course credit. Interviews occurred face to face between January and March 2019. Before the interviews, participants were given a short description of the study summarizing that it was examining perceptions they have about life with asthma. Participants were given a consent form and data privacy information they could consult with and agree to. Interviews were audio recorded using a mobile device and transcribed by the researcher. The average length of the interview was 50 minutes (the shortest interview was 40 minutes and the longest was 1 hour and 20 minutes).

3.3.5 Results and Discussion

This study identified common themes within patients' descriptions of their experiences with asthma. Analysis was atheoretical in order to allow the interpretation of unexplored, emerging themes [218]. Data obtained was coded independently by two researchers who then met to discuss arising themes. Where disagreement arose, researchers discussed until agreement was reached. Once the initial coding was finished, a further independent researcher replicated the coding process to check reliability using the codes that were developed. Topics identified comprised of: *internal perceptions*, including perceptions patients have about asthma, self-image, asthma management, impact of asthma on everyday life, attachment to inhaler and *external perceptions* (perceptions related to other people) that included assumed perceptions participants believe others have about them, other people's reactions to participants' asthma, and humour (see Table 3.1).

Themes	**Description and boundaries**	**Example**
Perceptions about asthma	Descriptions about how the participants see asthma. This included perceived prevalence of asthma and that asthma is 'a part of them'.	Jack: "Asthma is quite common. Everyone knows someone who has asthma." Shruti: "My mum always had a say in a way I perceived my disability." Jean: "I think it's because it's one of these invisible things you have, people only notice when start coughing."
Self-image	These are comments related to the way participants view themselves in the light of asthma, how they adjust to others' expectations. These include: disassociation, self-blame, self-pity and feeling strong.	Helen: "I think it's a more fragile, sick person that has asthma, whereas Helen of recent years without her inhaler has been a lot healthier and stronger. And I have this conception that I grew out of it as I no longer want to be that person." Amanda: "I'd probably remove myself, rather than say: 'can you not [smoke]?'... Because as far as they know they're not affecting anyone, they're not doing it on purpose. It's my fault that it's a trigger..."
Managing	Ways in which participants manage their condition: includes practical management; smoking and psychological management: departmentalisation, rationalisation and avoidance.	Jack: "I don't think smoking affects my asthma as much as it could. I started smoking 10 years ago, so my asthma had gone away after I started smoking. It's a bit weird that happened, isn't it?" Jean: "To me it's unpleasant, so I'd just rather forget it's there sometimes."
Impact	Comments describing thinking processes and feelings that emerged as a result of adjustments to living with asthma and limitations (physical and psychological) it imposes.	Camilla: "I guess it might affect in terms of what they keep in their house, in terms of pets, having plants, or anything that might trigger... Where they can go, where they spend time. Maybe physical activity, what they're able to do. Maybe possibly a job."
Attachment to inhaler	Feelings about the inhaler	Shruti: "I need to have it with me. I think it's psychological because I feel handicapped without it, even if I'm fine. It's a dependency [laughs]..."
Others' assumed perceptions	Assumed opinions of other people in case they know/find out participant has asthma. These include: stigma reflected in media impact, lack of understanding, impact on participants' job.	Amanda: "People would probably think you're more fragile, unfit, I think that's a main thing about how people are portrayed – lazy and unfit." Helen: "Well, especially in movies and stuff... It's always this dorky, fellow with the glasses usually and the braces, who walks around with his inhaler... So you feel kind of like that, you feel dorky. You feel a bit like you're... sick or something." Shruti: "I think there are actually a few jobs that people who have asthma couldn't do... I think it would be really weird to find out that the president has asthma."
Others' reactions	Comments that describe what others say or do if/when participants decide to disclose they have asthma. Includes the sense of community with other asthmatics.	Zoe: "I remember when I had quite a lot of friends who smoked, I'd say: 'Come on guys, can we just go to a non-smoking area', they'd be like: 'No, get over it', 'But I can't breathe!', and I felt like the outcast. They'd be like: 'No, you deal with it, not the other way around'." Amanda: "They (other asthmatics) just have more understanding of what you can and can't do."
Humour	People using humour when discussing asthma and ways in which this is done. Exaggeration of condition and use in interactions.	Nicole: "'Great, now I can't breathe either, f*** you asthma.' It's more rant-y, trying to make it grumpy, rather than a real talk."

External perceptions

Stigma

Rather than a single topic, stigma was the underlying link between several subtopics detected in interviews, including self-blame, others' perceptions that mentioned the role of media, jobs that asthma patients could not do, the issue of disclosure, others' reactions in the form of discrimination and negative humour. However, when asked directly, very few participants said that there is any asthma-related stigma present in their lives.

> *No, nobody that had asthma, they were never treated [differently]... In fact, they were the most popular kinds at school. (Tom)*

Nevertheless, stigma was frequently described in examples about how participants felt or imagined how they *would* be treated by non-patients. Through the interviews some participants assumed that others see them as 'weak' and 'fragile' and called their own condition 'a disability'. These examples highlight the underlying notion that asthma makes people different from the norm and and that there is 'something wrong with them':

> *I think when you see someone with asthma, you're more likely to think about the bad end of it. So if you say you have asthma, you're put into this 'unhealthy-something-wrong-with-you' box. (Nicole)*

> *If the CEO is giving a talk and then half way through used his pump, probably the traitors would be like: 'Oh my God, what's wrong with him?' (Zoe)*

Some participants even mentioned that they, themselves, see other asthma patients as 'weak', which could be a signal of internalised stigma. Enacted (experienced) and perceived stigma, previously reported in the literature, could also be identified [277]. Perceived stigma detected in interviews refers to opinions that patients assume other people have about them. According to participants, these beliefs are developed from portrayal of people with asthma in the media. A consistent finding was the perception that media creates a distorted image about asthma in TV shows, films and other media sources:

> *I think the only clear image I have about that is - in movies it's always the dorky, person with the braces who's with his inhaler. It's never the good-looking boy or girl who has asthma. (Hannah)*

Stigma characterised here is in line with the previous research that claims 17% of US films with asthma scenes present characters with asthma as 'wimps' and 'outcasts' [15]. However, our participants additionally claim that there are other ways media perpetuates stigma. For example, media often indicates that in stressful situations people with asthma are not able to cope well. Participants, therefore, indicated that media portrayal can reinforce misunderstandings about asthma, especially when they are portrayed as 'nerdy' and 'not being able to cope with stress', which leads to embarrassment for people with asthma.

> *And I think the perception is then: 'asthmatics will panic if they can't put their hands on their inhaler'. And they just make it worse... I know they always portray the worst-case scenario. Asthmatics will be on their knees, wheezing, gasping, totally blue, needing oxygen, whatever. (Jean)*

To expand on this idea, some participants claim this negative portrayal is exactly the reason why people with asthma rarely disclose their condition or discuss asthma. As a result, non-patients remain less knowledgeable, creating a vicious circle of lack of understanding.

> *I think people mostly don't understand asthma. Asthma doesn't come up much at all. Anywhere. So there's a lot of ignorance about asthma. (Amanda)*

Participants claim that this perceived lack of knowledge leads to further feelings of stigmatisation in the patients' group:

> *If someone didn't know so much about asthma and they meet someone who has asthma, they'd probably think like 'just go to the gym and you'll be fine'. Other times people who are less understanding would probably be like 'you're overreacting, just breathe'. (Tom)*

Enacted stigma was mentioned in situations when participants openly used their inhaler, were seen smoking, or working out in a gym:

> *I had people come up to me and say: 'I can't believe you smoke and you have asthma.' (Shruti)*

> *If I knew I would have some form of asthma related problem, then why did I push myself so hard as to get myself to that point. I feel like people would blame me. I feel like there would be a lot of judgement. (Helen)*

Types of stigma expressed in interviews signify the notion that stigma creates a separation between patients and non-patients, given that even when they are not directly discriminated against, patients may still *feel* marginalized. This may explain why perceived stigma was mentioned much more frequently than enacted stigma - apart from inhaler use, there are no indications that one has asthma, which tends to make asthma 'hidden'. This may, therefore, reduce potential discriminatory behaviours such as labelling, stereotyping and discrimination [7]. However, the legacy of negative portrayal and experienced discrimination may leave behind more permanent anticipated stigma and fear of judgements.

Disclosure

Many participants confided they would not disclose they have asthma even in their friendship groups and especially not at their workplace, due to a fear of judgement and potential perceptions that they would not be able to do their job as well as non-patients:

> *If I were a fireman I wouldn't go around telling all my colleagues I had asthma. If someone knew he had a defect it would stop him from doing his job. They wouldn't so much trust in him. If you have asthma, or some other disability, you don't really want to go flashing it about, because you want to be seen as someone you can rely on, count on. (Jack)*

> *In that position [a CEO] I would be so high up, it's a weakness in that position to let your employees know about asthma. (Shruti)*

The idea that other people would think one is 'unreliable' at their job because of asthma was often mentioned through interviews. This illustration of internalised stigma also appeared when participants spoke about professions that require 'strong and influential people', such as a president or a CEO. Some participants stated that these professions represent a symbol of power, as opposed to asthma, which is seen as a 'weakness'. As a result, one's identity as a 'powerful CEO' is not compatible with the 'asthmatics' identity. Therefore, the latter, according to some participants, should be a concealable identity and should not be disclosed [237].

Closely related to disclosure is the act of avoiding the use of the inhaler in public. Both of these topics related to hiding asthma with the latter having more negative effects on one's treatment adherence. This was illustrated by one interviewee:

> *...Self esteem wise it does affect you, and not taking your inhaler in public turns into not taking your inhaler - at all. (Helen)*

Based on this, it can be argued that non-adherence may be an indirect result of negative perceptions about asthma and patients attempt to conceal their 'asthmatic identity'.

Internal perceptions

Internal perceptions relate to how participants see themselves and their condition. Some participants who claimed there was no stigma present in their lives still used words such as 'defect' or a 'disability' to describe asthma. This choice of words emphasises negative perceptions patients have, not only about asthma, but also about themselves as they incorporate asthma in their lives. This was illustrated by one participant:

> *You're seen weaker because, in a way, you slightly are. (Zoe)*

Another internal perception that is relevant is self-blame. Self-blame could be interpreted as internalized stigma that was related to negative interactions with non-sufferers:

> *Lots of my housemates last year smoked and I asked them many times not to smoke in the house because of my asthma. But they often did. Maybe it was my fault for not saying it enough. (Amanda)*

Conversely, some participants reported positive perceptions such as perceived personal strength that is a result of living with their condition, which could potentially be interpreted as a coping mechanism. This was similar to humour, which was mentioned as a positive way of coping:

> *'I'd start coughing and say: 'Sorry I can't die quite as quietly.' If anything, it helps me as well. I like to make it as a joke.' (Jean)*

As opposed to positive coping, self-pity, as a negative self-perception, was used to illustrate how limited participants felt, not only due to physical, but also psychological boundaries:

> *You have to manage this illness. You might not be able to do much physical activity, but also, just the idea... It's an extra thing people have to deal with, are afflicted with or makes them lesser. (Nicole)*

Internal perceptions discussed could be observed as a part of one's self-concept, a composite view of oneself, which is continuously formed and re-formed through interactions with significant others [49]. The idea of a self-concept is relevant in this example, because positive or negative self-view could affect coping mechanisms and health-related behaviours. These results are also important as they highlight not only the existence of internalized stigma, but also ways in which it reflects on patients' perceptions about themselves.

3.3.6 Interim Discussion

This study builds on previous research that has identified perceptions people with asthma have, particularly related to stigmatization as a group of negative perceptions [7, 267, 151]. However, this work extends previous research by identifying further ways in which stigma manifests for people with asthma - by separating them as internal and external. Overall, external perceptions included anticipated and felt stigmatization, including themes: fear of disclosure, lack of understanding, others' perceptions and others' reactions. Internal perceptions were related to how participants see asthma and themselves in the light of asthma and these perceptions were reflected in topics such as self-image, management and impact. In this group of perceptions, participants mentioned departmentalisation and rationalisation as coping strategies to manage asthma that is perceived as limiting and unfair. Other internal perceptions were related to self-image, including the sense of empowerment though asthma, but also negative self-perceptions such as self-blame and self-pity. Stigma could be used to group several themes and it was identified as enacted, perceived and experienced stigma [7, 267, 151]. Perceived stigma was the most frequently mentioned type of stigma and it was described though hypothetical situations and assumed perceptions of others. For example, participants would choose not to disclose they have asthma at their workplace because of anticipated judgements. Internalised stigma was observed when participants spoke about other patients, because they were described as 'weaker' than non-patients.

These perceptions could have an impact on one's self-concept and adherence. Many of these perceptions were a direct or indirect result of patients' views about how society perceives them as asthma patients. However, this work only includes a small sample and views of asthma patients about asthma and stigma that may have prompted demand characteristics in how people responded. This is why this study was extended to examining more naturalistic data from tweets, in order to avoid desirability bias, include the voices of non-patients and to investigate how these findings manifest across a larger population. The second study is, therefore, focused on publicly available data, which was not created in a laboratory setting, and is created by both patients and non-patients. These public perceptions were them compared and contrasted with patients' perceptions from interviews.

3.4 Study 1b: Examining public perceptions about asthma based on content analysis of tweets

The goal of study 1b was to use a new frontier to assess public views on asthma, created in a less constrained real world situation, devoid from researchers' and desirability biases. It was considered that studying Twitter data may unearth previously unidentified perceptions, not necessarily consciously expressed, but evident in more naturalistic data, and widen the context by including the voice of non-sufferers. Ultimately, perceptions on Twitter were compared to perceptions expressed by patients in Study 1a, with the limitation that these two data set had different methods of data collection and different participants. Particular attention is given to where interviews and Twitter data converge or diverge.

It is, however, relevant to mention that Twitter itself has some limitations as users may choose to present particular versions of themselves for their online public, which may affect what and how they tweet about asthma [119]. In fact, some research claims that there can be an image-related utility of social media, which assumes that others' perceptions of themselves motivates users in terms of what content they produce [290]. Some authors even mention the discussion of whether our online self is our 'true self' or just another self - an act that is created for the context of social media [322]. However, even taking these characteristics into account, social media has proved to be able to provide insights about many population characteristics, such as personality traits [100, 224]. Before addressing the study design, it is useful to reflect on some of the ethical considerations with regards to Twitter data analysis.

3.4.1 Ethical considerations

Even though the use of Twitter data in academic research is becoming increasingly popular, there are still many partially unresolved ethical considerations, which directly impact conducted methodology. One of the main questions from this work was related to a discussion whether or not Twitter data is private or public. Twitter, as a social media tool, has binary options when it comes to privacy: accounts are either 'public' and, therefore, visible to even non-registered visitors, or 'private' (in which case they are visible only to the followers of the individual). Twitter's terms of service specifically state that users' posts that are public will be made available to third parties, and by accepting these terms users legally consent to this (Privacy Policy, 2018). Moreover, there is a previously established consensus that internet data that are freely and publicly accessible can be used for research [82]. However, this study will take into considerations a wider number of ethical challenges, even if we consider Twitter as a public channel. The first challenge is the possibility to identify an individual even when special measures, such as anonymization, take place.

Public or private?

Some authors claim that there is a need for a more strict approach than provided in 'legal' accounts of the permissible use of Twitter data [312]. This is specifically related to challenges that arise if we consider Twitter as a public channel. One of the main challenges is the possibility to identify an individual even when special measures, such as anonymization, take place. In the US, personally identifiable information is usually defined as individual's identifiable elements: their name, social security number, driver's licence or a credit card number [324]. EU law, on the other hand, has a more broad definition of one's identity and includes physical, physiological, mental, economic, cultural or social identity. This poses the idea that there is a potential threat of identification of individuals even when their anonymized text is in a set of thousands of tweets.

However, the legal basis for processing Twitter data is ‘public task’. In other words, according to ‘Lawfulness of processing’ chapter from Data Protection Act, Twitter collection operates under: ‘processing (that) is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller’, according to the Data Protection Act from 2018. In addition, Art. 85 GDPR (“Processing and freedom of expression and information”) states that ‘Member States shall by law reconcile the right to the protection of personal data pursuant to this Regulation with the right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression. Since processing data in this study is for academic purposes, it aligns with the following exemption: ‘For processing carried out for journalistic purposes or the purpose of academic artistic or literary expression, Member States shall provide for exemptions or derogations from Chapters: 2, 3, 4, 5, 6, 7, 9.’

Special category

Another ethical challenge that needs to be considered is related to a nature of data processed. Since this study is likely to capture some health data relating to individuals, this means that this data can be considered as ‘special category’ data. In addition, sensitive information is not only the information created by users, it can also come to light once the analysis is done and when the associations are identified between participants and their personal characteristics, by using algorithms [297]. Since it is already established that Twitter data is made public by the data subject, the conditions for special categories data apply: ‘processing relates to personal data manifestly made public by the data subject’ ("Data Protection Act

2018", 2018).

In relation to informed consent, it is stated that participants should always be clearly informed if their interactions are analysed and observed [312]. However, this information is *not obligatory* in case of certain exceptions or exemptions. This study represents one of these exceptions. According to the Data Protection Act (2018): The following provisions do not apply to personal data processed for — a. scientific or historical research purposes, or b. statistical purposes, to the extent that the application of those provisions would prevent or seriously impair the achievement of the purposes in question ('Data Protection Act 2018', 2018). Therefore, even though the data belong to the special category data, in this case, the processing is allowed.

It is also useful to mention that previously conducted survey shows a general lack of concern from users over their posts being used for research purposes, with university research that stated as attracting the least concern [312]. Lastly, analysis conducted in this study could be classified as passive analysis, that is, analysis of information patterns and interactions on discussion groups of which researchers have not been a part of. Having taken into account these ethical considerations, it can be concluded that the analysis of Twitter data conduced in this work can be labeled as of low intrusiveness [99].

As an additional measure, this work, including the work with Twitter data from other studies in this thesis, was reviewed and approved by an independent ethics committee. All the data collected using Twitter API was stored on the encrypted, password-protected server at the University of Nottingham. Access to this data was restricted to include only those who are directly involved with the research and this procedure was also explained to participants.

3.4.2 Design

Tweets that contained asthma keywords were collected, processed and analysed. Natural language processing techniques and descriptive analysis were used to obtain the most frequent topics discussed in relation to asthma on Twitter. These topics were then used in a convergent analysis, where social media data was combined with data from Study 1a. Social media data has been explored in various concepts and with different goals, however, there have been only several studies that try to converge the social media data with other data sources, with the attempt to use social media to support previous research and better understand individuals [120]. Similarly to the previous research, this study had a challenge – combining: 1) data from a qualitative research where perceptions are expressed in the form of a narrative with 2) tweets, which are limited to 140 characters and represent disjointed thoughts from various individuals. In order to compare this data, this research adopted a bottom-up approach with the goal of matching perceptions from the qualitative study with tweets that illustrate the corresponding perceptions, and identifying where perceptions did not match.

3.4.3 Participants

Since only less than 10% of Twitter users restrict access to their accounts, much of the previous work has recognised that Twitter can bring a better understanding of its users and impact on the society [196]. In terms of age restrictions, participants must certify they are at least 13 years old in order to use Twitter [259]. Most common age groups on Twitter are between 18 and 29 years old, whilst both men and women use Twitter in a similar proportion [210]. Lastly, it is relevant to highlight that tweets collected in this study were created by people who may or may not have asthma.

Materials

Tweets were streamed via Twitter's application program interface (API) for 6 months in 2018 and 2019. Considering that tweets are short bodies of text and contain casual language, abbreviations, acronyms and emoticons, it was important to make a clear choice of key words [113]. The keywords used in this study were: 'asthma', 'inhaler' and 'asthmatic', words that best reflect the topic of asthma on Twitter [239]. Tweets were processed using Python, the programming language and Structured Query Language (SQL) [35, 121].

3.4.4 Procedure

Data collection

Only tweets in English language were taken into consideration and there was no limitation in terms of tweet location. Over a million tweets were collected. 376.893 unique tweets remained after removing re-tweets. It was not possible to distinguish between people who do and do not have asthma based on their tweets, therefore, it is assumed that this study contains tweets from both groups. In addition, it is possible that some Twitter users wrote more than 1 tweet and some tweets could be coded as multiple themes.

Data Pre-processing

Since Twitter has grown significantly, it is now analysed by many organisations and services that aim to obtain information from text, so there are some standard techniques of text analysis that were applied in this study as well [149]. Natural language processing techniques were used to clean the data: every tweet was anonymized and only the full body of tweets' text was saved. All other properties of tweets: location, timestamp, username, short text, URL links (e.g. http://example.com), Twitter user names and mentions of other users were removed.

Manual Feature engineering

In order to investigate themes within tweets, a three-stage approach to analysis was undertaken (c.f [221, 273, 60]). Firstly, a random subset of 1.000 tweets was selected and tweets were manually coded into emerging themes. In the second stage, linguistic features that were common within a thematic category were established by reading tweets. Typical features identified were key words, phrases, common internet expressions and emoticons. For example, the theme labelled as 'lack of understanding' frequently contained the idiom 'just breathe' (to ease asthma symptoms) (indicating that the user does not understand how asthma is managed). Extracting common features from each theme was carried out in order to create rules for classifying additional tweets from the overall data set. Identifying features for each theme were developed into a code book (see Table 3.2) that contains discovered themes.

In a further analytical stage, additional tweets were obtained and allocated to themes within the code book in Table 3.2. This was done using Structured Query Language (SQL) techniques that relied on rules from the code book [121]. After these additional tweets were collected, there were overall 3.000 tweets that were all, each independently, coded by five more researchers who were first trained by the lead researcher in identifying themes using the code book (c.f. [60, 305]). Coding reliability between six coders was calculated using Cohen Kappa and the score obtained was 0.79, which was deemed as satisfactory. Analysed tweets were placed in one of eight themes, where six converged with themes identified in Chapter 3 and two were new themes.

3.4.5 Results and Discussion

Studies that involve Twitter data in the context of asthma are scarce even though it was stated that Twitter data can significantly contribute to obtaining a deeper understanding of public health [268]. However, this study offers a new perspective on public opinions about asthma by conducting a first content analysis on tweets related to asthma. Emergent topics in discussions about asthma on Twitter were: self-pity, lack of understanding, disclosure, sense of community, negative humour, attachment to inhaler, research and cost of inhalers (see Table 3.2).

Themes	**Key features (containing the words used to identify tweets**
Self-pity	'😫' or '😩' or '😭' and 'asthma' or 'inhaler'
	#f***asthma
	'sick of having asthma'
	'fml'
Lack of understanding	'just breathe'
	'asthma isn't real'
	'pretend illness'
	'can't f****** breathe'
	'believe me'
	'serious' and 'asthma'
	'asthma is a choice'
Disclosure	'I' and 'have' and 'asthma'
	'I' and 'got' and 'asthma'
Sense of community	'I got asthma too'
	'I have asthma too'
	'asthma squad'
	'fellow asthmatics'
Humour	'😂'
	'haha'
	'die'
	'lol'
Attachment	'I' and 'need' and 'inhaler'
	'where' and 'inhaler'
	'😭' and 'inhaler'
	'can't find' and 'inhaler'
Research	'research'
	'study'
Cost of inhalers	'cost'
	'price'
	'pay' and 'inhaler'

Table 3.2: Code book of manual features

Humour

The most dominant topic in tweets was humour. Notably, some features used to identify humour (e.g.‘🙄’) were also used in other themes, such as sense of community, which indicates some tweets simultaneously contain several topics. Often these tweets contain jokes that could be classified as stigmatization given that asthma patients were either objects of jokes or authors of humorous ‘confessions’ relared to asthma. Therefore, in many cases, this kind of humour is disparaging and contains derogatory connotations about both asthma and asthma patients, such as the following example [108]:

> *Calm down Lil Weezy*

In other cases, users use humour as a reference to something surprising or funny that happened and consequently caused asthma symptoms, or in a metaphoric way using the idea that asthma symptoms could have been experienced to emphasise something extreme.

> *IM GONNA HAVE AN ASTHMA ATTACK OMG* 😱😱😱😂😂😂😂 *cuz she’s actually gorgeous*

Self-Pity

This topic can, in the majority of cases, be assumed to be created by asthma patients, since tweets are mainly complaints about one’s asthma symptoms or feelings induced by asthma. Self-pity often included the sense of urgency, which might be interpreted as an help-seeking technique, since these tweets often elicit responses and comforting messages from others. A number of tweets from this category also described feelings of sadness and unfairness of having to deal with asthma. The following tweet illustrates that.

> *I'm supposed to be enjoying the night out with friends but here i am suffering from asthma attack* 😫😢😷

Lack of understanding

This theme represents a lack of medical understanding of asthma and its symptoms. These tweets are usually derogatory in tone and often contain judgements about patients:

> *I don't get it. There is so much air how can u get asthma*

> *Who tf still has Asthma ? That childish ass disease. Just breathe* 🙄

Since tweets are directed at patients, it can be assumed that they are primarily written by non-patients and are examples of discriminatory acts.

Disclosure

In some tweets users disclose they have asthma, which is often supporting an argument or an explanation. For example, a user might disclose they have asthma to explain why the air quality was important.

> *um i'm gonna need my work to stop making me clean the world's dustiest surfaces i got asthma and you try to kill me*

Stigma

Stigma can be interpreted as an underlying topic in many tweets, especially humorous tweets and tweets that signify the lack of understanding. Specifically, stigma can be found in the following forms:

1) Non-patients making fun of someone using inhaler/having an attack and or making fun of patients

> *When that boy took that inhaler out I bout died* 😂

> *nice inhaler nerd, where'd u get it the inhaler store*😭 😂

2) Using swear words to derogate someone

> *I'll never forget when it was a bunch of us in the box and this bitch decided to tell us at the last minute she had asthma*

3) Assumed patients making jokes about asthma:

> *I be hitting my asthma pump like its a blunt* 😭 😂

Sense of community

Sense of community represents the perceived support that people with asthma give to each other online:

> *If you like we can exchange experiences and treatments. The new ketamine inhaler appears to have promise. I don't want anything except to be able to help.*

This support is seen through encouragement, tips, but also warnings about triggers, such as weather changes.

Attachment to inhaler

Tweets often contain expressed need for an inhaler, frequently with dramatised sentiment:

> *SOMEONE GET MY INHALER* 😭😭😭😭😭

This can be perceived as a call for help or as a disclosure about an asthma attack. This topic is similar to humour as both mention the use or need for an inhaler. One of the key distinctions between this category and humour is in the choice of emoticons since humorous tweets mention using inhaler as a result of something funny.

Price

Some tweets contained complaints about inhaler prices and asthma prescription.

> *And too often, the inhaler prices jump for no good reason, only exceptional greed!*

Research

Research tweets were written in a more formal way and contain findings of the latest research and news and do convey no personal insights:

> *Our #CHORI researchers developed a nutritional supplement... found to improve asthma.*

Research tweets are different to perception tweets, because they are written using less colloquial language and contain no personal perceptions.

3.4.6 Cyberbullying

This study offers novel insights about public perceptions about asthma by capturing live data from the Twitter platform and developing a taxonomy to categorize emerging themes. Considering that emerging topics include disparaging humour, lack of understanding and self-pity, as the leading themes, it can be argued that Twitter, a key social media outlet, sheds light on potential negative perceptions of asthma such as stigmatization, highlighting users' psychological and emotional stress [79]. Some tweets can even be labelled as cyberbullying [214]. Cyberbullying is powerful as it is much easier to make hurtful and threatening statements online due to increased psychological distance between the victim and the bully and the reduced accountability [134]. This also means that the accountability for one's words is significantly reduced - especially, taking into account that many tweets are written anonymously. However, Twitter also offers insight into positive outcomes of social media in the context of asthma patients. Our data highlighted the sense of online community and support that patients can find online [152]. This insight indicates that there is a potential support network for asthma patients outside of traditionally explored 'off-line' support, by doctors and family [165, 192].

3.5 General Discussion

This study offers a new perspective on perceptions about asthma as the first study that combines the voices of both patients and the public, though the content analysis of Twitter data and thematic analysis of interviews. This approach enabled obtaining deeper insights about perceptions that were barely mentioned in interviews, such as self-pity or humour, which were more frequent on Twitter. Overall, there were six mutual themes, with different levels of occurrence in interviews and tweets: self-pity, humour, sense of community, disclosure, lack of understanding and attachment to inhaler. Stigma was more apparent on Twitter, however, its deeper consequences and coping mechanisms were described in more detail in interviews. Topics ‘research’ and ‘price of medication’ emerged on Twitter, but did not appear in interviews. This is to be expected given that topics discussed on Twitter were wider than the focus of the interviews. Conversely, some themes from interviews did not appear on Twitter, such as jobs asthmatics cannot do. The following sections offer a more nuanced discussion about mutual themes.

3.5.1 Consequences of stigma

The two main manifestations of stigma that can be discussed based on the findings are the trivialization of asthma as a health condition and patients' attempts at distancing from asthma due to negative connotations. Stigma and its underlying mechanisms have been mentioned as one of the damaging factors in relation to asthma treatment adherence [123]. However, stigma was expressed differently in tweets and interviews. While interviewed patients mostly only anticipated stigmatization by expressing the fear of judgements (in topics such as ‘disclosure’), tweets supported the presence of stigma in the form of discrimination in public tweets. Regardless of this dissimilar representation, there was a similarity between perceptions online and in interviews, given that both cases reflected the lack of understanding about asthma, coming from non-patients. For example, the following tweet contains the lack of understanding by (presumably a non-patient): *‘how is asthma even a thing just breathe harder smh’*

A similar perception (that asthma patients can be misunderstood) is also found in interviews:

> *Lots of people have asthma. But, I don't think there are many people who*

> *have it as bad as me. So I think people were thinking I was a bit of a wus. (Alex)*

Furthermore, some interviewed participants referred to other patients as 'weak', which is a *term* often found in tweets. These similarities between perceptions expressed in interviews and perceptions from tweets may indicate their interconnectivity and potential relationship [198]. These finding are important not only because they signify the presence of stigma in the public opinion (which was sometimes treated as questionable in the literature), but also because they could indicate that future interventions should not only be focused on changing patients' perceptions, but also on education of the wider public, especially because stigma was previously found to have adverse consequences in the lives of asthma patients [7], as well as being positively associated with poor asthma control [15].

Trivialization

Tweets indicated there is significant lack of understanding that Twitter users have about asthma. This is particularly evident in tweets that denote asthma is an 'imaginary illness' and that patients should 'just breathe'. This finding is relevant because it supports assumptions from previously conducted work about food allergies, that states that judgements and disparaging humour aid in trivialization of a health condition [2]. A quote from a patient with severe asthma reveals how this could be reflected when someone has an asthma attack:

> *I think when they see someone taking the inhaler they think they're overwhelmed. 'Oh he's or she's overwhelmed'. They think it's like when someone's in love. Many people don't have the understanding of what it is. (Amanda)*

Previous literature has mentioned a link between perceived non-importance of asthma and medication that lead to non-adherence [165]. This idea was mentioned in the previous work where a significant number of asthma patients underestimate the severity of their condition, as a potential consequence of public perceptions [230]. However, this signal of trivialization may also be a reflected in patients' discourse in public tweets, indicating the public often underestimates the seriousness of this condition. Furthermore, detecting trivialization of asthma could also indicate lower level of tolerance to discrimination given that when asthma is not perceived as 'serious', the acts of discrimination may seem less damaging [2].

Distancing

Another potential consequence of stigma could be patients' avoidance of association with negative perceptions about asthma. As mentioned in interviews, having asthma sets patients apart from non-patients, and stigma may further marginalise people with asthma. This is particularly illustrated in tweets that portray people with asthma as 'lesser' and different from a standard. Previous research indicates that those who feel too similar to a stigmatized group tend to distance themselves from it, to reduce negative associations with the self and to maintain a positive self-image [200, 2].

These behaviours could be interpreted as changes aimed at being more consistent with general perceptions [2]. This could range from not disclosing being diagnosed with asthma, using an inhaler in public, to not using an inhaler at all and non-adherence to treatment. Some interviewed participants argue they do not disclose they have asthma, due to 'hypothetical' judgements, which are on the other hand publicly expressed in tweets. The effects of this interplay between perceptions and their impact on adherence were illustrated by one participant:

> *Mot taking your inhaler in public turns into not taking your inhaler - at all.*

(Helen)

This supports previous qualitative research that labelled patients as either 'accepters' or 'deniers', based on patients' propensity to deny having asthma [5]. It is of vital importance for patients to establish positive perceptions about asthma, as these may relate to identity as someone with asthma, and ultimately – their medication adherence [143]. Negative public perceptions may play a role in creation of internal perceptions. Indeed, during the interviews, participants claimed that introducing asthma to one's identity could diminish a previously established identity of a strong, healthy person. Detecting any disassociation from identification as someone with asthma is important since it could lead to denial, which has a known link to non-adherence and the potential for life-threatening asthma attacks [4].

3.5.2 Community support

A topic that was only occasionally mentioned in interviews, but was more dominant on Twitter was a perceived sense of community between people with asthma. Tweets uncovered communication strategies created by people with asthma illustrated by 'self-pity' that can be interpreted as help-seeking, since tweets of that nature often attract supportive comments. Interviews illustrated, in more depth, the nature of support often received by patients:

> *Another friend of mine had asthma... We understood each other because of it. I never talked about it to anybody else. (Lesley)*

This kind of community between patients can be explained as assimilation, a creation of a group that develops its communication techniques and serves as support [200]. The community of asthma patients can be found on Twitter, containing a plethora of advice about coping with asthma and management of asthma symptoms, containing warnings and tips [216]. Perhaps more importantly, they often take the form of emotional support. Twitter fosters opportunities for people to create and engage with a community that shares similar characteristics and offer each other encouragement [75]. This can be illustrated through the following tweet:

> *Asthma can kiss my rear am I right fellow asthmatics?*

A similar sentiment was detected when an interviewed participant explained what she would tweet online:

> *It's stuff like: 'I'm dead now, someone come find me on the floor'... (But) if somebody else posted about it, like asking for support, I'd tell them my experience or my story, I would help. (Nicole)*

Tweets related to advice and emotional support are important, because they can counteract stigmatization, by raising awareness about asthma and sharing knowledge. Identifying existence of this kind of support available for people with asthma is novel in the literature. So far, research has only examined support patients receive from health professionals and family [245]. This study highlights the relevance of online support and community, as patients are also a part of a digital environment. These traits of social media can be useful implications for future interventions, as recent studies showed that social media may have potential to host patient education interventions and could even be used to create public health campaigns to engage Twitter users [22, 281].

3.6 Limitations and motivation for the following study

Participants in Study 1a and Study 1b were different, which may explain some of the differences observed in their perceptions. Twitter population tends to be younger than the general population [33, 210]. In addition, tweets this study were collected from anywhere in the world, from any English speaker, who may or may not be asthma patients. There also may be the cultural differences that may exist between participants interviewed in Study 1a and Twitter users (given there was no limitation in terms of the countries tweets were collected from). Considering that Twitter lacks context, it is not possible to distinguish between people with and without asthma online unless they explicitly say they have asthma in the tweet. Additionally, two studies had different methods of data collection and different sampling of participants, therefore, so we cannot clearly contrast perceptions. However, certain themes appeared to stem from one group or another (such as the lack of knowledge about asthma).

This study can be considered as an exploratory work that set a valuable basis for the several studies that followed in this thesis. This work identified perceptions that were then examined in more details in the following chapters, however it also had some limitations that motivated the remaining work. For example, this study investigated 3.000 tweets, however, these perceptions could be examined (and their presence could be established) on a larger data set, which was identified as one of the main motivations for the research conducted in Chapter 4. This work also identified the need to focus on measuring the (individual and cumulative) impact that perceptions detected in this study, have on non-adherence to asthma medication, which was in this work only inferred based on previous research and theory. The impact these negative perceptions have on non-adherence was, therefore, addressed in both Chapter 5 and Chapter 6.

3.7 Conclusions of Study 1

This work built on previous studies about asthma perceptions by comparing patients' perceptions obtained through interviews and public perceptions from tweets. In interviews, patients expressed perceptions that can be classified as internal and external, including stigmatization that could be used to group several themes. Twitter data was analysed using a content analysis and by creating a topology of perceptions, with prevalent perceptions being negative humour and self-pity. Convergent analysis indicated a significant overlap between the publicly written tweets that contain stigmatization of patients and stigma expressed in interviews, which could be an indication that some patients' perceptions may be informed by the public opinion. Stigmatization expressed in these two sets of perceptions could potentially lead to trivialization of asthma, as well as distancing from asthma by patients, both of which could negatively affect adherence to medication. However, this study also highlighted a perceived sense of community that exists online that could be an additional source of emotional support for patients. Most importantly, this work signifies that asthma patients should not be observed in isolation, but as a part of a wider social environment - including its digital component. Therefore, this study is the first study of this thesis that identified the need for asthma-educators to develop anti-stigma interventions and create more positive perceptions about asthma and asthma patients. Additionally, online sources should be recognized as an alternative form of support for patients.

Chapter 4

Perception detection using Twitter

Having undertaken a qualitative, exploratory investigation of perceptions about asthma in Chapter 3, the following step was to investigate the potential to quantify perceptions. Therefore, this study was build upon the following research questions: *how to extract, examine and measure the prevalence of asthma-related perceptions from self-expressed text data, specifically tweets*. This was done through two steps that include: delineation between perception and non-perception tweets in Study 2a and analysis of found perception tweets with the goal of obtaining the perception groups and measuring their prevalence, in Study 2b.

In order to detect perceptions from text, an important first step is 'perception detection' - essentially, creation of a filter that would provide the basis for the following steps of perception analysis. Detecting perception is, therefore, the first focus in this chapter and describes some of the main characteristics of Twitter data that have to be taken into consideration when extracting any kind of perceptions from text, not only those related to asthma. The subsequent study is more focused on the analysis and measuring the prevalence of groups of perceptions that were extracted from Twitter, with the goal to not only emphasise the existence of negative perceptions in the public, but to also highlight their commonality. In order to filter perception and then analyse and measure them, special attention has to be dedicated to revision of our definition of perceptions, in order to operationalize analysis. As such, the first part of this chapter is dedicated to additional examination of the shades of meaning within this concept, extending the discussion from Chapter 2 to consider differences between related concepts of: perceptions, opinions and emotions. Even though the differences between these concepts are often not the focus of social sciences (and are sometimes even considered as synonyms), such differences need to be more strictly defined for computational analyses - as previously mentioned, there is currently no clear method for identifying perceptions within big data.

4.1 Operationalizing a definition of ‘Perceptions’

In order to extract perceptions using the bottom up approach, it is important to create a definition of this concept. As previously detailed in Chapter 2, in the prior literature perceptions have been defined as either mental impressions or perceptions related to sensors such as vision, taste or hearing [291]. This research is concerned with perceptions related to mental concepts that include motivation, learning and experience - which is a topic that had been particularly relevant in the business context and the field of psychology.

4.1.1 Insights from other fields

Consumer and brand perceptions in particular are areas of research that have been dominant in the field of business, although practitioners were usually concerned with extracting perceptions using qualitative research or surveys [191, 238]. The main challenges of these approaches have usually been obtaining representative and sufficient sample size, due to cost and time. In terms of the work conducted with the goal of quantifying perceptions from a business context, two notable concepts arise: *semantic differential techniques* and *perceptual mapping.* The semantic differential technique is a common tool for measuring perceptions by asking participants to rate a topic on a set of seven point scales with polar opposites (e.g. smooth/harsh) [110]. Another technique is perceptual mapping, a well-known tool for establishing brand positioning (on a diagram that usually has two dimensions) [158]. These dimensions can be adjectives (e.g. price and quality) that help in establishing the gap between how the company views their brand as opposed to how customers perceive it. Both of these approaches emphasise that perceptions can be analysed using adjectives and not only though the intensity or sentiment. An example of a brand perception might be *good value for money*, which reflects both a positive perception and an aid to purchasing decisions [253].

Another field which is closely related to perceptions is research about emotions. An aspect of this research highly relevant to this framework is, therefore, 'emotion classification'. This field has been partitioned between researchers that classify emotions as categorical, dimensional or based on how the subject appraises its environment [275]. Some dimensions that have been leveraged are: positive vs. negative [255, 51] evaluation of emotions, power or intensity of emotions [310], activation [85] of emotions, attention, control, novelty, certainty, perceived obstacle and responsibility [275] of emotions. However, the only two dimensions that have appeared consistently throughout the literature are 'pleasantness' and 'level of activation' [275]. Level of pleasantness, or *evaluation* as it is addressed in this work, indicates whether the emotion is positive (e.g. happiness, joy) or negative (e.g. sadness, anger) [85]. Activity (or *activation*) is defined based on the motivation to take action (in the case of active emotions, such as frustration), or reject something based on passive emotions (such as boredom) [51]. These dimensions are represented in Figure 4.1 [310].

Another relevant part of the research about emotions is the recognised range of

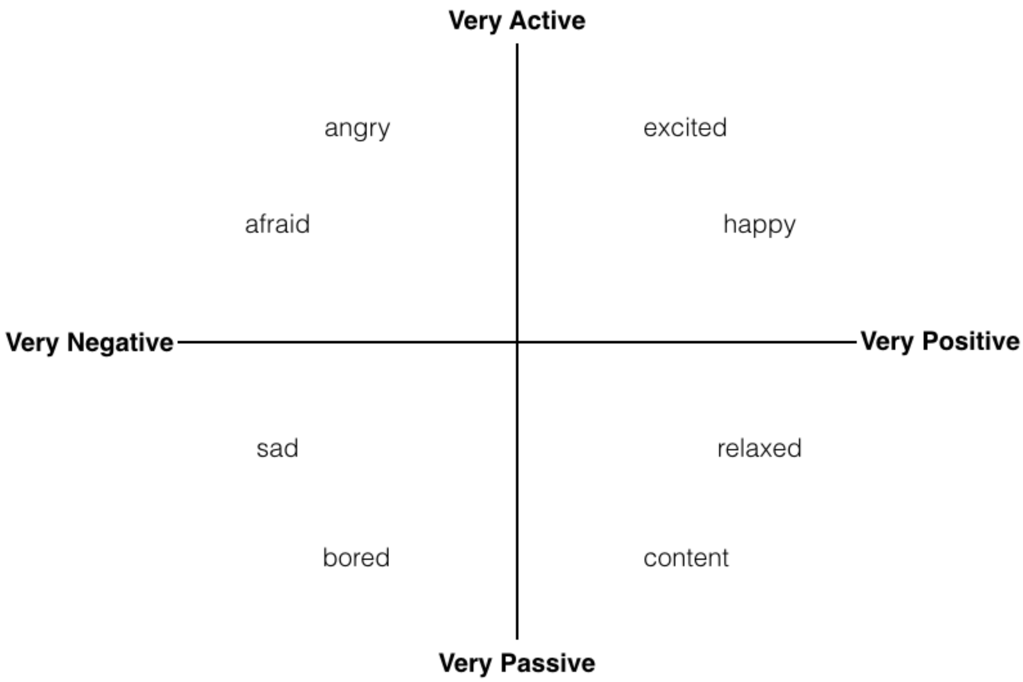

Figure 4.1: Two dimensions of emotions [310]

emotions. The most agreed-upon set of six basic emotions is: joy, sadness, fear, anger, surprise and disgust [96]. There is also a mention of secondary emotions, which are combinations of basic emotions [85]. However, Whissell (1989) suggests there are at least 9.000 words with affective connotations that can be used to express emotions [308]. This raises a relevant point: as per emotions which may require a large lexicon of words, to encompass perceptions, we may not be limited to a fixed number of predetermined words either. This idea served as a basis for the concept that is in the following sections described as 'modality': characteristics of text that represent persons' commitment to a claim [204].

4.1.2 Twitter, the place where perceptions live

Since no medical system is currently able to provide a completely holistic medical experience, patients frequently turn to online environments, in search of information sharing and support [152, 185]. As a consequence, patients often use Twitter and other social media channels to openly express how they perceive living with their condition [263]. Having in mind findings from the previous, Chapter 3, related to the significance of social media, and in combination with the argument that patients currently provide very limited and structured feedback to their medical doctors, it is clear that social media holds potential to enable a more effective flow of information about perceptions [152, 185]. For these purposes, Twitter has become an extremely popular choice in analysing user-generated content, with data readily available and more suitable for bottom-up approach in discovering patterns [210]. Twitter also has the nature of a random experiment, rather than a laboratory setting, which some authors have claimed adds to the external validity [261], which was described in more detail in Section 2.3.2.

However, Twitter also contains *a lot of noise*. Tweets can be ambiguous, contain slang, abbreviations and emoticons [113]. This creates substantial challenges to text analysis [78, 207]. Even more importantly, extracting perceptions for the purposes of sentiment analysis and opinion mining is often difficult due to a large number of tweets that contain - no perceptions. Twitter is a public platform that is open to not only individual users, but also companies that advertise their products, political parties and many other institutions that produce tweets that are not perceptions. Twitter, as opposed to Facebook, also attracts more professionals, who, as a result, accumulate a large following [235]. For example, in the case of a health condition, some tweets are created in order to promote both the latest findings in the research and to promote medications. The volume of tweets that contain no emotions or opinions is likely large enough to cause a decreased performance of traditional techniques, commonly used to analyse Twitter data [269]. For progress to be made, in terms of the analysis of tweets, filtering of perceptions from non-perceptions is an essential first step. Many analytical techniques have a tunnel vision in terms of obtaining insights - filtering of perceptions can only improve the precision of relevant information extraction [249].

The idea of removing 'noise' from tweets in this manner leads us to the concept of 'subjectivity'. Subjectivity has featured in the literature as a task of classifying sentences as opinionated or not opinionated [157]. It was used to improve question-answering in social media and in classification of online reviews [138, 36, 175]. A concept similar to perception filtering is also found within the field of emotion research, where it has been labelled as 'attention'. In this context, attention is the dimension that determines whether a stimulus should be ignored or attended [275]. Additionally, a similar idea appeared in a study that was concerned with predicting credibility in tweets. The focus therein was on linguistic features that are useful in predicting whether information in text is credible or not and in order to establish this, this study made the use of subjectivity, that differentiates opinions from facts [204].

4.1.3 Perceptions vs opinions vs attitudes

Of particular relevance to our goal of computationally detecting perception is a body of work related to a concept called Opinion Finder [313]. This concept, itself, is similar to subjectivity analysis. Opinion mining is related to perception detection proposed in this thesis, however, there are some notable differences between the two concepts. Firstly, perception tweet identification in this thesis aims to extract patients' perceptions, which includes distinguishing between tweets that express individual's perceptions, as opposed to tweets that are written by a company, PR agency or members of academia, with a goal of promoting their work. This is a subtle, yet important distinction, as Opinion mining (and subjectivity) focus on tweets that contain an opinion, regardless of a party that created the tweet.

Secondly, perceptions represent a broader class than opinions - because they do not need to contain an affective component [283]. Finally, subjectivity, as a task, has predominantly been used in the function of helping the accuracy of sentiment mining [58]. Yet, perception analysis in this thesis takes into consideration both sentiment *and* the content of tweets. If achieved, perception analysis has the potential to inform practitioners' interventions and provide a better understanding of the drivers underlying patients' sentiment. This means that the end goal of subjectivity and perception detection can be different. The methodology underpinning subjectivity detection has most frequently been associated with using a variety of lexical and contextual features for classification. The rest of this chapter will adopt a similar approach in the first part, however, additionally considering classifiers that utilize the body of text, as the main feature in perception prediction, as opposed to linguistic features.

Another concept that is closely related to perceptions is attitudes. Attitudes can be defined as one's tendency to act in a specific way due to their experience, that has a direct impact on individual's response to situations and subjects that are related to this experience [227]. Similarly, perceptions can be defined as a result of one's mental impression and interpretation of their experience, as it was stated in Section 2.1.4. 2.1.3. Both attitudes and perceptions can be substantially different from reality and they do not need to be aligned with behaviour - in the case of attitudes this is called Cognitive Dissonance [227]. However, attitudes may be seen as more closely related to behaviours than perceptions, as behaviour is one of the main components on attitudes (including also feelings and beliefs). Conversely, perceptions can be seen as a construct that is more related to the social surrounding of an individual, due to the complex inter-connectivity and nature of relationships between perceptions of one person and their social environment, which are explained in more detail in the Discussion, Section 7.1.

4.2 Towards Perception Detection

The literature from other fields (presented in Section 4.1) was used to develop a perception framework to be used for detection and analysis of perceptions in this chapter. This prior research is important, as it offers consilience around about the potential dimensions of perceptions and how they can be measured. In order to elaborate on how an operational framework can be created, the following is a summary of relevant conclusions, all drawn from the previous literature:

1. There is a need to make a distinction between perception texts and other types of content that do not express perceptions (especially on Twitter where such content is created by a range of different parties, where the inclusion of non-perceptions is likely to skew analysis);
2. Perceptions about brands can be presented visually using two main dimensions (such as price and quality in the case of brand perceptions), which is an effective visual representation;
3. The most commonly used categorization of emotions are evaluation and activation;
4. When perceptions are expressed, adjectives are highly likely to be evident.
5. Rather than pre-determining words that can describe perceptions, there is more utility in identifying a relevant lexicon via bottom-up approach.

At present, to our knowledge, there is no framework for measuring perceptions from text data or even a clear definition of perceptions in this context. Therefore, consiliented conclusions from previous research were used in the creation of the perception detection framework that can be used to investigate perceptions about people with asthma, using Twitter data. The methodology for this framework represents a combination of multiple machine learning classifiers incorporating the following tasks:

- Perception detection (or filtering) in order to filter tweets and obtain only tweets that relate to perceptions;
- Activation based on classifying tweets as either active or passive following the literature on emotions;
- Sentiment detection in order to determine whether tweets are positive or negative;
- Dimensionality reduction via topic modelling in order to obtain the most relevant perception words (and topics).

A high level view architecture of this framework is represented in Figure 4.2. The goal of perception analysis using this framework is to respond to several questions:

- How to distinguish between perception and non-perception tweets;
- Which group of perceptions are most prevalent on Twitter regarding people with asthma;
- Which group of perceptions potentially indicate non-adherence?

With this broad framework in hand, the following sections of this Chapter contain two studies that aim to respond to these questions. In the first study, a model is created to distinguish between perception and non-perception tweets, using asthma-related tweets as a case study. This consequently provides tweets that can be used in the second study, which is dedicated to the analysis of perceptions though steps of activation, evaluation and modality. The second study has the goal of uncovering the dominant groups of perceptions, interpret their meanings and prevalence in the data set of streamed tweets.

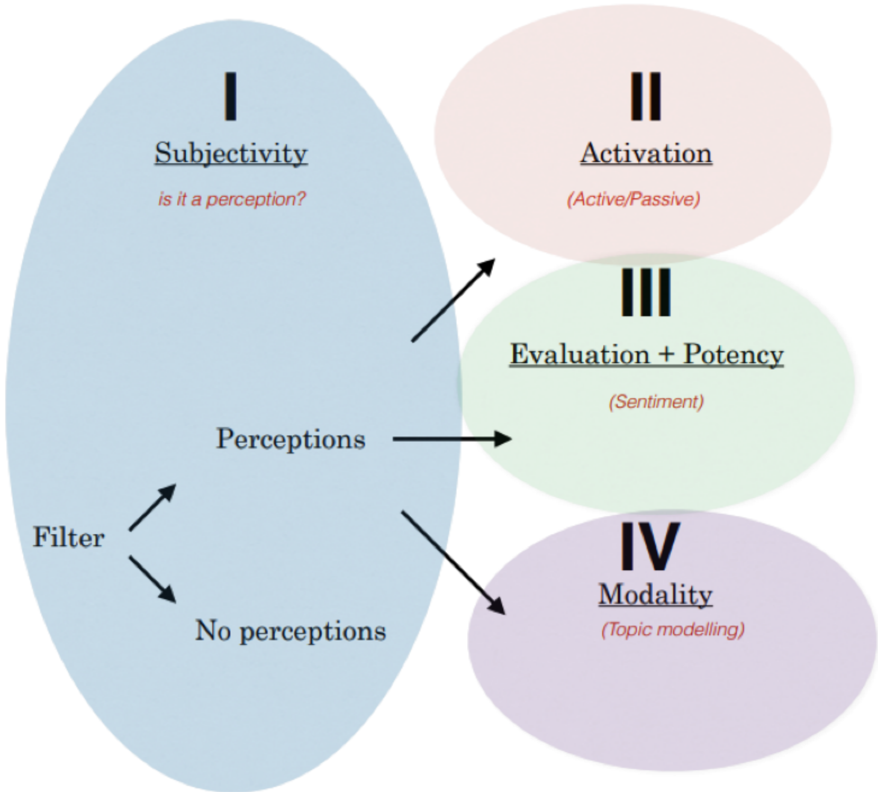

Figure 4.2: Perception Framework

4.3 Study 2a: Perception detection using Twitter

Study 2a represents the work that has been published under the name 'Perception detection using Twitter' in 2020 IEEE International Conference on Big Data [182]. The work is focused on detecting perceptions as they are expressed on Twitter and that reflect how asthma patients (and non-patients) feel about asthma and people who have this condition. The study examined methods that could be used to extract perception tweets. First, differences between perception and non-perception tweets were demonstrated in terms of their linguistic features, and the rest of the work was focused on filtering perceptions using the classification process.

4.3.1 Study Design

Study 2a is split into two parts. Using a set of ground truth labelled tweets, the first part starts to unpack the phenomenon of perceptions, investigating which pre-defined linguistic factors make perception tweets different from non-perceptions. In this section we hypothesised that perceptions tweets will be shorter, contain less punctuation, less hashtags and less capital letters. These features were then used to investigate their predictive power in terms of perception detection using traditional classifiers. Then, in the second part of the study, instead of using domain-led linguistic features to distinguish between the groups, word-embedded vectors based on the body of tweets were directly leveraged. This allowed machine learning models to use previously undetected patterns, rather than using prescribed features. A discussion of the data and the acquisition of perception / non-perception labels is common to both and discussed first in Sections 4.3.2 and 4.3.3.

4.3.2 Data

Data used in this study the same as data from Study 1b. Again, tweets were collected in 2019, and only those expressed in English language were considered, however, with no limitation in terms of tweet geospatial location. Out of over a million tweets, 376,893 unique tweets were analysed after removing all re-tweets. As mentioned, only tweets related to asthma were used, as collected though the following keywords: *'asthma'*, *'inhaler'* and *'asthmatic'*. As per the previous study data pre-processing was composed of: anonimisation of tweets, removing any duplicates or non-english tweets. In addition, lemmatization was done before transformation of documents into vectors.

4.3.3 Labelling

Using a sample of randomly chosen tweets from the collected data, a taxonomy of perception tweets was manually developed and this coding schema was then used by six researchers. Coding reliability between coders was calculated using Cohen Kappa and the score obtained was 0.79, which was deemed as satisfactory. Overall, 3,344 tweets were read and coded: 1,189 were labelled as non-perception tweets and 2,155 were labelled as perception tweets[1]. As an example, a perceptions tweet is: '*I missed class bc I had an asthma attacking this morning and I actually did my hw😭* ' or '*when i was complaining about my inhaler at work the other day and someone tried to offer me solutions???? like no, I'm here to complain not solve my issues bye*'. Examples of a non-perception tweet are: '*Volunteers needed to trial new COPD and asthma treatment*' or '*Systematic review on how primary healthcare workers in LMICs obtain information for prescribing (6) Quality of care for people with asthma | #HealthInfo4All #PHC #HealthForAll #prescribing*'.

[1]The dataset has been made available here: http://cs.nott.ac.uk/~pszgss/research/data/perception_detection_data.csv

Part 1: Summary of the first approach to perception classification: Detecting differences between perceptions and non-perceptions and classification based on their linguistic features

As previously mentioned, the first part of the study was predominantly concerned with examining the main differences between perceptions and non-perceptions and investigating whether known linguistic features could be used to successfully distinguish between the two groups. This analysis was informed by previous research that argued particular linguistic features were relevant predictors of whether information in text can be classified as a perception [204]. Based on this work, we hypothesised that perceptions tweets will be shorter, contain less punctuation, less hashtags and less capital letters. The following sections contain more insights about each of these linguistic features.

The most basic text feature that has been mentioned (as one of the Linguistic Inquiry and Word Count (LIWC) features) is the length of text [279]. Since Twitter seems to be a place for short personal expressions, the length of tweets was previously mentioned as a potentially useful feature for opinion detection [166]. In the light of this previous work, it can be argued that non-perceptions tweets may be longer than perception tweets, potentially because they contain more information.

Another potentially valuable feature that was mentioned in previous work is the number of hashtags, meta-data tags about the relevant topic [269]. Previous research stated that neutral tweets contain less hashtags [242]. Using hashtags is frequent on Twitter as tweets with hashtags are visible to more people. This can, therefore, be seen as a way of advertising for research studies, as well as for companies. Based on this, it was assumed that non-perception tweets might have more hashtags, which could potentially differentiate them from perception tweets.

Previous work also argued that tweets contained more misspellings and slang than reviews [195]. In general, by reading a sample of tweets, it could be suggested that tweets written by companies or members from the academic environment and other non-perceptions, were written without slang or abbreviations. These tweets follow grammatical rules, which means that they had a correct use of punctuation signs and they also normally start with a capital letter. For this reason, the number of punctuation signs and capital letters were investigated in this study as linguistic features.

In summary, features that were explored as potentially discriminatory were: the length of tweets, number of hashtags, punctuation signs and capital letters (see Table 4.1). Descriptive analysis was first used to establish whether there are differences between the two groups based on each of these factors. A machine learning process was then used to 1) examine the power of these factors in distinguishing perceptions from non-perceptions and 2) to examine which factors really 'reveal' a perception. Traditional classifiers such as Logistic Regression, Random Forest and Multinomial Naive Bayes were used for this task. Data used in this task was split using random sampling into training (75%) and test sets (25%) and the performance was tested on the held-out set. In order to assess the success of classification, performance was evaluated using classification accuracy with higher resolution analysis of confusion matrices [228].

	Label	Length	Punctuation	Hashtag	Capital letter
mean	0.5	128.23	4.35	0.83	7.55
std	0.5	79.95	4.5	2.12	14.37
min	0	1	0	0	0
25%	0	60.25	1	0	1.25
50%	0.5	108	3	0	4
75%	1	197	7	1	9
max	1	281	50	23	229

Table 4.1: Descriptive statistics for linguistic features

Part 2: Summary of the second approach to perception classification - classification using the body of text

The second part of this study used a different approach in detecting perceptions: it did not use pre-determined individual features, but rather the raw text of tweets, as the input for classification. To examine classification ability, Random Forest and Multinomial Naive Bayes were chosen as the first set of classifiers using the word-based frequency vectorizations (bag-of-words approach). Logistic Regression was used as a baseline classifier because it is the most simple classifier (considering it is focused on linear relationships) used in this task [92]. Additionally, since neural networks previously achieved success in accurately capturing semantics and the overall context of texts, they were also tested on this task [115]. Tweets, in the form of vectors, using word-based frequency vectorization, were used as input for neural networks' models.

Both RNN and BRNN were used in this task. Additionally, the type of RNNs used in this study, LSTM, controlled which information was stored or forgotten [163]. As previously mentioned, the only input features in this task were word-embedded vectors, created from the labelled Twitter data set. In terms of the meta parameter tuning, there was one embedding layer and transfer learning (based on pre-trained word vectors) was used as there was not so many labelled tweets. The neural network architecture had one LSTM layer, which also contained 15 LSTM nodes. Global Max Pool layer was used to output a single classification probability score. Therefore, results were probabilities allocated to each tweet, with a range between 0 (if the model predicts the tweet is a non-perception tweet) to 1 (when the model predicts that tweet is a perception tweet). This well known architecture was selected due to its good empirical performance. Different numbers of internal LSTM nodes were manually tried, with 15 used based on the performance on a validation set.

4.3.4 Part 1: Results of descriptive analysis and using linguistic features in classification of perceptions

Descriptive analysis was undertaken as a preliminary form of feature analysis. The first characteristic of tweets that was compared in both perception / non-perception groups was the length of tweets, with the unit of length being a word. Based on Figure 4.3, it is apparent that the majority of perception tweets had around 50 characters, whereas non-perception tweets were, on average, much longer. This indicated the potential relevance of this feature in distinguishing between two groups.

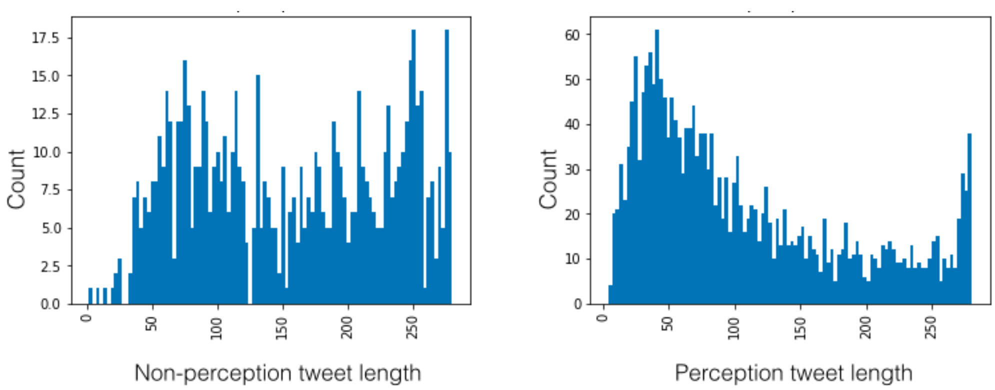

Figure 4.3: Number of words in perception and non-perception tweets

It was hypothesised that tweets written by individual users, where they express their personal opinions and beliefs, tend to be written in a casual, even grammatically incorrect way. This means that the use of punctuation, which is a sign of a grammatically correct text, would be more prevalent in tweets written by academics, or in formal communication by companies. The preliminary finding in Figure 4.4 indicated that non-perception tweets, indeed, had slightly more punctuation signs.

As stated, hashtags can arguably help in increasing the visibility of companies' tweets. For example, a representative tweet would be: '*The #flu and #cold #viruses can worsen #Asthma and can cause #sinusinfections in people with al-*

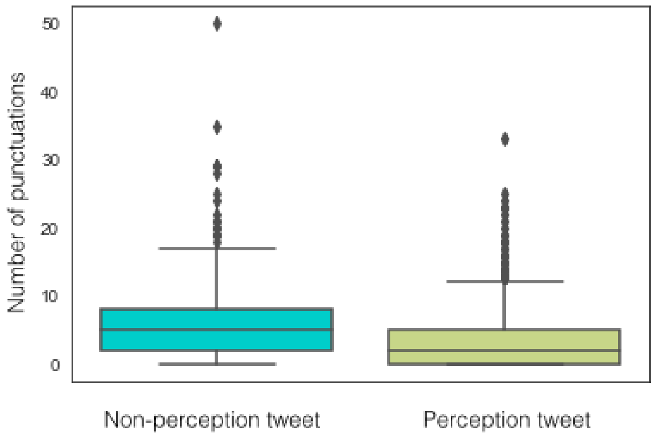

Figure 4.4: Number of punctuation signs in perception and non-perception tweets

lergies. See the advice from below to prevent the spread of cold and flu.' As seen in this example, the hashtags were used on the most relevant words in the tweet and they were arguably used less in promoting tweets written to express patients' perceptions. The results indicated that original hypothesis can be confirmed - perception tweets on average contained less hashtags. Figure 4.5 represents the difference across the two groups. This was, therefore, another feature that was seen as potentially useful in predicting perceptions.

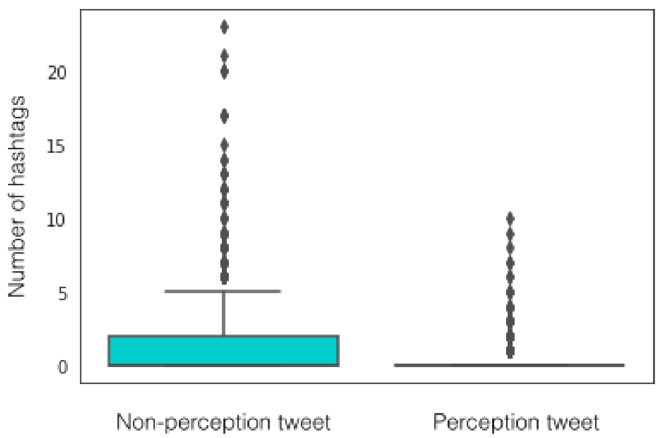

Figure 4.5: Number of hashtags in perception and non-perception tweets

It was also hypothesised that perception tweets would have less capital letters at the start of the sentence or in other, grammatically appropriate places [113]. This is another argument that was validated using the descriptive analysis of tweets. Figure 4.6 implied that non-perception tweets contained slightly more capital letters.

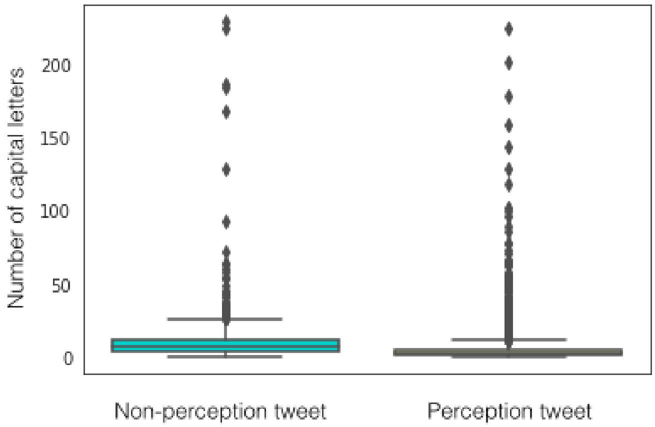

Figure 4.6: Number of capital letters in perception and non-perception tweets

Based on these preliminary findings, perception tweets appear to be, on average, shorter than non-perception tweets; and they had smaller number of punctuation signs, hashtags and capital letters. In order to examine the co-variation of these individual features, a correlation matrix was generated. The 4 by 4 correlation matrix in Figure 4.7 indicated that features chosen for this analysis are either mildly or strongly correlated between one other, reflecting the existence of some shared information. Some of this shared information is intuitive - longer tweets contain more opportunity for punctuation, while the relationship between use of hashtags and punctuation, for example, was more surprising.

Figure 4.7: Correlation between linguistic features

After this preliminary descriptive analysis, machine learning classifiers were trained to build a model that is able to distinguish perceptions from non-perception tweets based on previously mentioned linguistic features: number of words in tweets, number of punctuation signs, number of capital letters and number of hashtags.

Several algorithms were investigated in this process. Random Forest had the best accuracy of prediction (74%), Multinomial Naive Bayes produced the second best result (70%) and the slightly worse accuracy of 69% was produced using Logistic Regression classifier. The more detailed results of the prediction task were documented in Table 4.2, in which recall and precision for the perceptions class are reported as well. Additionally, based on the results of the best classifier, the permutation feature importance was obtained in order to determine which factor had the highest predictive power. The results showed that the length of tweets best indicated which group a tweet belongs to, whereas the least important factor was the number of punctuation signs.

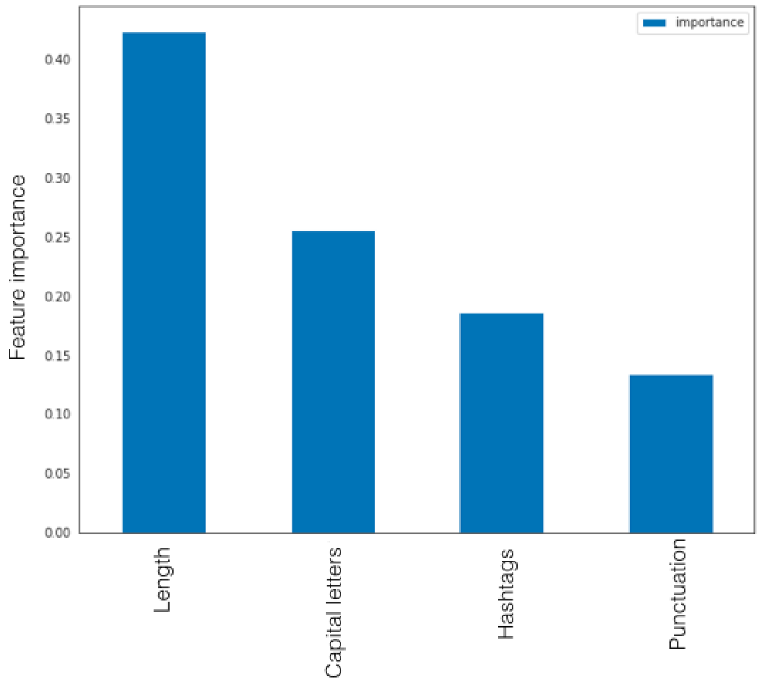

Figure 4.8: Permutation feature importance for linguistic features

Despite the simple modelling approach taken, the results of Part 1 highlight the potential to identify perceptions via rudimentary structural variables. In order

Classifier	Accuracy	Precision	Recall
Logistic Regression	0.69	0.65	0.82
Multinomial Naive Bayes	0.70	0.65	0.85
Random Forest	0.74	0.74	0.74

Table 4.2: Classification results using linguistic features

to contrast this approach of perception classification via linguistic features, the following, Part 2, is focused on classification that is not based on predetermined features, but rather on raw text.

4.3.5 Part 2: Results of classification using the body of text

As previously mentioned, Part 2 of this study followed the approach that was concerned with the overall potential of the body of text in tweets to distinguish between perception and non-perception tweets. The bag of words approach was employed with the same traditional classifiers that were used in Part 1. Table 4.3 contains the results, reporting the accuracy, precision and recall with the target class defined as 'perception'. The results indicated that Multinomial Naive Bayes achieved a significant success in classifying tweets. Prediction accuracy, when tested on a held-out set, for Multinomial Naive Bayes was 85%, whereas Logistic Regression and Random Forest accomplished slightly worse results: prediction accuracy for Random Forest was 80% and for Logistic Regression it was 83%. Considering the success of Logistic Regression (which has previously been described as a model that is focused on linear relationships), these results may also be an indication of a linearly separable problem with less evidence for subpopulations of behaviour (i.e. 'one size fits all'). Additionally, these results also indicated that bag of words approach had a significant improvement in terms of perception detection as oppose to features in Part 1.

Classifier	Accuracy	Precision	Recall
Logistic Regression	0.83	0.82	0.85
Multinomial Naive Bayes	0.85	0.84	0.86
Random Forest	0.80	0.76	0.86

Table 4.3: Classification results based on bag of words approach

Classifier	Accuracy	Precision	Recall
Recurrent Neural Network	0.88	0.89	0.86
Bi-Directional Recurrent Neural Network	0.90	0.90	0.90

Table 4.4: Classification using neural networks

Since neural networks have been shown to be highly successful in the field of text classification [65, 113], deep learning models were also trained for this study. In general, a neural network consists of an input layer, hidden layers and an output layer [113]. Two main variations of neural networks were applied to this classification task: Recurrent Neural Networks (RNN) and Bi-Directional Recurrent Neural Networks (BRNN). The main difference between RNN and BRNN is reflected in the fact that RNN processes input in a strict temporal order. This means that the current input contains the context only of the previous input and not of the future. Unlike in the RNN process, BRNN duplicates this processing chain by creating a two-directional processing that goes in both forward and in reverse (forward pass and backward pass) [304]. This means that BRNN had the access to the future context as well.

For this classification task, the best performance in distinguishing between perception and non-perception tweets had Bi-Directional Recurrent Network (90% of predictions accuracy, where predicted output was higher than 0.5). RNN also accomplished good classification accuracy (88% prediction accuracy, where outputs were also over 0.5). Based on this classification task, Bi-Directional Recurrent

Network was used on the overall sample of tweets to predict perception tweets. As a result, out of 376,893 unique tweets, 271.779 were predicted as perception tweets (see Table 4.5) and were used in the following steps of the analysis.

4.3.6 Discussion

Central to this study was identification that perception tweets structurally differ from non-perception tweets (such as tweets that are used to advertise companies and health research). According to these findings, *the length of tweets* was the most powerful feature in distinguishing perception tweets from non-perception tweets. This can be a useful insight for future research, as introducing a boundary of tweets' length could be a simple and effective first step in filtering perception tweets from non-perception tweets (with a note that this is specific to the medium examined). Preliminary findings indicated that perceptions tweets had the median length of 83 characters (as opposed to non-perception tweets that had a median length of 138 characters, see Figure 4.3). The optimal boundary for a potential cut-off between perceptions and non-perceptions can be explored in future research.

This work also highlighted one more characteristic of perception detection in text. One of the main challenges of creating a labelled training set for the supervised learning task is that it can be highly time consuming. Since this study implied that examined linguistic features provide a useful contribution to perception prediction, they might be rapidly leveraged to expedite annotation of larger amounts of data. Even though the predictive success of linguistic features was less powerful than the approach to classification in part 2, these generated characteristics of tweets still yielded a notable success in predicting perceptions. Based on this, it can be argued that there is a potential to further explore linguistic features that can be used in perception prediction with the goal to even further increase classification accuracy [228].

The most powerful results in perception detection were accomplished by text vectorisation and word embeddings. This is relevant because some prior research relies on feature construction (as done in the Part 1 of the study). However, it is arguably significantly harder for humans to find all the relevant patterns in terms of linguistic features that aid distinguishing between two sets of texts. Even when the vast majority of the linguistic features are detected and the prediction is close to the one that is accomplished in Part 2, other challenges that might appear in this manual feature extraction are the curse of dimensionality and data sparsity problem. However, when the features are generated purely on bag-of-words approach or neural networks, these are usually black-box models and interpretability can be affected. Based on these results, it can be concluded that both approaches have their own benefits and that the future work could attempt to combine them. However, by far, the most relevant highlight of this study is that perception and non-perceptions *can* be detected in text and this can represent a filtering approach (that was utilized in this Chapter in order to estimate the prevalence of detected perceptions).

4.3.7 Conclusion and Future work

This work first analyzed several linguistic features that were hypothesised to discriminate between perceptions and non-perceptions, in order to elicit a better understanding of defining characteristics of perception tweets. Results indicated that it was possible to offer strong predictions as to which tweets were perception tweets and which were not, based on their linguistic features: length of text, number of punctuation signs, capital letters and hashtags. The most discriminatory feature was the length of tweets. While these linguistic features in perception detection obtained a notable success in classification prediction, even better results were accomplished when non-linear machine learning models were trained to classify tweets using only the body of text itself, as the main input. In this context, the deep learning architectures demonstrated their state-of-the-art performance, with BRNN, in particular, showing the best performance, with prediction accuracy of 90%. Future research could investigate whether these two methods could be complementary to improve the perception detection even more. However, the main success of this work is that it served as a successful basis for Study 2b: by extracting perceptions and removing the noise from Twitter data, it enabled the following study to proceed with the analysis of perception tweets, in far greater detail.

4.4 Study 2b: Measuring perception from tweets based on activation, evaluation and modality

This section represented the basis for the paper that used a version of this method and has been published under the name 'Big changes start with small talk: Twitter and climate change in times of Coronavirus pandemic' in 2021 Frontiers in Psychology [112]. As mentined, Study 2b represents the extension of the previous section. Filtering perceptions, from Study 2a, was an important pre-step in analysis, not only for simply extracting perceptions from online sources, but even more so to: 1) obtain insights about what kind of perceptions those tweets contain and 2) to measure their occurrence in big data set. As discussed in Section 2.3, qualitative research can be used to investigate which perceptions are self-declared within groups of patients. However, to upscale that endeavour, this study relied on formalizing a definition of perceptions and descriptive dimensions of perception in the big data - with the ambition to not only examine what kind of perceptions exist, but to also quantify the prevalence of detected groups of perceptions, on a much larger sample than the one in interviews.

4.4.1 Study Design

As mentioned, the second part of this chapter is dedicated to the analysis of perception tweets, including the following *three steps*: activation (detecting whether a perception tweet is active or passive), evaluation (providing the sentiment of the perception tweets) and modality (establishing the topics that arise within the perception tweets). The following sections describe in more detail the method of executing each step and how they lead to uncovering the main groups of perceptions about asthma on Twitter.

Activation

Activation is a concept based on the assumption that all emotional states are associated with an action [275]. Level of activation implies whether people are more or less likely to take action under a particular emotion and this is why emotions can be described as either active or passive [51]. The dimension of activity is relevant in this framework because perceptions are assumed to be strong determinants of future behaviour. In the case of asthma medication adherence, for example, if people express sadness (low activity), this is described as a passive emotion and likely related to non-adherence [275]. Activity was previously examined in relation to consumer behaviour, because it is argued that some emotions have an effect on consumers' behavioural responses to brands [252]. For example, customers that are angry (expressing an active emotion) are more likely to participate in a negative activity such as complaining or campaigning against the company, whereas people who express discontent will probably not engage in such negative activity (because discontent is a passive emotion) [252]. Activity has also featured in computational analyses, but only for detecting emotional states from speech, rather than in text [85].

Evaluation or Pleasantness

Evaluation indicates whether a person has positive or negative attitudes about a topic of interest [275]. It is often described as the most basic classification of emotions [255]. Evaluation has been a point of interest in many fields, such as psychology and business, however, recently it has also become increasingly popular in computer science. In particular, this occurred as social media developed, together with the expression of emotions via language in this kind of communication [126]. Indeed, there is a substantial body of research that explores emotions through analysing text and language, very much based on the observation that text contains similar amounts of emotions as face-to-face communication [314]. Evaluation has mostly been tested on text using sentiment analysis, which uncovers the latent information about emotional states from text [116]. Sentiment analysis is today treated as a core part of analysing customers online by tracking reviews, but it is also used in other fields that track political trends or even stock markets [116].

Modality

Finally, modality is defined in this framework as the process of extracting the main themes or topics of conversation from tweets. Modality became relevant as natural language processing started developing and the polarity of text was no longer the only point of interest. In other words, it became possible to understand not only *how* people feel about something, but also *what and why* they feel this. Modality as a step in this framework is reflected through the use of topic modelling with the aim of obtaining structured information from tweets. In essence, it is a statistical algorithm that deems especially useful in extracting latent concepts that are potentially useful in the discussion about perceptions [33]. Topic Modelling has particularly been used in the previous literature to explore types of arguments in tweets, online news media bias, perceptions about tobacco products and many other topics [104, 91, 210]. One of the main implications of combining sentiment and topic modelling is the potential to imply whether the overall sentiment is positive or negative, tracking these changes over time, including the potential to investigate which features or themes are the most relevant [302].

4.4.2 Summary of the experimental plan

This section contains a short description of the overall process of analysing tweets in this chapter. Following the results of study described in Chapter 4, collected tweets were labelled as either perception or non-perception tweets. In the steps of activation, evaluation (pleasantness) and modality, *only perception tweets were analysed*. In the activation step, the goal was to classify tweets as either active or passive, whereas in the step of evaluation, the aim of the analysis was to distinguish between positive and negative perception tweets. Specifically, both activation and evaluation tasks were classification tasks with only two classes in each case (active/passive; positive/negative). The features used for both activation and evaluation were generated using the bag of words approach - each tweet was represented with a vector of the word counts and it was further normalised so that the importance of common words was scaled down [42]. Following the classification tasks (on the same perception tweets) in activation and evaluation steps, as the result, four groups of tweets were obtained: active positive; active negative; passive positive and passive negative. The final step (modality) was applied in order to obtain topics that are representative in each of the four groups.

Based on the previously described work, there is a significant value in knowing whether perception tweets are positive or negative (and following this - the prevalence of positive and negative perceptions). This would indicate whether asthma-related tweets are in general positive or negative and whether there are signals of the potential stigma presence. The activation step was added since active tweets could signify the presence of emotions that drive people to take action, which is of special interest considering that adherence to medication is also 'an action'. Of particular interest would be active *negative* perceptions since they are related to active negative emotions such as anger and may indicate stigmatization. Conversely, passive *negative* perceptions may drive non-adherence, since they contain emotions such as sadness that discourages people from taking action. However, the pure classification of tweets into these four groups may not be sufficient since there is a value in knowing *which topics* are discussed within each of these groups since this may uncover the nature and themes that are the underlying reason behind these sentiments and emotions. This is why the final step of this work is dedicated to modality, which was conducted four times on each of the four data sets (belonging to active positive; active negative; passive positive and passive negative tweets). The more detailed process of each of these steps is described in the following sections and the Figure 4.9 is the visual representation of this analysis.

4.4.3 Labelling

Activation

As mentioned, data used in the activation step was comprised only of tweets that were labelled as perceptions in Part 1 of this chapter. In the activation step, tweets were labelled as either active or passive. In order to arrive at the distinction between active and passive tweets, a categorical approach to emotion classification was used: each tweet was labelled based on which basic category of emotion it expressed and based on whether this emotion is described in previous literature as either active or passive. The model used for the classification of basic emotions as either active or passive was used from similar previous studies [310].

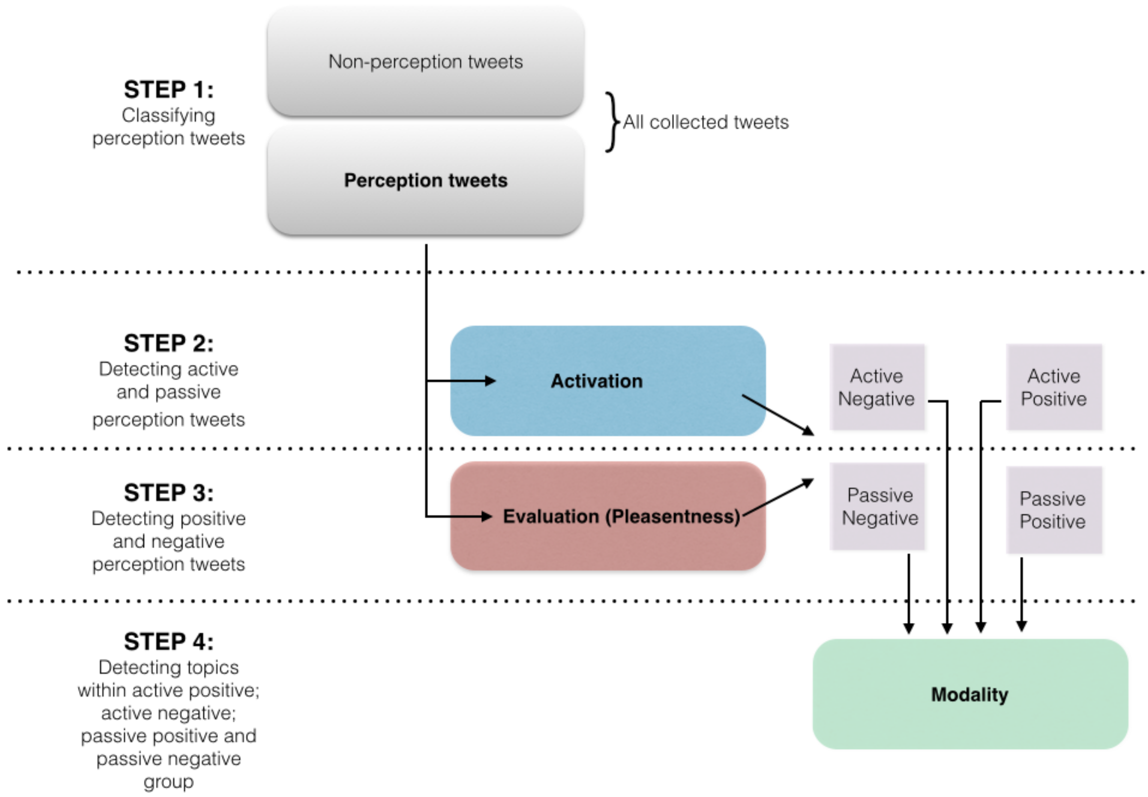

Figure 4.9: Overview of the experiment plan

A random set of 2 153 perceptions tweets was labelled as either active or passive, based on whether they contained emotions that the literature considers as either active or passive. For example, the following tweet contains excitement and this is an active emotion.

> *Dang i had a breathing treatment with albuterol for the first time in like yearsss and i feel like i can run a marathon right now.*

Conversely, the following tweet was labelled as passive as it contains sadness, which is a passive emotion.

> *Home sick today due to air quality issues. Decided to watch THE 100 two part season finale. Guess what doesn't help near asthma like symptoms :(*

In order to distinguish between anger and sadness, anger was labelled as an active emotion in cases when it was aimed at other people. If there was no solution given, the tweet was labelled as 'sad', and, therefore, passive. Tweets were labelled as active if they contained all capital letters or exclamation sign (in which case emotions were recognized as either anger or excitement). Positive emotions could be both active and passive, however humour was labelled as an active emotion as researchers consider humorous tweets to be closer to the sentiment of excitement (active) than content (passive). After labelling all the tweets, it was established that 1.025 tweets were labelled as active and 1.128 were passive.

Evaluation

Evaluation refers to extracting sentiment from tweets. Despite the potentials of sentiment analysis, there are still several challenges in this field that need to be addressed: one of the most eminent issues is related to assessing words that have different meanings in different contexts or words that carry a strong sentiment but might be used to express irony or sarcasm. Additionally, Twitter often contains slang, usually grammatically incorrect spelling and Twitter-specific conventions such are hashtags, which can all be challenging for this kind of analysis.

In general, there are two main approaches to sentiment analysis: 1. Using a previously created corpus that contains sentiment for a large number of words; 2. Manually labelling parts of text and then using supervised machine learning that builds classifiers based on those labels from the ground up [33]. It is generally recognized that manually annotating data sets gives better results than using rule-based approaches [86]. It also enables authors to obtain sentiment following their own definitions and without relying on third parties' interpretations of sentiment words [233]. Based on these considerations, the manual labelling was used in this methodology. 996 tweets were labelled as either positive or negative. The neutral category was avoided since it was assumed that 'perception tweets' contain sentiment of some form. The following is an example of a positive tweet:

> *Well I am glad to hear the improved diagnosis... sorry he has to use the inhaler, but at least its going to help. Give him my best!* 😉

This is an example of a negative tweet:

> *The housework is never ending today and my asthma is acting up. Help!*😭

When tweets contain mixed sentiment the label was based on the sentiment at the end of the sentence. More 'power' was given to the sentiment of emojis that appear in the tweet as opposed to sentiment in words. When tweets were neutral, contained sarcasm or only one word, they were labelled as positive (or based on the face value of the sentiment they express). 431 tweets were manually labelled as positive and 565 as negative.

Modality

Modality was mentioned as a task that has the main goal of expanding the previous two steps by uncovering the nature of the *content* behind the perception tweets. For this task, Latent Dirichlet Allocation (LDA) was applied as a form of topic modelling, which was used to investigate modality. LDA is one of the most standard methods for unsupervised machine learning. It is one of the machine learning techniques that assume there are a number of latent topics in a corpus [91]. Following the previous literature that defines an argument as an adjective-noun pair and the assumptions of this work that describe perceptions as adjectives, a parser was applied in this step, which means that only nouns and adjectives were kept as keywords [302, 158]. In order to prepare the data to be used by an LDA model, a gensim library was used to create a mapping between each word and its unique ID since LDA uses numbers as input and not text. Prepared datasets were then fed into the LDA model. The optimal number of topics was obtained based on interpretability. The coherence score was consulted, however the manual decision making was deemed as more appropriate and the optimal number of topics was depended on the iterative process of manually reading the most representative tweets until the interpretation made sense. This process was repeated for each of the four groups of tweets (e.g. active positive, passive positive, etc). The results for each topic were keywords that best describe the topic, and their weights that signify how relevant that word is in the allocated topic.

4.4.4 Classification process and results

Data was split using stratified random sampling to training (70%) and test sets (30%) and each models' performance was tested using the held-out set. Results for activation and evaluation tasks showed that Random Forest and Multinomial Naive Bayes (MNB) yield good results in classifying tweets. For activation prediction, Multinomial Naive Bayes accomplished the accuracy prediction of 62.29%, whereas Random Forest had slightly worse result (59.96%). However, neural networks were also used and for the tasks of activation and evaluation they had the same architecture that was also used in perception tweets detection (more detailed description about this architecture is in Section 4.3.3, second paragraph). Neural networks had much more significant success than the traditional classifiers: BRNN accomplished 64% accuracy prediction and RNN produced even better results with the prediction accuracy of 66%. Since RNN had the best result, this model was used to label the remaining data set. In terms of results, out of 271.779 perception tweets, 164.626 were labelled as active and 107.153 were labelled as passive tweets.

In the case of evaluation, similarly to previous work, neural networks had the best results. It is relevant to mention that the nature of the task often means more labelled data is needed, especially when neural networks are used. In order to resolve this issue, the classification model was simplified. This means that the number of units in LSTM (the dimensionality of outer space) was reduced to 8. Following the trade off between accuracy and loss and validation accuracy in 100 epochs, it can be concluded that the learning curve flattened around the accuracy of 63%. RNN accomplished the accuracy of 62.65%, whereas BRNNs had more success and the accuracy was 63.45%. Once the model was trained, BRNN model was applied to remaining tweets in order to obtain sentiment for the whole data set (271.779 tweets). As a result, 112.000 tweets were labelled as positive and 159.779 as negative. These results are stored in Table 4.5.

Class	#Overall Tweets	#Perception Tweets	%
Perception	376.893	271.779	72%
Class	**#Perception Tweets**	**Length**	**Punctuation**
Active	271.779	164.626	61%
Passive	271.779	107.153	39%
Positive	271.779	112.000	41%
Negative	271.779	159.779	58%

Table 4.5: Overview of total tweets, perception tweets and activation and evaluation analyses

The final set of results is related to modality or Topic Modelling. Result were represented through 4 groups of perceptions: *active positive, active negative, passive positive and passive negative perceptions.* For each of these groups a number of topics was obtained via topic modelling. The number of topics in each of the four groups was chosen based on the qualitative interpretation and these results are presented in Tables 4.11, 4.12, 4.13 and 4.14.

Active positive

There were 65.495 active positive tweets (24% of the overall perception tweets). Four topics were chosen for the active positive group of perceptions as they were the most interpretable. The words that had the highest frequency in this group were: 'attack', 'good', 'clean air', 'matter' and 'friend'. The blue circles in the topic modelling graph illustrated in Figure 4.11 are well separated and relatively large, and cover all four quadrants, which means the topics are distinctive. As it is evident from the Graph 4.11, it can be challenging to make a good interpretation of results based on a single word. Similar graphs for the remaining groups are presented in the Appendix, Chapter 8, along with the graph representing the main words in topics. In order to aid interpretation, for each of the topics, the most representative tweets (those that contributed to a certain topic the most) were extracted and read. The results of this process for active positive tweets is presented in Figure 4.10.

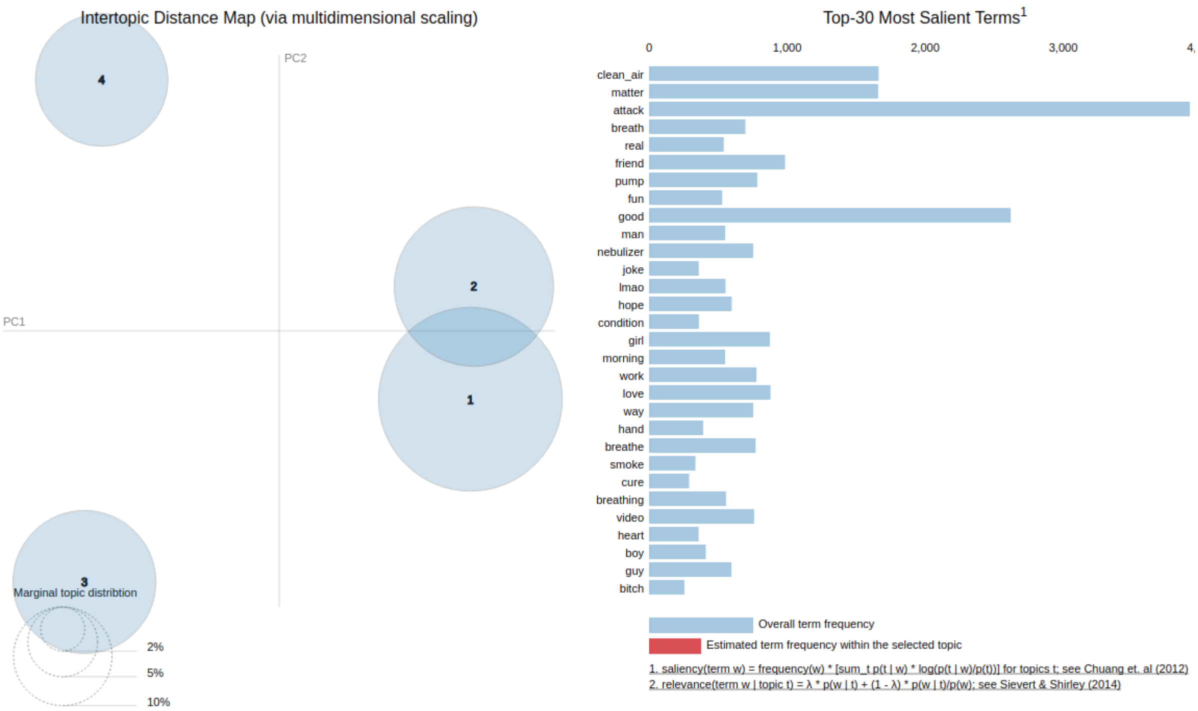

Figure 4.10: Active positive group of perceptions and its four topics and salient terms

Group: Active positive				
Topic name	**Brief Description**	**Tweet example**	**Topic size (%)**	**Topic words**
Practical tips	Advice and tips from (presumably) people with asthma to other patients	*i had a cough for 3 months went to doctor 3 times got antibiotics and an inhaler. it didn't go away till i started drinking hot tea with a lot of honey and a bit of lemon. hope his helps* ❤️	37.33%	good, day, bad, today, allergy, nebulizer, week, med, doctor, child
Empowering support / Disparaging humour	Messages of support, however also illustrations of jokes about asthma	I think my favorite thing was *a girl pulling out an* inhaler and this *other girl came up to her and asked for a hit* 💀😂 *she deadass did it too*	18.51%	attack, time, love, work, people, girl, man, guy, heart, condition
Support by raising awareness / Positive humour	Support for people with asthma by raising awareness about triggers and asking for actions	*know anyone with asthma? This is why clean air matters. see what's happening in St. Louis.*	22.59%	attack, matter, clean_air, st_louis, great, pump, hard, breath, video, breathe
Support via sharing experiences	Confessions and personal experiences of people with asthma.	*Today, someone gave me 20 fantastic free books for home schooling--and I'm watching my husband put up a shelf for them--my son is training for marathon--and my asthma isn't so bad --my daughter is balleting --and this year I really AM going to be organized --so, yeah. Life is good*	21.55%	good, year, kid, thing, life, friend, cat, school, time, real

Figure 4.11: Labelled topics and topic words within active positive group of perceptions

Overall, the results from this perception group seem to be related to tweets of support that are created and aimed at presumed asthma patients. The largest topic in this group is dedicated to support tweets formed as practical advice. To illustrate, a representative tweet from the topic 'Practical tips' is:

> *I use peppermint oil, tea tree oil and eucalyptus. Lavender is pretty good as well I have really bad allergies and asthma*

Three other topics in this group are related to additional forms of support: through sharing experiences related to asthma, providing encouragement and hope. An example of a support tweet is:

> *I hope you feel better!! Asthma is no joke and can be very scary. Get some rest if you need to. Your health comes first.* 💕

Lastly, one topic is also dedicated to support through raising awareness - these tweets are presumably aimed at non-patients and are in the function of reducing the lack of knowledge and misconceptions about asthma. Nevertheless, some tweets in this group of perceptions also contain negative humour. These tweets are examples of stigmatization through humour. There is also an aspect of expressing personal experiences and excitement about asthma improvement, which was sometimes expressed though sarcasm, which is another form of humour. This can be described using the following example:

> *Hahahahah love having an asthma attack at work and not being able to finish your shift and also feeling super grim and weak as a result, love it love it*

Passive negative

Passive negative tweets made 22.32% of the overall sample of perception tweets. Topic modelling results indicated that four topics in this group had the best interpretability. These results are stored in Table 4.12. In the largest topic of this group, patients used tweets to express their dissatisfaction with the current situation with health care systems and politics. This topic reflects the worry and sadness regarding the cost of insurance and medication - topics that are most frequently discussed in the US (there was no limit on the location of tweets in terms of streaming). These are politically charged tweets that reflect despair and inability of users to help themselves and/or other asthma patients. An example of a tweet from this topic is:

> *I am just so very sorry. We are supposedly the richest ; greatest Country on 🌏, but people are dying because they can't afford their medications. This is Criminal! We should sue politicians for screwing our insurance up so much ; enabling Big Pharma to charge ridiculous prices!*

The second largest topic in this group contains offensive swear words. This topic represents complaints aimed at people who trigger patients' asthma symptoms and reports of events when other people have been inconsiderate or made fun of people with asthma. These tweets can be interpreted as passive since they usually mention no specific action, especially since the events mentioned in tweets happened in the past and the expressed emotions are usually annoyance and sadness. The following tweet illustrates this:

> *THIS BUILDING SMELLS BAD ITS AGGRIVATING MY ASTHMA. I CANT BREATHE. ITS NOT JUST ME?!?!?" *some middle aged*

Group: Passive negative				
Topic name	**Brief Description**	**Tweet example**	**Topic size (%)**	**Topic words**
Politically charged dissapointment	Worry and sadness people feel regarding the health care	*I pay for health insurance out of pocket. For myself. For my son. We both have asthma ..*	31.5%	people, good, life, insurance, doctor, sick, kid, month, health, med
Complaints about inconsiderate others	Discrimination and inconsiderate actions of non-patients	*I gotta stop hanging around Gaby when she smokes i be taking a puff of my inhaler as if I was hitting a blunt bitch I'm just tryna breath*	24.7%	bitch, smoke, breath, ass, man, lung, breathe, air, lmao, baby
Sadness due to limitations	Complaints about poor quality of life	*Having asthma and anxiety sucks couse I never know why u can't breathe is it asthma or alergies and I can't just take my inhaler to see if it's alergies couse if ur anxiety's high ur inhaler makes ur heart race making anxiety worse and then u can't breathe anyway*	21.6%	attack, shit, breathe, bad, bc, fucking, mom, omg, girl, night
Sadness due to asthma	Complaints about asthma	*Cried today bcos of my skin asthma and eczema. This really hurts.*	22.1%	time, day, today, year, nebulizer, thing, love, guy, work, week

Figure 4.12: Passive negative group of perceptions and its four topics and salient terms

> *white bitch* "No, it is. Go outside then". I'm setting myself up to get my ass beat this morning.*

The remaining two topics illustrate the poor quality of life, challenges and limitations (assumed) patients face in their day-to-day activities, due to asthma. In particular, the smallest topic represents complaints about life with asthma in general:

> *hello i have a chest infection but ive alreayd had to do ovr time every day in work this week plus tomorrow and i have to work from home all weekend and i need an inhaler just to get up the stairs and honestly id rather just die at this point it would less painful*

Passive positive

Passive positive tweets were the smallest group of perceptions, making 17.11% of the overall sample of perception tweets. This group of perceptions contained overall positive sentiment, reflecting on memories and experiences that were either empowering or humorous. Results are presented in Table 4.13. Topic named 'Positive humour' is the largest topic and these tweets describe situations and memories or hypothetical situations that (assumed) patients found so exciting and/or funny that they almost triggered their asthma. An example tweet illustrates this:

> *My phone made a video of all the of bf with "rock music" and I laughed so hard I had an asthma attack.*

The other three topics demonstrate positive experiences patients had in relation to either other people, animals or their own empowering experiences related to asthma. A tweet from topic 'Gratefulness for other people's actions' represents situations in which non-patients were supportive:

> *One time I fell out of Morgan's tree and needed my inhaler cuz I got the breath knocked out of me and couldn't breathe. immediately takes off in a dead sprint to my house to get it no questions asked. that's love*

Active negative

Active negative perceptions made the largest group: 36% of all perception tweets were active negative perceptions. Three topics emerged as most representative in describing this group of perceptions. Tweets designated as 'Frustration with non-patients' are largely related to expressions of anger, frustration and annoyance,

Group: Passive positive				
Topic name	**Brief Description**	**Tweet example**	**Topic size (%)**	**Topic words**
Positive humour	Humorous tweets about memories when (assumed) patients laughed and almost triggered asthma sympthoms	*This made me laugh so hard I'm crying. The Ted Larson part ; Parker part almost gave me an asthma attack from laughing so hard. This is a winner thank you*	44.9%	attack, breathe, omg, love, funny, video, hard, lmao, work, laugh
Gratefulness for other people's actions	Illustrated help or support from others that makes patients feel content and grateful	*TY, I swear, I'm good.* ❤️ *They hooked me up with a nebulizer treatment in the ER, gave me a rescue inhaler before I left, and a prescription for another. Everybody's been extremely kind today, offering me extra inhalers and even prescriptions. Y'all are the best damn people.*	17.9%	good, friend, guy, people, bitch, man, life, breath, big, nice
Content related to animals	Description about how closeness of animals helped patients overcome their struggles	*So very happy with my new nebulizer! I go into a panic every time I need my meds (side effect) but this puppy will keep me happy!*	14.9%	nebulizer, day, bad, girl, today, kid, good, great, hope, tweet
Feelings of empowerment	Personal experiences about feeling supported or being able to fight asthma symptoms	*Nebulizer gang* 💨 *WYA* 😤 *let's drip drop* 💦 *this albuterol sulfate* 💉 *; get this shit* 👄 *SHAKIN* 💅 *time to clear out* 🧹 *those BRONCHIAL tubes* 🎺	22.3%	time, shit, thing, year, attack, albuterol, smoke, bro, mom, ass

Figure 4.13: Passive positive group of perceptions and its four topics and salient terms

targeting the manner in which other people treated users who wrote these tweets (presumably asthma patients). Results are saved in Table 4.14. Some tweets from this set illustrate this:

My Aunt's brother died of an asthma attack way back. It's not a joke!

Sometimes (assumed) patients also express anger towards *asthma*, in tweets, revealing a topic centered on the unfairness of having to go through things that non-patients do not have to deal with. In addition, these tweets mostly relate to negative feelings about one's body parts such as lungs, which are the object of frustration as they are perceived as 'non-working'.

I hate the heat....my fucking allergies and asthma doing the UP MOST right now.... 🤬 🤬

Group: Active negative				
Topic name	**Brief Description**	**Tweet example**	**Topic size (%)**	**Topic words**
Anger about asthma	Expressed negative emotions people have about asthma and their own bodies due to asthma symptoms	*U know how embarrassing asthma is?? Like I'm really 21 pulling out a bright red inhaler in public because my lungs are shitty*	28.8%	time, day, bad, good, today, sick, year, doctor, week, air
Frustration with non-patients	Tweets expressing anger towards asthma patients	*i swear 2 god if this bitch coughs in the fitting room one more time i'm busting the door down and shoving an inhaler down her throat*	46.8%	people, life, year, problem, medication, severe, month, med, kid, child
Frustration with patients	Expressed feelings of annoyance related to how other people treated authors of tweets	*My last real job (stress leave after a bad asthma attack due to construction dust) I was phoned daily by the head of human resources and grilled on when I was coming back to work. I finally quit. Never worked again really.*	24.3%	attack, lung, bad, breathe, allergy, thing, breath. chest, pump, cough

Figure 4.14: Active negative group of perceptions and its four topics and salient terms

However, the annoyance reflected in these tweets is not only expressed by patients, but also by people who (presumably) do not have asthma and this is denoted as ‘Frustration with patients’. This is the largest topic in this group. Tweets in this topic can be highly offensive as illustrated in the following example:

> *Next one buy her an inhaler ur dumb bitch*

4.4.5 Interim Discussion

Combining the components in this framework: activation, evaluation (pleasantness) and modality enabled obtaining four groups of perceptions that are described as either active or passive; positive or negative and in addition - themes were detected within these groups and described though topic modelling (as the step called modality). This method enabled the gathering of insights that neither of these concepts could achieve on their own: evaluation enabled obtaining the measure of sentiment for a large number of tweets (as positive or negative); activation enabled insights about the tweets that are active or passive and lastly - modality enabled the emergence of themes or content of discussion within these tweets. A simplified version of results is presented in the Graph 4.15. The more detailed results for each group are in the Appendix (Chapter 8).

Active positive

This group contained traces of disparaging humour, which was in the function of ‘making fun’ of people with asthma. However, the most prevalent concept was, in fact, ‘support’, indicating that this group is most closely linked to the perceived sense of community which was one of the findings from Chapter 3. This group of tweets also tallies with the previous work that states that the use of social media and technologies, in general, can produce multiple benefits for one’s health

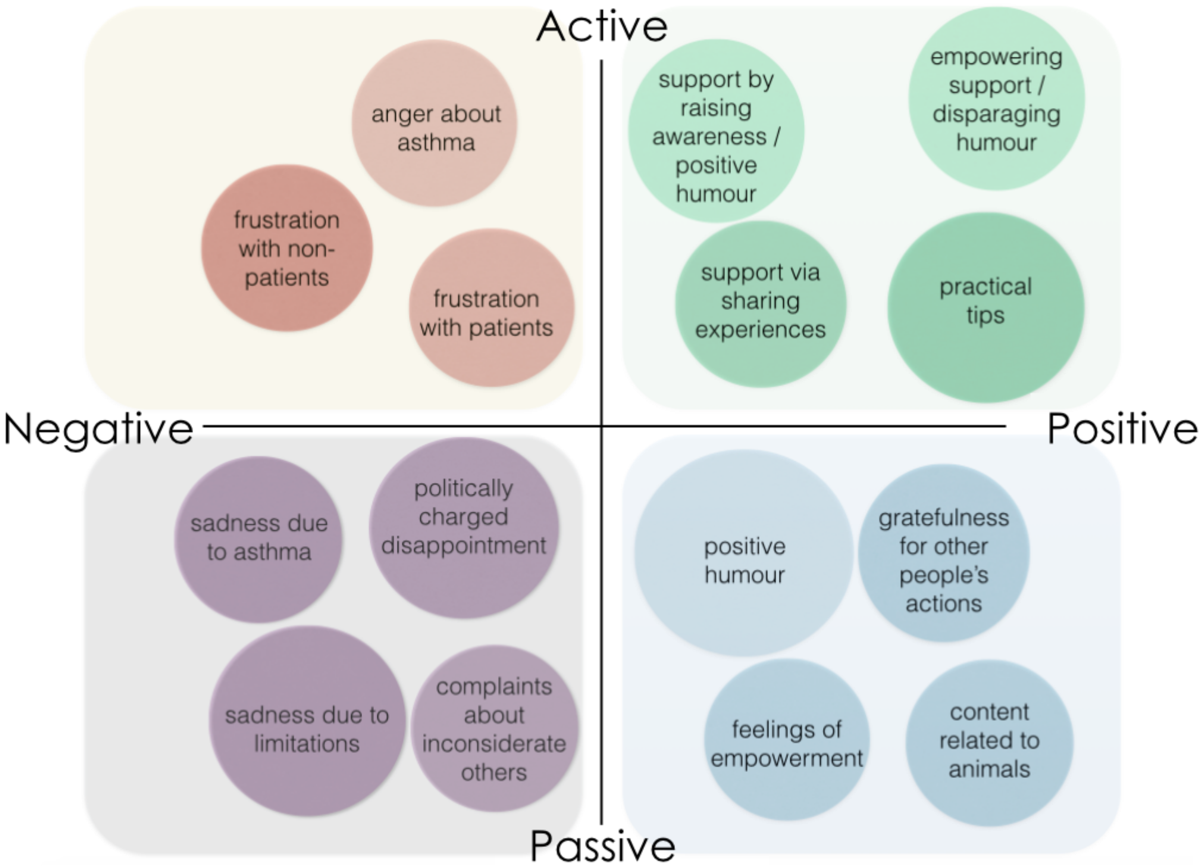

Figure 4.15: The diagram of resulting groups of perceptions (based on their activation and evaluation measure and modality)

care, through trigger interpretation, warnings and general help with asthma management [216]. Active positive tweets highlighted that there are also potentially underestimated *psychological benefits* that arise when people with asthma engage with a community of people who share the same characteristics (offering each other support and encouragement) [75]. The activity of these tweets is reflected in physical recommendations and activities, ranging from practical tips that suggest concrete activities themselves, to exuberance when sending messages of support. Active positive tweets support the findings from the previous study that hypothesised the presence of the sense of community online. In fact, this group highlights that this community is based on activities such as advice sharing and positive humour.

Passive negative

Modality analysis of passive negative tweets illustrated the inability of (presumed) people with asthma to be able to take action, whether buying necessary medication or living life without limitations, which made them feel sad and discouraged. Such tweets were mostly in the form of confessions about negative experiences with other people and asthma itself. There are also politically charged tweets that express the dissatisfaction with the healthcare system, which is not surprising, as it is well known that Twitter is the place where people express political opinions [168]. Authors of such tweets also complained about events that have happened in the past, expressing emotions specific to sadness, annoyance and discouragement.

Having asthma is often the main reason why patients may feel negative emotions (especially when it comes to perceived limitations). However, there are also forces outside of patients' control that put them in a passive negative position. In fact, following the sizes of these topics, it can be argued that in a larger number of tweets within this group (approximately 56%), presumed patients feel passive negative emotions due to others (such as inability to purchase medication or due to feelings of discrimination), rather than due to asthma itself. Felt stigma is, for example, expressed in 'Complaints about inconsiderate others'. Based on this, it can be argued that the purpose of writing tweets that contain complaints about not only asthma, but also 'outside' factors may well lie in seeking help, echoing the communication techniques that were recognised in Chapter 3. Specifically, previous study reflected the need of patients to seek support via tweeting, and this theme uncovers the potential reasons that caused them to ask for support in the first place.

Passive positive

Passive positive tweets expand the findings about the perceived sense of community, by uncovering that positive humour is the main strategy used in describing asthma on social media - and their positive effect could be reflected in normalising asthma (and hence fighting stigma). In other words, passive positive tweets reflect the humorous side of social media (44% of tweets in this group were related to positive humour). The most expressed topic is related to humorous tweets that describe experiences from patients about things that made them laugh so hard that they 'almost experienced an asthma attack'. There is a strong similarity (and overlap) with the group of active positive tweets since they are both related to the perceived sense of community, however, as opposed to active positive tweets, this group marks the emotional support, encouragement and feelings of empowerment, that do not necessarily lead to action (in contrast to 'Practical tips', a topic from the active positive group). The mutual topics are, however, the 'empowerment' and expressed gratefulness of patients for actions of others. The group of passive positive tweets is significant because it demonstrates a strong signal of online support that exists for asthma patients, which is supported through humour and empowering messages - highlighting that social media is another source of support for asthma patients (unlike traditionally examined support from friends and family) and could be focused on in future interventions.

Active negative

Active negative group of tweets is almost entirely dedicated to stigmatization: expressed discrimination against asthma patients; expressed *perceived discrimination* by asthma patients and feelings of embarrassment or strong, negative feelings towards one's 'dysfunctional body' due to asthma symptoms. This group could be considered as one of the most relevant groups especially since it contains *enacted stigma*, which is usually expressed though the acts of discrimination and unfair treatment by others. This insight is particularly relevant since it demonstrates the power of unobserved data collection - it is very rare that one would admit they engage in discrimination against asthma patients; however, Twitter data is not collected in a laboratory environment and people more freely express their real opinions. Modality was the step that uncovered having asthma is the basis for discrimination and is seen as a weakness. Moreover - it was also discovered that one of the main 'causes' for discrimination may arise due to expressed asthma symptoms such as loud coughing, or using an inhaler, which may 'irritate' non-patients. In addition, this group also contained a significant signal about internalized stigma, expressed though hatred towards asthma and even patients' own bodies when they are struggling with asthma symptoms. These tweets could also be interpreted as 'self-pity', a theme also detected in Chapter 3.

4.5 General discussion

The 'size' of stigma

Stigma represents negative perceptions about asthma, which were present in almost each four groups on the diagram. Topic related to stigma were: 'Anger about asthma', 'Frustration with non-patients', 'Frustration with patients', 'Complaints about inconsiderate others' and 'Disparaging humour'. Considering the size of each topic and taking into account potential overlapping between the topics and noise as a result of classification prediction, it can still be argued that *approximately 42%* of perception tweets were related to either stigmatization in different forms (e.g. through negative humour, discrimination) or confessions and complaints about felt, perceived or internalized stigma. This is a significant number, which highlights that stigma about asthma is not only 'alive and well', but also present in a significant amount on social media. This can be interpreted as a signal that publicly expressed perceptions about asthma need to change.

Some tweets could be labelled as cyberbullying, which was mentioned in the previous study from this thesis [214]. These kinds of messages, that are publicly available for anyone to see, could potentially aggravate feelings of embarrassment, isolation and further marginalization. Patients who are subjected to this kind of messaging online may be seen as high risks regarding non-adherence to medication, because reading negative tweets about asthma could lead to sadness (which is a passive emotion). Furthermore, it is arguable that patients who feel sad about their condition are less likely to take action and, in this case - use their medication. Additionally, insights from this Chapter provided a further confirmation that it is prudent for future interventions to address the vastly prevalent misunderstandings, misconceptions and lack of knowledge about asthma, and pay increased attention to ameliorating the 'cyberbullying' of asthma patients.

Conversely, the method used in this chapter has also enabled us to observe the prevalence of previously mentioned sense of community. It can be argued that active positive tweets (excluding the topic that contains disparaging humour) could be a good representation for the perceived sense of community. Aproximately 24.09% of all perception tweets used in this analysis could, therefore, be considered as a part of the sense of community topic. Additionally, passive positive tweets could be seen as the positive result of support asthma patients feel, where topics sucha as 'Gratefulness for others actions' or 'Feelings of empowerment' illustrate this more closely. Perceived sense of community emerged as one of the potential topics in Chapter 3, but this chapter has extended this finding, by uncovering the prevalence of perceived sense of community on a significantly larger dataset. This study also uncovered ways in which patients feel most supported: through sharing their own experiences with asthma and asthma symptoms, actions of help and messages of encouragement from other people (including non-patients) and even useful presence of animals - something that would probably be missed in interviews. This finding is important because it signals that the perceived sense of community online does not only help in fighting the ever-present stigma, but it also leads to action (e.g. via practical suggestions related to help with symptoms). Additionally, this confirms arguments from the previous study that digital environment should be seen as an alternative medium of support for patients. And increasingly - it will benefit future interventions to consider social media as a tool for supporting asthma patients.

4.6 Limitations and motivation for the following studies

A version of the method presented in this work was applied in the paper investigating the Twitter users' actions related to climate change in times of Coronavirus pandemic [112]. This paper also demonstrated that the future research could benefit from experimental method presented in this study, when assessing perceptions related to a another case study, where both sentiment and activity are relevant. However, there are still some limitations related to this method, that can be acknowledged. Following the results of the prediction accuracy from perception filtering (which had the best results), evaluation and modality, it can be argued that the models deployed on the overall data set may have introduced some noise. This could result in overlapping between topics that were obtained for each of the four groups of the diagram.

Additionally, even though Twitter is considered to be a more naturalistic setting (than interviews used in Chapter 3) it can also be argued that people often want to present a particular version of themselves on the social media, which may differ from their private self. Lastly, some tweets used in the manual labelling for evaluation and activation may have been missclassified since some emotions are similar in the written form (e.g. anger and sadness) and as a result, this would introduce additional noise. However, regardless of these limitations, this work detected a significant prevalence of stigma online. Considering that stigma is not only found in traces, but significantly present online, it can be assumed that this level of negative perceptions could still significantly affect patients' treatment adherence. Therefore, based on the findings of Chapter 3 and this work, that express not only the existence, but also the large prevalence of stigma, a hypothesis was created: there is link between stigma features and adherence to medication. This hypothesis served as a basis for the work conducted in Chapter 5 and Chapter

6.

4.7 Conclusions of Study 2

This research created a novel framework with the goal of extracting perceptions from text and detecting their prevalence. Tweets were processed through several stages of analysis: subjectivity, activity, evaluation and modality. As a result of an iterative machine learning approach, 4 groups of perceptions were isolated for analysis: active positive, active negative, passive positive and passive negative. Active perceptions were dominated by tweets where active emotions reflect either *perceived support* (e.g. in the case of excitement) or perceived and *enacted stigma* (e.g. in the case of anger). Passive perceptions could be described through feelings of *empowerment* (when positive) and *sadness* about having asthma (when passive). In terms of implications for practitioners, it is relevant to recognise that positive perceptions express the impact social media has on one's coping responses. On the other hand, the largest number of negative perceptions were related to stigma, either felt or enacted, which means that more focused interventions could be developed with the aim to identify and support high risk patients. Most importantly, implications of this work echo findings from the previous study in Chapter 3 and express the need to educate and change negative public perceptions due to their potential impact on patients' perceptions of their own condition.

Chapter 5

Stigma mechanisms and their impact on adherence to asthma medication

> "Who am I then? Tell me that first, and then, if I like being that person, I'll come up; if not, I'll stay down here till I'm someone else."
>
> - *Lewis Carroll, Alice's Adventures in Wonderland / Through the Looking-Glass*

Previous work in this thesis gave rise to the idea that stigma about asthma is not only still present in lives of asthma patients, but may also have a strong impact on their behaviour. Some previous research indicated that stigma about asthma affects adherence to medication. However, as it is stated before, there are some contradictory findings in this field. Additionally, Chapter 4 reflects the idea that stigma is not only present, but is also expressed in different 'shapes and forms'. However, in the current literature, stigma is often examined as a single, irreducible concept. Having examined (in the previous two studies) which perceptions about asthma exist, stigma about asthma was interpreted though several concepts. Hence, relying on previous studies and previously mentioned stigma-related concepts, this Chapter explored the following research question: *how do individual stigma-related features (discrimination, exogenous perceptions, denial, media impact, disparaging humour and internalized stigma) affect adherence*. In other words, the main goal of this study is to dissect the effects of stigma on adherence by focusing on factors that have traditionally been associated with stigma and examine their individual and cumulative association with adherence.

Responses from 511 asthma patients (including both mild and severe asthma) were collected using a quantitative survey that examined level of adherence and perceptions about asthma. Linear regression was used to investigate the power of individual stigma-related features in predicting adherence. In addition, in order to emphasise the varied nature of these stigma features, a mediation analysis assessed the role of coping mechanisms in mediating the relationship between stigma features and adherence. This work is important because stigma is a complex issue, and this work hypothesised that different stigma features have different mechanisms by which they affect patients. This is particularly relevant for future interventions as their design may be informed by insights about relevant patient and non-patient related stigma features.

5.1 Background

Stigma about asthma was described in more detail in Section 2.1.10 in the literature review Chapter, where stigma was described as a complex concept and it was mentioned that current research has many contradictory findings.

Another problematic characteristic of stigma research is that it has largely overlooked the delineation between aspects related to stigma, by treating sigma as a single, overarching concept [172]. As previously stated, in Section 2.1.10, studies mention the existence of three facets of stigma, that are not only typical for stigma related to asthma; they were also mentioned in research related to mental health and other health conditions such as HIV [201]. Particularly relevant in this context is enacted stigma - this is stigma that is expressed in discriminatory behaviours and actions of non-patients [278]. Enacted stigma emphasises the importance of taking into account the fact that the sources and targets of stigmatization are mutually connected and usually situated in the same social environment [172]. Ultimately, this means that stigma caused by others should not be only addressed by changing perceptions of patients.

Additionally, previous work also mentioned internalized stigma and perceived, felt, or anticipated stigma, that includes the anticipation of negative treatment if others find out about patient's condition [7, 93, 278, 9]. This work refers to perceived or felt stigma as *exogenous perceptions*. These kinds of perceptions can be interpreted through the *looking glass theory*, that states an individual uses reactions of others to estimate their own views [5]. Particularly, we can assume that a large amount of perceptions patients have, are created and reinforced based on the perceptions that exist in the public sphere. Damaging effects of enacted, internalized and anticipated stigma on adherence to medication has been mentioned in the case of tuberculosis and HIV patients using quantitative methods [9, 93]. However, the effect of stigma subtypes on adherence to asthma medication remains largely unexplored.

5.1.1 Denial

Another key concept related to stigmatization is denial. It has been argued that, in order for patients to adhere to medication, they need to accept they have the condition [5, 240]. Deniers were characterized as patients who struggle to incorporate the identity of an asthma sufferer with their everyday identities [219]. Simply said, people who experience the effects of stigma around asthma are also likely to deny that they, themselves, have asthma [5]. Denial has been recognised as harmful with regards to adherence to medication, corrosive to one's well-being, leading to decreased, delayed and interrupted healthcare utilization, negative asthma outcomes and even death [237, 219]. However, similarly to stigma, denial has mostly been investigated though qualitative studies that imply that denial could be one of the reasons for non-adherence [123, 246]. Conversely, a quantitative study implies that adherence score and denial were not significantly correlated [70]. Moreover, another work implies that denial was not even significantly correlated with the feelings of stigma [52]. This means that currently there are no congruent findings in relation to the impact denial has on asthma patients and their behaviour.

There has been a debate regarding whether denial, as a construct, can be seen as congruent to repressive coping (a defensive style characterized by high levels of anxiety and low self-reported anxiety) [70]. However, one study found that the association between denial and repressive coping is not strong enough to consider them equal [70]. Denial can rather be interpreted as an attitude to asthma self-management, which means that denial can be interpreted as closely related to perceived need of treatment, where lower belief about their necessity for treatment leads to lower adherence rates [188, 137, 69].

5.1.2 Coping mechanisms

Coping mechanisms were described in more detail in Section 3.1.2. In addition to previously mentioned repressive coping, which is a defensive coping style, it was mentioned that previous literature has identified appropriate and dependent coping styles. Unlike negative coping styles, appropriate coping mechanisms, such as information seeking is useful for patients as it enables people to obtain knowledge and even more importantly, gain a sense of control over their condition [278]. One study classifies coping strategies as: restrictive lifestyle, hiding asthma, worry about asthma, information seeking, positive reappraisal, and ignoring asthma [1, 295]. In particular, hiding was tested in one study that examined the effect coping strategies have on quality of life of adolescents with asthma and found that agreeableness is fully mediated by reappraisal and hiding coping strategies [295]. However, findings about association of coping mechanisms on adherence to asthma medication is scarce. According to a review papers [24, 41], only two studies had investigated the relationship between adherence to asthma medication and coping: a qualitative study found deniers under-used preventer medication [5]; another study claimed that adherence was negatively associated with avoidance as a coping style in patients with hypertension, however, not in asthma patients [107].

5.1.3 Media

When discussing stigma, it is also very important to address concepts that have an impact on how an individual might see themselves, but which are outside of their control. Such a perception-creating concept is media, that is often seen as one of the main perpetuators of stigma. Previously mentioned analysis suggests, out of 66 US films that contain asthma scenes, more than 17% of them portrays characters with asthma as 'wimps' and 'outcasts' [15]. Another content analysis claims that more than one fourth of newspapers contained stigmatization [151]. Media has been recognised as pivotal in creating stigma, as it can fuel the lack of education and fear of medical conditions, leading to further ostracization of patients [7]. This can be reflected on patients' treatment adherence because depending on how people interpret media portrayals of a medical condition, this informs their identity amid their cultural environment and could strongly impact the use of their inhaler in public [62]. However, the impact that media has on adherence has not been quantitatively investigated in the case of asthma patients.

5.1.4 Humour

Humorous portrayals of asthma patients in the media, lead to previously mentioned disparaging humour, previously mentioned in 3.1.2, where it was described that disparaging humour is thought to reinforce negative perceptions, prejudices and stereotypes held in the society [147, 108], hence aiding stigmatization. The effects of disparaging humour were also mentioned in Chapter 4 and it is implied that disparaging humour could be a significant stigma-related feature when it comes to its impact on non-adherence. Previous studies have found that negative humour leads to the trivialisation of potentially deadly food allergies [2], but also to the fear of being targeted by ridicule [147]. We could, therefore, hypothesise that asthma patients may avoid using their inhaler in public, out of fear of being made fun of. Similarly, a study conducted with European patients claims that a significant number of asthma patients underestimate the severity of their condition, as a potential consequence of trivialization caused by negative humour [230]. Nevertheless, the effects of disparaging humour on adherence to asthma medication have, to our knowledge, not been examined before.

5.2 Current work

Despite its negative power, stigma is not easily assessed, especially due to its versatile nature and limitations, such as desirability bias [190]. The literature also implies that research related to stigma and impacts on adherence are largely scarce. There are many contradictory findings and the majority of research has usually focused on investigating stigma without distinguishing between its underlying mechanisms and without differentiating between non-patient and patient-related stigma features. To address current gaps in the literature, we conducted a quantitative survey that is focused on examining the relationship between individual stigma-related concepts and the power they have in predicting adherence to asthma medication. We hypothesised that individual stigma-related features and adherence would have negative relationships.

It is also suggested that as a result of negative experiences, patients modify the significance of their experience through different coping mechanisms [137]. As a result, self-care is not only an adjustment in terms of behavioural patterns and coping with the limitations that asthma imposes, it is also linked to the kind of perspective and belief system that people have around asthma [192]. This why we also hypothesised that there is a significant mediating effect of coping strategies between relationships between stigmatization and adherence.

5.2.1 Participants

UK participants, older than 18, who received an asthma diagnosis at some point in their lives, were eligible for entry into the study. The majority of participants who took part in the study were between 25 and 34 years old (155); 67.34% of participants were female and 32.63% were male. The largest number of participants had had asthma for more than 15 years (267), whereas only 16 participants had asthma for less than 1 year. 85% of participants had mild asthma and 15% had severe asthma.

5.2.2 Procedure

Information for participants and the consent form were included on two pages before the survey questions. Participants were required to complete the consent form in order to confirm they understood the nature of the study and agreed to participate. Upon starting the survey, participants were asked whether they were diagnosed with asthma and if the response was positive, they could proceed to other questions. Each participant was rewarded with points by Prolific, equivalent to £2. Survey respondents were also given an option to email the researcher at any time, for information about the research or related to their consent. No participants made further inquiries.

The study was administered online. It was created using Google Forms and advertised on the Prolific website, an online platform that enables data collection. The study was described as an exploration of patients' experiences about life with asthma. 545 individual responses were collected in the period between the 1st and the 5th of May in 2020. 511 valid responses were used in the further analysis. Time needed to complete the survey was between 11 and 23 minutes. No information about participants related to their location, name, or any other identifying information was collected. Data was stored on the encrypted, password-protected server at the University of Nottingham.

5.2.3 Materials

Three main parts of the survey (that were of interest in this study) were: demographics question (e.g. age, gender, education); questions about perceptions of asthma (e.g. questions from the stigma adjusted scale, questions about the role of media) and questions about related behaviours and coping mechanisms (e.g. adherence to medication, smoking, exercising). The full question list appears in the Appendix, Chapter 8. For the purposes of this analysis, the focus was on the relationships between stigma-related concepts, adherence behaviours and coping mechanisms. The following section includes measures for these concepts and where applicable, Cronbach's alpha was used as an indicator of internal reliability of scales created. Additional statistics are stored in Table 5.1.

Construct	Questions	Response scale	Cronbach's alpha	Mean	SD
Adherence score	Do you sometimes forget to take/use your medication? Thinking over the past two weeks, were there any days when you forgot to take your medication? Have you ever cut back or stopped taking your medications without telling your doctor, because you felt worse when you took it? Did you take your medications yesterday? When you feel like your health condition is under control, do you sometimes stop taking your medications? Do you ever feel hassled about sticking to your treatment plan? Prior to COVID-19, did you find it difficult to take all your medications simply due to daily life? Do you sometimes forget to bring your medication when you travel, go to work or socialise?	Yes/No	/	5.13	1.65
Denial	Even though I am diagnosed, I think I may not have asthma. My asthma is not as serious as my doctor and my diagnosis say it is.	1 - Strongly disagree; 2 - Disagree; 3 - Neutral; 4 - Agree; 5 - Strongly agree	0.71	2.02	1.01
Discrimination	I have been discriminated against at work because of my asthma. Sometimes I feel that I am being talked down to because of my asthma. I am angry with the way some people have reacted about my asthma. I have had any trouble with people because of my asthma. I would have had better chances in life if I had not had asthma.		0.72	1.94	0.79
Exogenous perceptions	I do not use my inhaler in public because people might make fun of me. I prefer if people did not see me using my inhaler. I worry about how people might react if they found out about my asthma. People would treat me differently if they knew I had asthma. I worry about telling people I use an inhaler. I do not tell people at my workplace that I have asthma.		0.78	2.13	0.83
Media	Media (films, TV shows, etc.) generally portray asthma in a negative light. Society does not perceives people with asthma as strong.		0.72	3.59	0.82
Disparaging humour	People sometimes joke about asthma to put me down.		/	1.80	1.10
Internalized stigma	Some people with asthma are weak.		/	2.23	1.20
Information seeking	I often try to find out more about asthma. I take note of new updates in the media concerning asthma.		0.78	3.11	1.11
Ignoring	I try to forget that I have asthma. I pretend that asthma does not bother me at all.		0.69	3.24	1.03

Table 5.1: Description of Major Study Features

5.2.4 Adherence score

The Morisky-8 item medication adherence questionnaire, a popular scale for measuring adherence to medication was used [215]. The scale is composed of eight questions with 'yes' or 'no' responses, including items such as '*Do you sometimes forget to take/use your medication?*'. The obtained adherence score is a sum that ranges from 0 to 7, with higher scores indicating higher adherence [248, 144].

5.2.5 Stigma measures

Discrimination stigma (or experienced stigma) was measured by an adjusted five-item scale, previously used to assess discrimination of mental health patients. This scale included items such as *I have been discriminated against at work because of my asthma* [160]. Participants rated their experiences from 1 (strongly disagree) to 5 (strongly agree). The mean of these items formed a reliable scale (alpha =

.74).

Another set of perceptions related to stigma were exogenous perceptions, previously described as beliefs that patients have, based on what they perceive others think about them. These perceptions include anticipated stigma that may or may not be accurate reflections of beliefs held by non-patients. Items measuring these perceptions were based on a scale for measuring stigma related to HIV and mental health (e.g. '*I worry about how people might react if they found out about my asthma*'; '*People would treat me differently if they knew I had asthma*') [176, 160]. The score for exogenous perceptions of stigma was created using a mean for the six scale items which formed a reliable scale (alpha = .77). The item that measured internalized stigma was adapted from the Stigma Scale for Mental Health and the word 'dangerous' was replaced with 'weak': '*Some people with asthma are weak.*' [160]. This item was measured with a scale from 1 (strongly disagree) to 5 (strongly agree).

Stigma related to media was examined using two items (e.g. '*Media (films, TV shows, etc.) generally portray asthma in a positive light.*') which formed a reliabile scale (alpha = .74) [201]. Denial was defined using items '*Even though I am diagnosed, I think I may not have asthma*' and '*My asthma is not as serious as my doctor and my diagnosis say it is.*' which formed a reliable scale (alpha = .71) [219]. Lastly, disparaging humour was measured using one item: '*People sometimes joke about asthma to put me down.*' [72]. All items were measured with a scale from 1 (strongly disagree) to 5 (strongly agree).

5.2.6 Coping mechanisms

Coping mechanisms used in this study were measured using the adjusted Asthma Specific Coping Scale and they included: restricted lifestyle, hiding asthma, positive reappraisal, information seeking, ignoring asthma, and asthma worry [1]. Some coping mechanisms were excluded, given their scales had low internal reliability with Cronbach's alpha lower than 0.6. Namely, questions that made the scale '*worrying about asthma*' had Cronbach's alpha of 0.49; '*positive reappraisal*' (0.51); '*hiding*' (0.50) and '*restrictive lifestyle*' (0.49). Coping mechanisms that were *included* in this analysis were: '*information seeking*' (alpha = .79) and '*ignoring*' (alpha = .71). More details about these two features can be found in Table 5.1.

5.3 Results

Statistical analysis was performed using SPSS (version 27) and Python, programming language. Mediation analysis was conducted using PROCESS for SPSS to test the mediating effect of coping mechanisms on individual relationships between stigma-related features and adherence [132]. Missing values were handled using median imputation. The traditionally used cut-off of p-values < 0.05 were considered statistically significant. In the first step, we examined the relationship between features related to stigma and adherence to asthma medication using Pearson's correlations. Secondly, Multiple Linear Regression was conducted to assess the relationship between stigma factors and adherence. Additionally, a similar analysis was applied on coping mechanisms in order to assess how they related to adherence. Summaries of regression diagnostics include regression coefficients, SEs and p values that correspond to 2-sided testing of zero regression effects. Finally, mediation analysis was conducted to examine how coping mechanisms help explain the relationship between stigma dimensions and adherence.

5.3.1 Descriptive statistics

Descriptive statistics were produced for the whole sample (N = 511). Distribution of the adherence measure indicated that the largest number of participants had the adherence score of 5 (see Figure 5.1). The mean adherence level for the sample was 5.13 (SD = 1.65). Out of all stigma related concepts, media had the highest mean value of 3.59 (SD = 0.82), whereas disparaging humour had the lowest mean of 1.80 (SD = 1.10). Denial had a mean of 2.22 (SD = 1.01). The additional statistics are stored in Table 5.1.

Adherence was negatively correlated with denial, exogenous perceptions and discrimination (see Figure 5.2). The strongest Pearson correlation was between adherence score and denial ($r = -.24$, $p < 0.001$). There was a less significant negative correlation between adherence and internalized stigma and humour ($p < 0.05$). Media and adherence had no statistically significant correlation ($p > 0.05$). In terms of correlations between individual features, discrimination and disparaging humour had the strongest correlation ($r = -.50$, $p < 0.001$), followed by discrimination and exogenous perceptions ($r = -.47$, $p < 0.001$).

5.3.2 Stigma-related factors: Correlations and Multiple Linear Regression

Multiple linear regression was performed to examine relationships between stigma-related concepts and adherence. The six stigma-related variables were entered simultaneously and they were treated as continuous input features; adherence score was a continuous output feature, see Table 5.2. Results indicated a statistically significant model ($F = 11.47$, $p < .000$) where the predictors explained 8.1% of variance in adherence score. The previous work mentioned that demographic features, such as age and gender, asthma traits (e.g. duration) and perceptions from

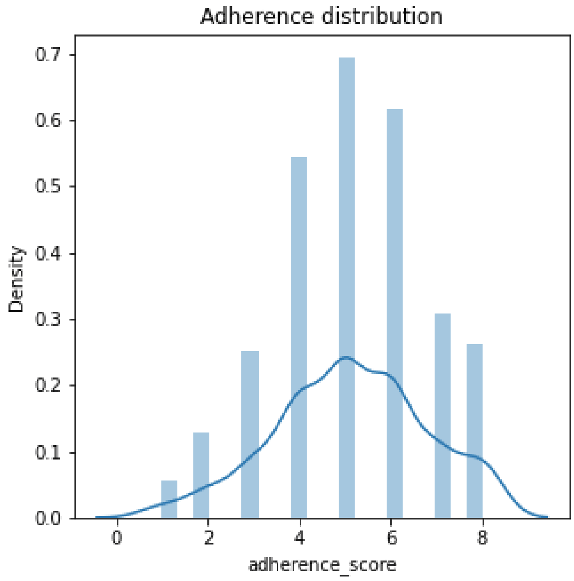

Figure 5.1: Distribution of the adherence score

the Self-Regulatory Model (e.g. certainty about asthma status, beliefs about antecedent causes) were able to predict 28.7% of the variance in adherence [150]. However, the result that was obtained in our analysis can still be considered as significant, because it explains 8.1% of variance while only remaining on perceptions about asthma (does not include other groups such as demographics) and it only relies on perceptions related to stigma (does not use other perceptions such as perceived sense of community). Based on this, it can be argued that stigma features on their own still generate a significant result in terms of their impact on adherence.

In terms of individual stigma features that have, specifically contributed to this result, denial had the most dominant role, as it was successful in predicting adherence with a statistically significant result ($p = .000$). Exogenous perceptions and discrimination had a significant negative direct correlation with adherence score, however in the regression model they yielded p values that were not statistically significant ($p > 0.05$). This ndicated that there may be some overlap in variance

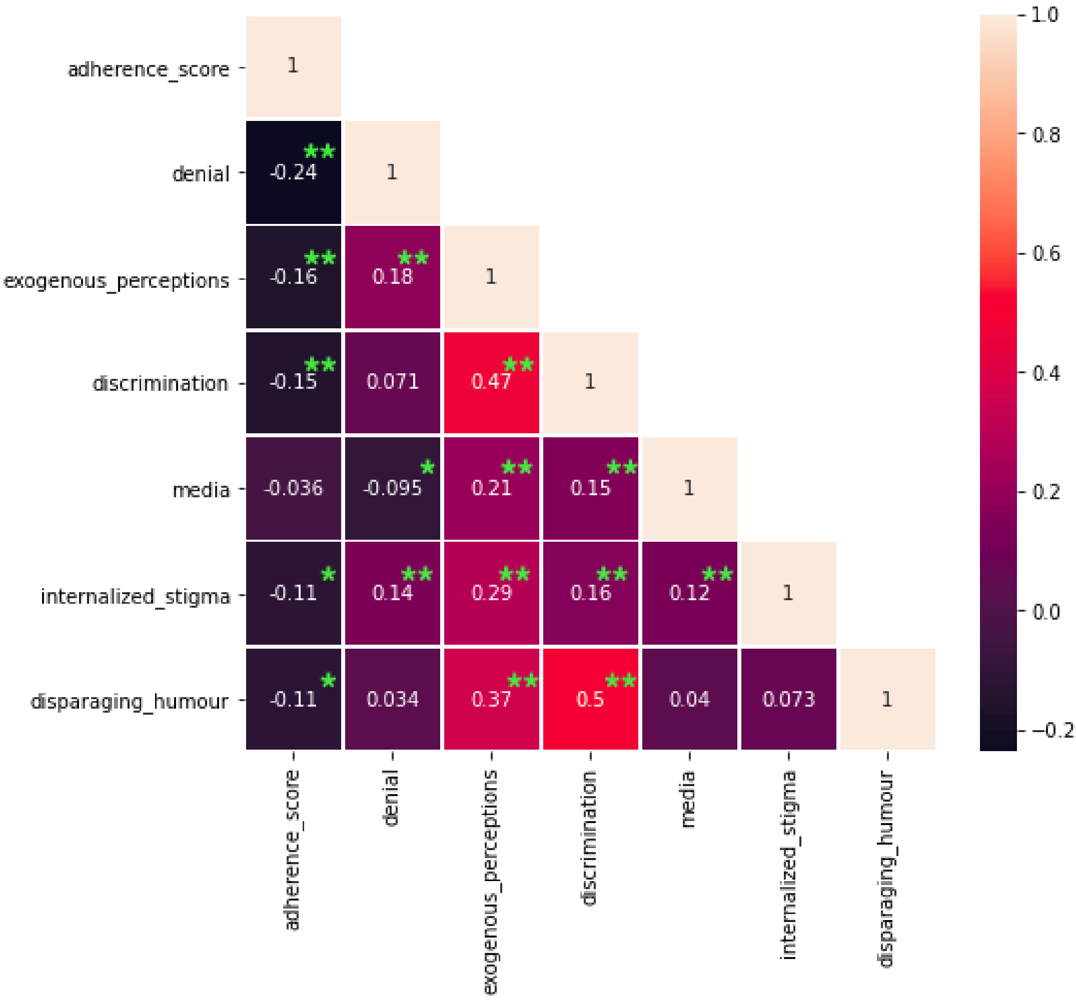

Figure 5.2: Correlation matrix of the Pearson correlations between factors Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

explained by other predictors that are included in the model.

One of the issues that can arise in linear regression is multicollinearity [232]. All stigma facets scored a lower VIF than 2.5 [11]. However, Table 5.2 demonstrates that exogenous perceptions and discrimination had significant correlations with other stigma features, which may indicate that there is a level of variable masking and a difficulty for a model to partial out the variance in the dependent variable, uniquely attributable to each predictor.

				Correlations	*Collinearity Statistics*	
Variable	**Coefficient**	**SE**	**p Value**	**Zero-order**	**Tolerance**	**VIF**
Constant	6.808	.383	.000			
Denial	-.355	.072	**.000**	-.237**	.937	1.067
Exogenous perceptions	-.105	.103	.310	-.162**	.669	1.495
Discrimination	-.151	.110	.173	-.145**	.648	1.544
Disparaging humour	-.061	.075	.417	-.108*	.724	1.381
Media	-.053	.089	.550	-.036	.923	1.083
Internalized stigma	-.066	.062	.285	-.111*	.899	1.112

Table 5.2: Linear regression results for stigma features Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

5.3.3 Coping mechanisms: Correlations and Multiple Linear Regression

We examined the relationship between between coping mechanisms (information seeking and ignoring asthma) and adherence to asthma medication using Linear Regression; direct correlations are also provided for comparison. Results showed a statistically significant positive correlation between information seeking and adherence (0.096; $p < 0.05$); and a significant negative correlation between ignoring asthma and adherence (-0.170; p<0.01). There relationships are illustrated in Figure 5.3.

The results of Linear Regression implied there was a statistically significant model (F = 9.78, p<.000) and both coping mechanisms yielded statistically significant p values, see Table 5.3. Higher level of information seeking were related to higher adherence scores. In contrast, ignoring asthma had a negative association with adherence, which means higher asthma ignoring was related to lower adherence.

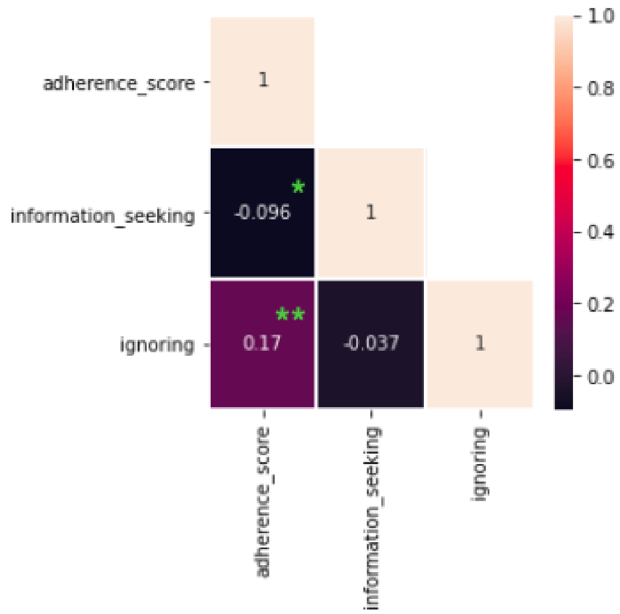

Figure 5.3: Correlation matrix of the Pearson correlations between adherence score and coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

5.3.4 Mediation analysis

A series of mediation analyses was conducted to examine whether coping mechanisms mediate relationships between stigma features and adherence using least squares path analysis. Mediation effects were tested using Hayes' PROCESS (model 4) [315]. Regression values were calculated to estimate the direct effect

				Correlations	Collinearity Statistics	
Variable	**Coefficient**	**SE**	**p Value**	**Zero-order**	**Tolerance**	**VIF**
Constant	4.775	.283	.000			
Information seeking	-.134	.065	**.039**	-.096	.999	1.001
Ignoring	.268	.070	**.000**	.170	.999	1.001

Table 5.3: Linear regression results for coping mechanisms features Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

(without controlling for mediators) and indirect effect (after controlling for mediators). Specifically, to test the hypothesis that coping mechanisms mitigate effects of stigma related features on adherence, bootstrap regression analysis was carried out. Mediation models assessed the mediating effect of ignoring asthma and information seeking as coping strategies on the relationship between each individual stigma related feature previously identified (denial, exogenous perceptions, discrimination, disparaging humour and internalised stigma) and adherence to asthma medication. Media was not included in this analysis because it did not have a significant relationships with adherence in the previous regression analysis. In all mediation models, associations between coping mechanisms and adherence were statistically significant. As in the previous regression analysis, the relationship between information seeking and adherence was positive and between denial and adherence the relationship was negative.

Our first mediation examined the relationship between denial of having asthma and adherence to asthma medication medicated by information seeking and adherence. Greater levels of denial were related to lower levels of information seeking (standardized beta = -0.146, SE = 0.048, 95% CI (-0.241, -0.051)) and higher levels of ignoring (standardized beta = 0.216, SE = 0.044, 95% CI (0.129, 0.303)). The direct effect of denial on adherence was significantly negative (standardized beta = -0.33, SE = 0.072, 95% CI (-0.472, -0.189)). The indirect effect of denial through ignoring was significantly negative (standardized beta = -0.043, SE = 0.019, 95% CI (-0.087, -0.010)). However, information seeking was not a significant mediator (standardized beta = -0.014, SE = 0.011, 95% CI (-0.040, 0.052)). This is reported in Figure 5.4.

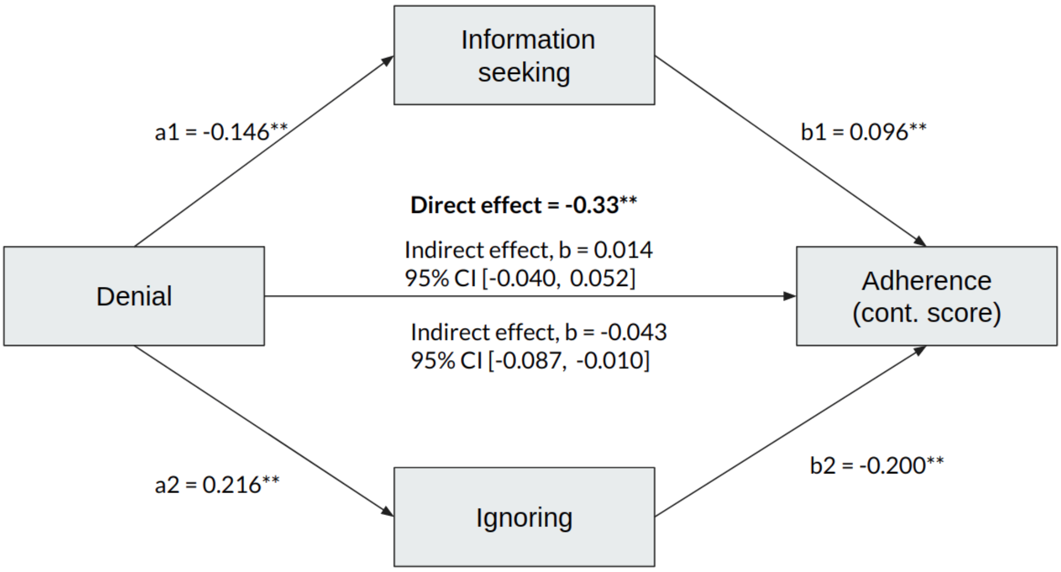

Figure 5.4: Conceptual framework of mediation effect of denial on adherence via coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Figure 5.5 represents the mediation model where exogenous perceptions are related to adherence to asthma medication, mediated by information seeking and the extent to which people ignore their condition. Exogenous perceptions were negatively associated with adherence (standardized beta = -0.255, SE = 0.091, 95% CI (-0.434, -0.077)). There was no statistically significant link between exogenous perceptions and information seeking (p = 0.103), whereas the association between exogenous perceptions and ignoring asthma was significant and positive and the beta coefficient was higher than in the case of other stigma dimensions (standardized beta = 0.392, SE = 0.052, 95% CI (0.298, 0.493)). The indirect effect of exogenous perceptions on adherence, through ignoring was significant (standardized beta = -0.079, SE = 0.033, 95% CI (-0.147, -0.018)). The indirect effect though information seeking was not statistically significant.

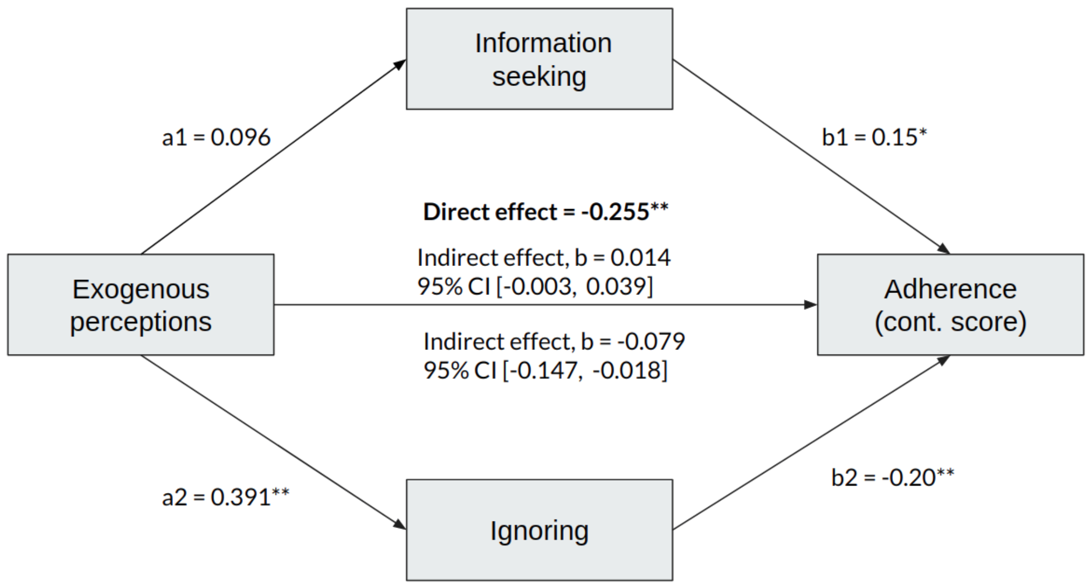

Figure 5.5: Conceptual framework of mediation effect of exogenous perceptions on adherence via coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

A further mediation model examined the relationship between discrimination and adherence to asthma mediated by ignoring and information seeking. Discrimination had a direct negative association with adherence (standardized beta = -0.341, SE = 0.093, 95% CI (-0.523, -0.159). In terms of coping strategies, discrimination had a significant positive relationship with information seeking (standardized beta = 0.336, SE = 0.060, 95% CI (0.217, 0.453), but no relationship with ignoring. The indirect effect of discrimination on adherence through information seeking was significant (standardized beta = 0.065, SE = 0.026, 95% CI (0.019, 0.120), see Figure 5.6; ignoring was not a significant mediator in this model (standardized beta = -0.025, SE = 0.017, 95% CI (-0.063, 0.003)).

We also examined the relationship between disparaging humour and adherence to asthma medication, mediated by information seeking and ignoring. Disparaging humour was negatively linked with adherence (standardized beta = -0.173, SE = 0.067, 95% CI (-0.306, -0.041)). In terms of coping mechanisms, disparaging humour was positively related to both information seeking (standardized beta

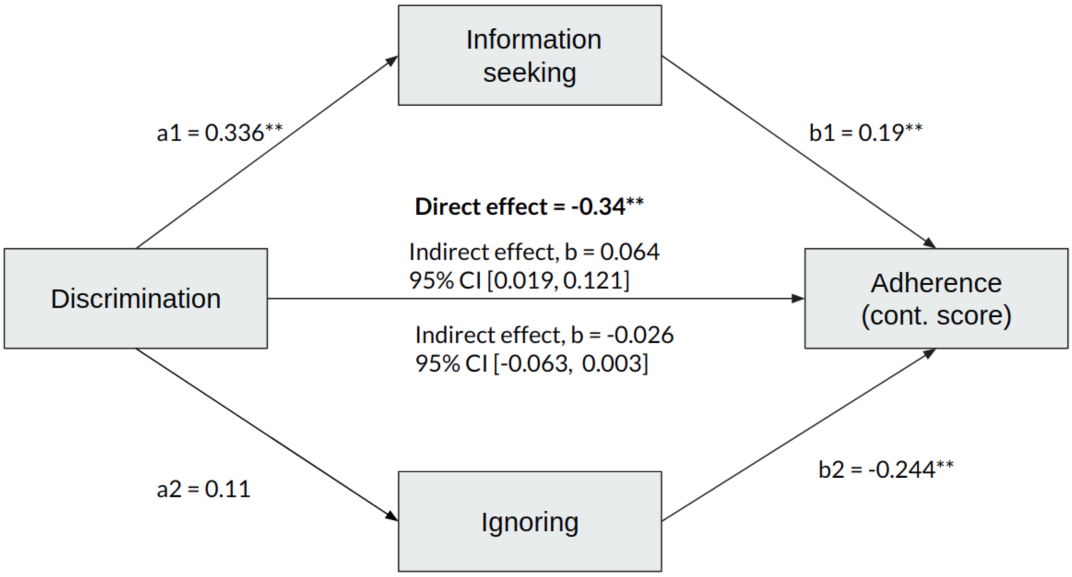

Figure 5.6: Conceptual framework of mediation effect of discrimination perceptions on adherence via coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

= 0.232, SE = 0.043, 95% CI (0.147, 0.318) and ignoring (standardized beta = 0.121, SE = 0.041, 95% CI (0.042, 0.202). The indirect effect of disparaging humour though information seeking was positive and significant (standardized beta = 0.041, SE = 0.017, 95% CI (0.009, 0.078)), and the indirect effect though ignoring asthma was negative and significant (standardized beta = -0.029, SE = 0.015, 95% CI (-0.063, -0.006)), see Figure 5.7.

5.4 Discussion

This is the first study that examined the complex and nuanced nature of stigma related to asthma, through investigating individual relationships that features associated with stigma have with adherence and how coping mechanisms aggravate or alleviate these relationships. Stigma has previously been described as a significant contributor to suffering by causing challenging stereotypes, perceived and experienced discrimination [15]. However, stigma, as a concept, can be broken

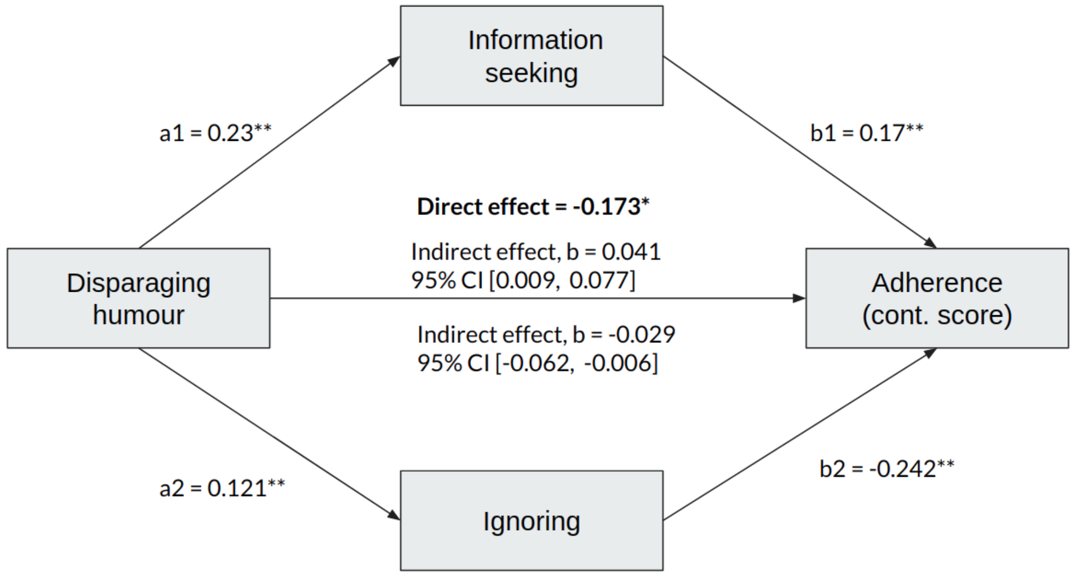

Figure 5.7: Conceptual framework of mediation effect of disparaging humour on adherence via coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

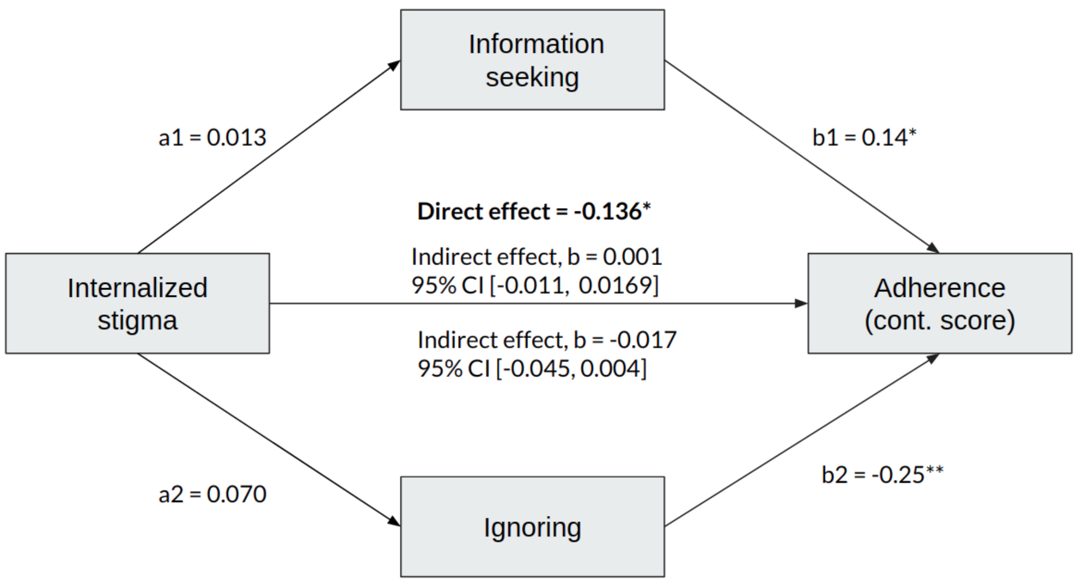

Figure 5.8: Conceptual framework of mediation effect of internalized stigma on adherence via coping mechanisms Note: **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

down into a number of features, including external dimensions - stigmatization originating from non-patients (such as discriminatory attitudes and actions) and patient-related or internal dimensions. These underlying mechanisms of stigma were less examined. This study, therefore, contributes to this body of research by assessing stigma dimensions and their power to predict adherence to medication and how these relationships can be mediated though information seeking and ignoring asthma as coping mechanisms.

The key findings of this research is that denial has the significant negative impact on adherence. Apart from denial, exogenous perceptions (e.g. anticipation of discrimination) and discrimination were significantly, negatively correlated to adherence. However, in the linear regression, with adherence as the output, their impact was not statistically significant. Both coping mechanisms had a small, however, statistically significant association with adherence. The mediation analysis found that higher levels of denial, exogenous perceptions and humour were related to a greater tendency to ignore asthma. Information seeking was negatively related to denial, but conversely - it was positively related to discrimination and disparaging humour. This implies a varied nature of stigma-related features, which means that future interventions could use different strategies for different stigma features, in order to be more effective in fighting stigmatization.

Previous findings about the relationship between denial of being asthmatic and adherence to asthma medication are mixed [70, 219]; our results align with previous qualitative work that claim that denial has a corrosive impact on adherence [123]. This negative relationship between denial and adherence can be understood though the assumption that people who deny having asthma do not perceive the need to use their medication. However, this is also why denial is one of the most 'dangerous' factors: when people do not accept they have a medical condition, this creates a significant barrier to getting appropriate medical attention, care and support.

It is also important to discuss the nature of denial and its causes. Denial could be interpreted as a patient-related factor (as opposed to discrimination which is related to actions of others). However, denial may be a interpreted as a *result* of patient's inability to negotiate the gap between their public and their 'asthmatic' identity. Specifically, denial can be seen as an attempt to minimize the effect asthma has on patients' social identity [231, 260]. This means that deniers would presumably avoid using an inhaler since that is the main signal of asthma existence. However, conversely, the regular adherence to medication is the optimal way of preventing worsening asthma symptoms and hence, obtaining 'normality' and 'healthy' identity. Some previous work even states that people with asthma are 'pushed towards' non-adherence simply by their interpersonal sensitivity [107]. This raises the notion that has been mentioned a couple of times in the thesis - future interventions and education strategies should not only be focused on patients, but also on changing wider societal perceptions that aggravate discrimination and disparaging humour.

5.4.1 Ignoring

Ignoring asthma could be considered to be similar to denial that a person has asthma. Our data shows that these constructs are positively related but not collinear. The main difference is that ignoring asthma involves acknowledgement that the person does have asthma, whereas denial does not acknowledge that the person has this condition. Unlike denial, ignoring was not previously suggested as negatively associated with adherence within the group of asthma patients [107]. However, both denial and ignoring belong to the group of inappropriate coping strategies, that have been found, by previous studies, as related to negative health outcomes [70]. Following the results of the mediation analysis, denial, exogenous perceptions and disparaging humour were found to be positively associated with ignoring asthma. Therefore, stigma appears to be associated with ignoring asthma as a coping strategy. This could be interpreted as creating a distance from asthmatics' identity. As discussed in previous literature, stigmatization creates a perception of 'weakness' that many patients do not want to be associated with, therefore, based on our analysis it can be assumed that a level of disassociation takes place in this process, presumably out of fear of being judged or being seen as an object of humorous remarks [82].

5.4.2 Information seeking

Information seeking could be interpreted as a normalization strategy, by which patients with asthma attempt to continuously identify the norms of the target group, so that they could minimize the differences and maintain their belonging to a 'normal' group [231]. This coping mechanism is also related to the perceived sense of community that exists online, where patients engage in sharing information and experiences [75]. This kind of emotional and information-sharing support that was also mentioned in Chapter 4 can be valuable for patients' well being, as it was documented in previous literature, even though this literature was often based on investigating 'offline' support groups [15]. The findings of this study also imply that there is usually a positive link between information seeking and adherence, which indicates that patients who seek information about asthma tend to have a higher level of adherence to asthma medication.

In terms of individual stigma factors, it is perhaps not surprising that denial of having asthma had a negative association with information seeking, since people who do not believe they have asthma, presumably, also have no need to seek information about it. However, it should be noted that perceptions relating to actions of others (discrimination, disparaging humour) have a *significant positive link* to information seeking. In other words, the higher level of stigma one experiences, the higher are their information seeking behaviours. It is possible that the exposure to stigma drives some patients to obtain information about their condition as a way of counteracting stigma effects, to feel more empowered or re-gain a sense of control. However, our analyses cannot rule out the reverse causal effect that may exist between information seeking and stigma. This effect could then be explained by a potential negative side effect of information seeking - and finding not only useful information, but also more stigmatization.

A large part of information seeking is likely to be conducted online, the same place where many people choose to express opinions that are not necessarily supportive of others. It is, therefore, possible that asthma patients can be exposed to negative views and stigmatization, and even bullying from others while they are looking for useful information about asthma [216]. In fact, some previous research has emphasised the negative outcomes of engaging in online environments, due to the ease of technology use and lack of accountability [134, 214]. These findings imply that future interventions should be mindful when choosing platforms or channels for communication with patients or when encouraging them to seek support in online communities - there is always a danger that the negative side of online environment could expose patients to further stigmatization.

5.5 Limitations and motivations for the future work

This work also faced some limitations. One of the factors that may contribute to the unrepresentativeness of the sample is the selection bias toward higher socioeconomic levels, considering the sample was collected using an online survey [213]. Another characteristic of this research is that patients tend to under-estimate their non-adherence to medication, which may have impacted our treatment adherence measure. Participants were also aware that the focus of research is on perceptions, which may have introduced bias in terms of how they respond to questions related to stigma. Lastly, since this is not a cross-sectional analysis, the established associations may have not been stable over time [155], which could be a focus of future research. Future research could also examine causes of denial, and other stigma-related features that negatively impact adherence, as well as the relationship between public and individual perceptions.

5.6 Conclusions of Study 3

This study followed the idea that stigma is not an atomic concept and that there is a range of stigma-related factors, where each has its own nature and ways they affect patients' adherence. Findings indicate that denial is a strong predictor of non-adherence, which is characterised by one's disbelief they have asthma. Our findings demonstrate significant relationships between a range of stigma factors and adherence to asthma medication. These findings have key implications for a paradigm related to adherence to medication that should be extended as it currently focuses mostly on the physical components of asthma [152]. We highlight that socio-cognitive perceptions also affect adherence and should be acknowledged accordingly.

This study also found that relationships between stigma factors and adherence can be changed though coping mechanisms: ignoring asthma was associated with both stigmatization and adherence to medication, which may represent disassociation that patients create in order to avoid stigma and judgements. However, information seeking had different associations with individual stigma-related features - while it was positively associated with discrimination and disparaging humour, it had a negative relationship with denial. Based on this, it can be argued that different stigmatization mechanisms may be affecting patients in different ways, driving them to adopt different coping styles and strategies. This supports the idea that better understanding of nuanced stigma mechanisms could lead to better designs of targeted interventions and development of a mutual language between health professionals and their patients. Future research could also further investigate causes of denial, as well as other stigma features.

Chapter 6

Using Model Class Reliance to measure group effect on adherence to asthma medication

> "It's taken me *years* to become like 'You know what, I've got it'. People are going to look, maybe it doesn't matter.' That takes a while to reach. If I'm confident and I don't care, and if I'm finding peace with myself with it, with my struggle, then it shouldn't matter. Perceptions shouldn't matter."
>
> \- *Donna (32)*

The previous work in this thesis implied that perceptions play a significant role in affecting one's level of treatment adherence. This claim was made following insights from qualitative interviews in Chapter 3, text analysis from Chapter 4, and based on examination of stigma-related perceptual factors and the impact they have on adherence through classical statistic analysis, in Chapter 6. The goal of this study is to make a full circle and re-assess the role of perceptions in one's treatment - through a holistic approach. Therefore, the main research questions that this study had was: *what is the importance of individual groups of known factors that affect adherence; and what is the value of perceptions in comparison to other groups*. Based on the work so far, we hypothesised that perceptions would have a leading role in predicting adherence. The resulting work conducted in this chapter has been accepted at the 2021 IEEE International Conference on Big Data, under the name 'Using Model Class Reliance to Measure Group Effects on Non-Adherence to Asthma Medication'.

As identified in Chapter 2, the motivation for this research was the gap in our understanding related to factors that drive adherence: while various drivers for non-adherence have been considered in isolation, interactions between demographic, socio-cognitive, behavioural and situational factors have never been modelled in concert. However, previous studies of this thesis also provided a basis for this Chapter. Firstly, work from Chapter 3 indicated, based on qualitative interviews, that one's perceptions are essential in their decision to adhere. Secondly, in Chapter 4, we could imply that perceptions (and especially, negative perceptions) about asthma may have some negative impact on patients. However, since the analysis was conducted on Twitter data, it was not possible to measure this negative effect, since there was no connection between individual tweets about perceptions and their levels of treatment adherence. Lastly, Chapter 5 indicated some collinearity between stigma-related features, in the data collected from the survey. This is relevant because traditionally used statistical techniques mostly focus on linear relationships and the application of traditional variable importance methods to examine explanatory factors is not possible. This is often due to insights being obfuscated by extensive shared information and non-linear interactions occurring across variables.

Previous work has also emphasised that insufficient attention has been paid to stigma and stigma-related features, that could play an even greater role in one's adherence, than previously examined features such as age, education or personality traits. In addition, this study represents the final, methodological approach that extends previous findings obtained from interviews and via classical statistics, by taking into account a more complex nature of non-linear relationships. This was done by introducing a first Grouped Feature approach to Model Class Reliance (Group-MCR), quantifying the importance of variable sets in underpinning explanations, in order to establish which groups of features are the most relevant in predicting adherence. In other words, this work is used to re-iterate the notion this PhD thesis mentioned though out this work: regardless of their often ignored nature - (for better or worse) perceptions *do* matter.

6.1 Research gap

Many drivers of non-adherence have been proposed: patients' use of medication may be impacted by demographics and lifestyle choices, as well as being highly influenced by contextual and psychological factors. Yet, there exists little consensus as to which of these groups of factors are *most central* to non-adherence behaviours [177]. Cross-sectional data is lacking, and while many covariates have been examined in an attempt to explain patients' reticence to medicate, studies have generally occurred in isolation, considering only one or few variables at a time. Additionally, studies vary in methodological approaches and results include some contradictory findings with respect to whether factors predict adherence to medication [89]. For example some studies found that personality traits have no impact on adherence [135], whereas other claim psychological traits such as neuroticism have a negative effect on one's treatment adherence [18].

Identifying which group of factors is the most relevant to non-adherence is valuable not only to practitioners, but to policy makers and those implementing health interventions. Understanding the interplay between potential drivers is key. Better identification of the categories of features most linked to non-adherence promises to provide not only a better understanding of the condition itself, but holds potential to generate actionable early warning signals, identifying those who may be at risk. Central to this is that different factors related to non-adherence require different approaches and different channels of communications with patients to act upon [177, 29]. For example, interventions may have to focus on reaching out to patients in rural areas, increasing education in poorer communities, or creating behaviourally focused therapies [29]. Therefore, understanding which group is dominant is essential, as this could inform more efficient interventions. Creating this insight could ultimately lead to a more proactive paradigm in medicine, better patient outcomes and lowering personal and stakeholders' costs [206].

6.2 Background

A wide range of approaches have, of course, been applied in an attempt to better understand non-adherence risk factors (see Section 6.2.2). However, traditional methods have struggled to cope with the non-linearity and high information sharing occurring between input features (the most simple version of which is multicollinearity). In particular, methods in social sciences and pharmacology have problems analysing group effects, with remedial solutions being prone to information loss (due to elimination of features that do not contribute to a 'compound scale' [162]). Yet, there also remain deficiencies in methods from Computer Science, with explanatory approaches such as grouped permutation importance providing insights only about the mechanism of a single model, rather then generative processes of the underlying phenomena - again, multi-collinearity and interaction effects can often resulting in misleading conclusions [125, 276].

To meet these challenges, this study takes a non-linear modelling approach, leveraging a dataset from Chapter 5 - data collected from a quantitative survey that had 511 participants. We additionally introduce the use of Grouped Model Class Reliance (Group-MCR). In contrast to group permutation importance, this extension is able to take into account multiple models to uncover underlying relationships between groups of factors and indicate each group's utility in predicting adherence. The rest of this chapter is structured as follows: Section 6.2.1 details related work with regard to risk factors, Section 6.2.2 highlights the challenges that exist with regard to applying recent/traditional approaches to achieve the aims of this work and Section 6.3.1 details the data set and the proposed methodology/experimental design, including the use of Group-MCR. Results are then provided in Section 6.4, followed by a discussion and conclusion in Sections 6.5 and 6.6 respectively.

6.2.1 Previously examined factors

While never assessed in concert, various potential drivers of non-adherence to asthma medication have been considered in isolation and they are described in more detail in Section 2.1.3 in the literature review. Many factors share a common thread, mostly speaking to patients' ability and motivation to obtain effective support or acquire knowledge, however, some are also related to obstacles that are outside of patients' control [17, 250].

An extensive stream of work, specifically considering asthma non-adherence, has focused on socio-demographic factors. Previous research, however, has not yet identified clear and consistent relationships between medication usage and demographic variables, such as gender and age [177, 135]. There is some evidence indicating that patients with better adherence behaviour tend to be female [39], from older demographics [29] and are most frequently married [135]. In addition, patients who report a higher social status [39], do not only adhere more strictly to their medication regimes, but are also reported to have fewer hospitalizations [241]. Despite these links being made, studies show that socio-demographics and clinical factors can only account for a small amount of variance in adherence [135] - and that additional other factors are relevant. In particular, lifestyle factors, such as smoking and lack of exercise are seen as negatively associated to adherence to prescribed medication [122].

The fact that socio-demographics and lifestyle features have only managed to yield partial explanations for non-adherence has increasingly led to the examination of the impact of factors such as personality traits. Individual differences such as personality traits are the focus of several studies on adherence, yet consensus is still to be achieved on what characteristics may relate to non-adherence. For example, neuroticism as a personality trait is found to have a negative effect, while conscientiousness and agreeableness have a positive effect on adherence [18, 282]; however, contradictory results state that personality does not have significant impact after all [135, 45]. Other psychological studies have considered the impact of the emotional states of patients, or affectivity. Negative affect, a common symptom of depression, has been correlated with low adherence to asthma mediation [20], a result supported by recent work evidencing that emotions are also predictive of *intentions* to adhere [159]. Other studies have focused on specific emotional states. Evidence of fears of asthma consequences, for example, have been strongly associated with reduced willingness to use medication [40] - whereas 'hedonic capacity', has been positively correlated with effective asthma control [20].

Less well examined, however, are the relationships between perceptual factors and adherence to asthma medication. This is perhaps surprising for asthma, which has been associated with issues of both stigma and avoidance, however, as previous studies state - in a limited capacity [19]. As previously reported in Section 2.1.9, patients decisions to adhere are potentially linked to patients' notion of their own personal identity, while maintaining an existing quality of life may dominate many individuals' experiences of asthma [89]. Perceptions related to asthma, and perceived 'weakness' some attach to the condition, also have effects on coping strategies [135], also described in Section 3.1.2.

Numerous potential factors described above, have been found to influence non-adherence and underlying drivers of non-adherence are deeply embedded in people's daily lives, ranging from behavioural to situational circumstances, psychological to demographic. However, it is relevant to note one of the key challenges of this kind of research - it is very possible that these groups of factors work together and even inform each other. In other words, the difficulty of this work is reflected in the fact that there may be a number of (potentially non-linear) interactions that exist across indicators. An example is age and stigma: as individuals age, their risk of hospitalization due to asthma increases; yet the stigma they anticipate when carrying a preventative inhaler in public decreases, reducing risk. Such interactions, and the range of mediating and moderating factors at play across patients' lives, are the motivation behind this study, which takes a non-linear and inductive approach to the problem domain. Leveraging machine learning techniques, this study models a new data set of 80 dependent variables to interrogate explanations of non-adherence.

6.2.2 Identifying actionable feature groups

Due to the complexity of the non-parametric approaches common in machine learning, derivation of stable model explanations is often non-trivial. A goal of this work, therefore, is not only to construct an accurate predictive model of adherence to asthma medication, but to use outputs to elicit robust insights into specific *groups of risk factors* that might inform future medical interventions [217]. To achieve actionable variable groups of this nature requires feature importance techniques. Importantly, research into model interpretability can also be dichotomised into two approaches in assessing the predictive performance of feature groups:

- *a priori combination:* features are combined prior to modelling. This corresponds to the tradition of feature engineering in machine learning and includes, but is not limited to, compression techniques such as PCA; Multidimensional Scaling; Kernel PCA; Maximum Variance Unfolding; and Partial Least Squares. All such techniques identify components (or latent variables) that reduce dimensionality - however, they are notoriously hard to interpret, restricting explanatory insights about an application domain [296].

- *posteriori combination:* feature combinations are assessed after modelling has occurred. Such approaches are particularly appropriate when input features sets are associated with specific groupings from the outset - either based on the theoretical or domain knowledge, or due to processes of data collection. Increasingly, improvements in algorithmic efficiency have reduced the need for exhaustive feature compression, allowing group importances to be assessed after the fact in both linear and non-linear models (e.g. via grouped permutation importance) [125].

The main focus of this research is to provide explanatory analysis of non-adherence to asthma medication in order to inform medical interventions to promote adherence. Use of *a priori* combination is, therefore, less desirable, due to the complications related to interpretation. Moreover, the dependent variables used in health research are often categorized (e.g. therapy-, condition-, patient-related), making a *posteriori* approach to group analyses more attractive.

A priori reduction of features into actionable sets is actually a common task in 'survey-based' research. Here, individual items are often combined into scales (constructed from variables that were designed to measure the same theorised underlying concept) with reliability assessed using measures such as Cronbach's alpha [311]. A high Cronbach's alpha supports the conjecture that features are interrelated to one another, and can be combined into a compound variable by taking a mean value, or sum [124]. While extremely common in social sciences research, this process has numerous drawbacks. First, features that are linearly correlated cannot be allocated to the same compound 'group', even if they link thematically. Second, items of varying types, such as categorical or continuous features, cannot be incorporated. Finally, Cronbach's alpha is only valid for data that is linear and normally distributed [311]. Worse, it is common for variables that cannot be grouped to be eliminated from modelling altogether, a situation that preserves the integrity of analysis, but at the cost of information loss and predictive accuracy - and consequently weakening the case for the validity of explanations produced [162].

Increasingly, machine learning approaches are eschewing a priori feature construction of this form, using approaches such as grouped permutation importances, performed on models trained on all original features [125]. Groups of features are simultaneously permuted to measure joint effects as their information is disrupted. In a similar fashion, feature exclusion and retraining methods[1] [106] include or exclude groups of feature and assess the significance of their inclusion in the model (although use of methods such as split counts and SHAP values [187] remains non-trivial with feature sets).

A major weakness of these approaches is that they only tell us about the mechanism of a single model - and not the underlying mechanism itself. It has, therefore, been noted that the explanations they output can be misleading regarding an underlying generative process [106, 276]. This is due to the fact that input features typically share predictive information (i.e. multicollinearity) - focusing on learning a single functional relationship, despite many possible, and equally predictive, relationships existing within the data set; this can therefore provide

[1] this includes step-wise approaches though these are more typically used for variable selection

deceptive results [106]. To address these issues, Model Class Reliance (MCR) has recently been proposed [276, 106]. This method implicitly considers the full set of models with equal predictive performance to the learnt 'best' model, known as its *Rashomon Set*, reporting variable importance bounds for each variable. This is particularly relevant for the field of current research, as it is highly possible that observed groups of features have intricate interactions.

6.3 Current research

Due to the goal of finding actionable explanatory insights for non-adherence, we use MCR within this study, extending and demonstrating its use in understanding grouped features. This work, therefore, examines two research questions via an inductive modelling approach: *What are the main groups of risk factors that predict non-adherence to medication?* and *Which individual features within these groupings are particularly representative of non-adherence?*. However, as previously stated, this is a distinct methodological challenge since indicators of non-adherence are not necessarily being mutually exclusive [177] and, in reality highly likely to interact with one another. Therefore, measuring the performance of predictive factors should be recognised as a multidimensional problem.

6.3.1 Participants and Materials

Data for this study was the same data used in Chapter 5. This means that data was collected via an online survey, and only participants from the UK, who were older than 18 and possess a current asthma diagnosis were retained as part of the study. However, while Chapter 5 only focused on the adherence score and the group of perceptional factors, this study examined additional groups of features that participants were asked about.

As mentioned in the previous study, adherence was measured via the Morisky-8 item questionnaire [208], which is composed of 8 questions with 'yes' or 'no' response options but producing a compound adherence score ranging between [0,7] (with higher scores implying higher adherence). However, in accordance with previous literature, scores can be further categorized into a binary adherence score using a cut-off point of 6, with a score < 6 corresponding to low adherence and a score ≥ 6 to high adherence [208]. The survey's other questions can be categorized into seven groups, typically measured using (1-7) Likert Scale (Table 6.1 details exemplar items from each group and includes some key research references. The full question list appears in the Appendix, Chapter 8.

- *Asthma traits:* characteristics of asthma and its symptoms were measured via: condition type (mild/severe); duration; inhaler type; and severity of symptoms [288].
- *Demographics:* Socio-demographics questions comprised of: age; gender; education; employment; and income (factors that previous work has identified as potentially indicative of adherence outcome [192, 39]).
- *Perceptual Factors:* based upon qualitative results from a variety of studies, perceptions measured included perceived stigma of having asthma [15]; perceived sense of support [216]; negative impacts of media and disparaging humour [108].
- *Coping mechanisms:* Coping mechanisms were examined using the adjusted Asthma Specific Coping Scale, which is composed of questions relating to: information seeking; hiding; ignoring; worrying about asthma; restrictive lifestyles; and positive reappraisal [1].
- *Emotional Affect:* Propensity for positive and negative emotions measured in the survey based upon the PANAS scoring system [306]. Emotions in

relation to asthma that were measured were: feeling sad, strong, scared, hostile, ashamed, nervous, determined, feelings of guilt.

- *Lifestyle Factors:* Items relate to lifestyle behaviours were collected, all relating to factors within patients' realm of control, with questions identifying habits such as smoking and exercise.

- *Personality traits:* Personality traits were collected using the Ten Item Personality Inventory (TIPI) version of the well-established Big Five: neuroticism, openness, extroversion, agreeableness and conscientiousness [74, 199].

A total of 511 valid individual responses were collected from the survey. 344 (67.3%) participants were female and 165 (32.3%) were male. 212 participants (41.49%) had high adherence, with 299 participants (58.51%) having low adherence. Missing values were handled using median imputation, and the issue of class imbalance was solved by down sampling [180], resulting in low and high adherence classes, both containing 214 data points (N=428).

Variable Group	**Description**	**References**
Adherence Score	Morisky-8 question adherence scale (0-7)	[208, 8]
Asthma Traits	Characteristics of asthma impact on the individual: • asthma severity (mild/severe) • duration of condition (years) • inhaler type (categorical)	[288]
Demographics	Age, gender, education, employment and income	[192, 39]
Perceptual Factors	Asthma perceptions influenced by exogenous factors • adjusted Stigma Scale • perceived sense of support • perceived effects of media and disparaging humour	[15, 216, 108]
Coping mechanisms	Asthma Specific Coping Scale	[1]
Emotional Affect	Range of features based upon PANAS scoring system	[306]
Lifestyle Habits	Identification of Patient controllable behaviours: • smoking • exercise	[122, 89]
Personality traits	Big-5 trait measurement scale	[74, 199]

Table 6.1: List of features groups, descriptions and exemplar references

6.3.2 Experimental Design

While the output of a classification task is a prediction from a number of finite categories, regression deals with predictions of continuous values. This means that in order to establish how effective the model is, different performance metrics are used in the case of classification and in the case of regression. The performance metrics for the regression task are focused on predicting the error, which summarizes how close prediction is to its expected value. However, in practice predictive models are often used to take an action, for instance to consider an intervention if an individual is predicted to be likely to non-adhere. In this case, the desired metric is not the error of the predictor but how accurate it is at predicting the categories which will drive the action. In the case of asthma analysis considered in this chapter it is the latter which is of interest and as such the task was cast as a binary classification task with non-adherence (the target class) coded as zero, following the previous work [148]. Data was stratified into a training (75%) and test set (25%), with the performance of each model examined against the held-out test data. Classification accuracy was used as the standard metric against which results are assessed.

More importantly than establishing trust and transparency between machine learning algorithms and scientists, it is, sometimes, even more valuable to understand the importance of individual features that drive the prediction [114]. Even though Neural Networks had a significant success in the previous work in this thesis (Chapter 4), the complexity of the model means a number of explanatory methods to understand the model/phenomena are not available. This is why, when the transparency of the underlying rules is secondary and the prediction power of a model is a primary goal (which was the case in Chapter 4), neural networks are a good choice. However, for this study, the main emphasis was on understanding the predictive power of different groups of features, which can be done using a novel technique that is currently only developed for a limited number of traditionally used algorithms (such as Random Forest). Therefore, for this study, three classes of more traditionally-used models (Logistic Regression, Support Vector Machines and Random Forests) were trained and evaluated.

Meta-parameters for each model class were optimized via a grid search and 10-fold Cross-Validation[2], with training data being further split to obtain a 10% validation set for each fold. Once an optimal set of meta-parameters was identified, the corresponding model was then re-fit to the full training data set and evaluated against the hold out sample. While nested cross-validation could be employed, a split sample approach was preferred here due to the explanatory focus of the analysis and the subsequent goal of isolating an optimal reference model against which variable importance analysis could be performed. However, to confirm that the optimal reference model found was representative of generalized classification accuracy, the experiment was re-run in its entirety ten times to ensure the representativeness of the reference model's performance.

With an optimal model now in hand, factors underpinning non-adherence were examined through the lens of feature importance analysis. Due to the grouped nature of the feature sets used, risk factors were assessed, via permutation importance, within their typed groups. In contrast to the proposal in [125], where group permutation importance scores are divided by the number of features in each variable set (thus 'normalizing' them), this study reported only total increase in error. This is motivated by the fact that our setting is not tasked with *minimal feature selection* (as per [125]), but addresses the challenge of *maximal information collection*. That is, if practitioners must focus their data collection efforts on one pre-defined feature set, whether psychological, behavioural or physical, it is crucial to identify which ought be prioritised. This distinction is key, as the minimal feature selection problem favours inclusion of smaller groups of features

[2] See Chapter 8 for full details.

per step - even if such groups contribute less predictive power than a larger group overall. Thus, for present problem setting, only total mean decrease in accuracy for permuted feature sets was assessed.

Permutation importance can lead to misleading results if practitioners are unaware that outputs only encode a single functional relationship between input feature and response feature - which is not necessarily representative of feature importances in the underlying phenomenon. MCR was introduced to help address this problem [106], requiring that analysis considers not one arbitrary solution, but all models that can provide optimal predictive performance. By computing MCR scores [276] for this Rashomon set, we identify those features that contain indispensable information, and those which are readily substitutable for another feature. As well as moving our understanding towards explanations of the underlying phenomenon being modelled, rather than the workings of an individual model instantiation, this also mitigates against any feature selection bias that might occur during model building.

6.3.3 Grouped Model Class Reliance

To this end, not only is the individual MCR analysis applied in this study, but the existing MCR methods are also extended, by adding functionality that allows analysis of groups of features within the framework (echoing the addition of grouping mechanisms to naive permutation methods developed in [125]). Group-MCR is realized though a modification of the RF-MCR algorithm made available in [276] RF-MCR implicitly constructs a Rashomon Set from a reference model and then determines MCR+ and MCR- for each feature. For MCR+ (MCR-) the algorithm considers a feature of interest, v_0, and derives a forest where v_0 is maximally (minimally) relied upon.

To locate this forest, two transforms are computed using the reference model at runtime. First, for each tree in that model, transform (1) constructs a set of structurally equivalent trees by exhaustively learning surrogate-split features. These are then substituted where possible to force (avoid) the use of v_0 at each decision node (assessing impact on performance using a permuted version of v_0). This allows us to find the tree for which v_0 is maximally (minimally) relied upon, storing the mean decrease in accuracy (MDA) of permuting v_0 as a measure of this. This MDA score is then input into transform (2), which uses pre-computed lists of tree 'prediction equivalences', to construct a new forest that produces exactly the same predictions, but is maximally (minimally) reliant on v_0. For more information and proofs of the validity of this approach see [276].

This algorithm is extended in this work so that it can consider feature groups (i.e. $v, ..., v_n \in G$) rather than a single feature of interest (v_0). This requires modification of the transform (1). Specifically, when computing MCR+ (MCR-) we must assess the MDA of structurally equivalent trees. Instead of forcing use of v_0, we consider use (avoid) all elements in G, retaining the one that maximises MDA for the tree (and selecting an arbitrary if multiple equivalent options are available). Following this procedure, transform (2) continues as before, with the proofs provide in [276] unaffected, resulting in identification of Grouped-MCR+ and MCR- for the G.

While the proposed Grouped MCR provides an understanding of the indispensability of both individual variables and groups established, Shapley additive explanation (SHAP) values aid variable interpretability, shedding additional light on the relationship that exists between adherence score and input features [111]. While no method to extend SHAP to groups or MCR currently exist and are beyond the scope of this work, such an analysis remains of interest and so SHAP results are presented when focusing on the effect of individual factors within groups indicated to be of interest by the MCR analysis.

6.4 Results

Following parameterization via 10-fold cross validation of Logistic Regression, Support Vector Machine and Random Forest model classes, and evaluation against the held-out test set, Random Forest models produced best predictive performance, with classification accuracy of 71% (precision = 0.73; recall = 0.69); SVM produced the second best result with the accuracy of 64% (precision = 0.67; recall = 0.56). Logistic Regression Classifier produced an accuracy of 63% (precision = 0.67; recall = 0.52). As the optimal classifier, the Random Forest model was first analysed using unconditional permutation importance [13], examining the importance of various feature groups to its predictive mechanism. Results of permutation importance are illustrated in Figure 6.1a (left). Perceptual factors provided the strongest influence on successful predictions of non-adherence, followed by emotions. Lifestyle choices and coping behaviours were of moderate importance to the model, with other features of only minor relevance to model outputs.

Next, Group-MCR analysis was applied, with Figure 6.1b (right) illustrating the arbitrary nature of raw permutation importance analysis. Group-MCR analysis takes into account all equally predictive models from the model class, with importances boundaries being notably at odds to results in Figure 6.1a. First we note that, as expected (and by definition), the results of analysis of the reference model lie within the Group-MCR bounds. However, the Minimum Model Class Reliance (MCR-) and Minimum Model Class Reliance (MCR+) scores imply that there are clear differences between competing 'optimal' models in terms of the variable groups they leverage. This indicates that information sharing is occurring between the groups - in some instances feature groups are reflective of the same information, and could be use interchangeably. The only exception to this is the group of Perceptual factors, which not only contributes the most to any optimal model, but contains information (as indicated by a non-zero MCR- score)

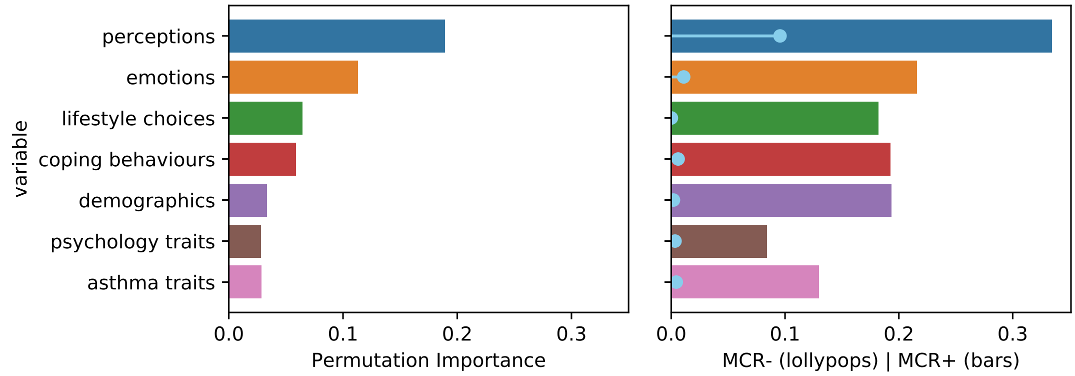

Figure 6.1: Comparison of unconditional permutation importance (left) and Group-MCR (right).

that is indispensable to prediction of adherence.

The perceptual factors, which showed strongest import to the model class, were then considered at an individual feature level via RF-MCR analysis. Individual features here correspond to individual questions from the survey, and have been ranked in Figure 6.2 with respect to their importance ranges across all models. Immediately, we recognize that much information is shared across questions, as one would expect given they all relate to patient perceptions of their condition. However, statements reflecting *Denial* are noticeable in achieving the highest available MCR+ levels: *'Even though I am diagnosed, I think I may not have asthma'* and *'My asthma is not as serious as my doctor and my diagnosis say it is'*. Another notable perception is *Perceived Discrimination*, as statements that illustrate this perceived stigma had a high MCR- and MCR+ score (e.g. *'I have been discriminated against at work because of my asthma'*).

In order to provide an indication of the relationship that exists between adherence scores and individual perceptions [111], Shapley additive explanation (SHAP) values were also computed based on the reference model. Figure 6.2b illustrates SHAP values for 28 individual features within the perceptions variable set, each item reflecting an individual question from the survey. Interpretation of these

SHAP values indicates that predominantly, negative perceptions, indicative of discrimination and denial, correspond to low adherence scores (note that in the figure negative perceptual factors are predominantly indicated in red - but this is dependent on exact wording of individual questions).

6.5 Discussion

Results have shown that Perceptual factors are the most important category of predictors indicative of non-adherence. This conclusion is supported by both permutation importance and Group-MCR methods, with the factors' high MCR-score indicating that the information contained by perception features cannot be replaced. Even in the predictive model where they are needed the least, they still play a key role. These observations are supported by previous work in this thesis, and ideas mentioned in the literature review which report that a patient's mental map of their illness/medication is the basis for a constant feed back loop with their behaviour [44]. Group-MCR analysis of all other feature groups showed information overlap, with all other groups being almost wholly exchangeable by another feature category. However, most importantly, consideration of perceptions in understanding asthma cannot be considered a side issue - they are essential and should be taken into account in all models.

In retrospect, these results are intuitive. Perceptions relate to people's interpretation of their world, including how they see themselves and their condition, which in turn informs their actions. Perceptions also have an indirect effect on behaviour, with negative perceptions having been evidenced as leading to higher rates of anxiety and depression [53], which are in turn linked to non-adherence. This group of factors is also of particular importance because once established, health-related perceptions are not easily altered [29]. Results of this work have an

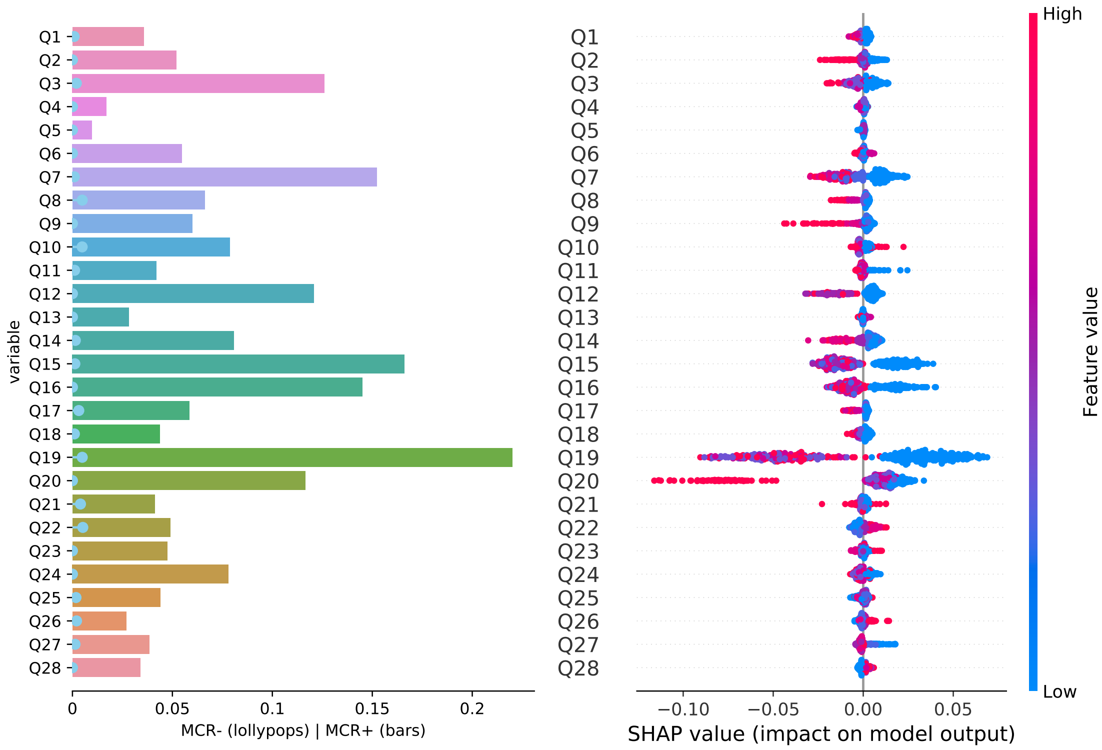

Figure 6.2: Perception questions (features) only: MCR Plot (left) and a SHAP Summary Plot for a single, arbitrary, model from the Rashomon Set (right). Higher adherence is indicated by positive SHAP values, and non-adherence by negative SHAP values.

Q1. People sometimes joke about asthma to put me down.
Q2. People would make jokes about asthma in front of me.
Q3. I do not use my inhaler in public because people might make fun of me.
Q4. Media generally portrays asthma in a negative light.
Q5. Society perceives people with asthma as weak.
Q6. People would treat me differently if they knew I had asthma.
Q7. I have never been discriminated against at work because of asthma.
Q8. Sometimes I feel that I am being talked down to because of asthma.
Q9. I worry about telling people I use an inhaler.
Q10. Some people with asthma are weak.
Q11. I do not think people understand what asthma really is.
Q12. I think people with asthma are not as reliable co-workers as anybody else.
Q13. I worry about how people might react if they found out about my asthma.
Q14. I would have had better chances in life if I had not had asthma.
Q15. I do not tell people at my workplace that I have asthma.
Q16. I prefer if people did not see me using my inhaler.
Q17. I am angry with the way some people have reacted about my asthma.
Q18. I had trouble with people because of my asthma.
Q19. Even though I am diagnosed I think I may not have asthma.
Q20. My asthma is not as serious as my doctor and my diagnosis say it is.
Q21. Using an inhaler means you are not coping well with your asthma.
Q22. I would not ask others to change their behaviour...it is my own problem.
Q23. I get valuable information on how to cope with asthma from people online.
Q24. I feel more understood about asthma problems by people online...
Q25. I don't think that support groups for asthmatics are of any use to me.
Q26. I am concerned that I might get incorrect information...online.
Q27. Realizing that my experience is not unique helped me cope better.
Q28. I would rather suffer from smoke than explain that I have asthma.

implication for future intervention strategies indicating they are likely to benefit from greater focus on actively addressing the reputation of asthma, which continues to be caricatured as a sign of 'weakness' in the TV/film media. Finding this particular group of factors as the most relevant is important because perceptions are often exogenous, as mentioned in Chapter 5. This means that they may not only originate from patients, but also from the public and can be constantly reshaped based on public (and other people's) opinions. Therefore, unlike for some other groups of factors that focus on patient-characteristics, such as demographics, interventions aimed at targeting perceptions should not only focus on patients regarding stigma, but also the general public as well [15].

However, apart from stating the relevance of perceptions, this study also gives rise to other findings which have not been addressed in this thesis before. Specifically, group-MCR analysis highlighted the existence of models in which some feature groups have MCR- score of 0 (e.g. lifestyle choices). This insight is important as it emphasizes the dangers of only considering results of permutation importance of a single model. Permutation importance (Figure 6.1) implied lifestyle choices could be considered as the third most relevant category of features. Yet, Group-MCR demonstrates that this is not necessarily representative of the phenomena and underlying generative process; there are models that simply do not need to use this group and can still achieve the same prediction accuracy. The difference in terms of MCR- and MCR+ scores indicates that groups have intertwined relationships and likely share a significant proportion of information between each other. Instead of focusing on changing patients lifestyle, it is likely more impactful to focus on changing attitudes and sentiment in the first instance. The differences between graphs in Figure 6.1 also emphasize that when features share information, ranking of features can be highly arbitrary - demographics, for example, was an unimportant group in the reference model, but achieved the third highest MCR+ score.

In the breakdown of individual perception importances (see Figure 6.2), RF-MCR finds strongest predictive power to be reflected in the question: *'Even though I am diagnosed, I think I may not have asthma'*, illustrating *denial* of the condition [219]. SHAP values associated with statements about denial show the negative association denial has with adherence. This echoes previous findings from Chapter 5, which have identified denial as one of the most 'dangerous' aspects of stigma; whether it was the result of mental health challenges, or simply an inability of a patient to accept the identity of being an 'asthmatic' (mentioned in 2.1.9), denial was mentioned as a significant barrier on the path of getting appropriate help and support [29].

Many individual perceptual factors have a very low score for MCR- (again see Figure 6.2a). This is to be expected, as many questions in the survey are related to each other - and could potentially be used interchangeably, due to high information-sharing between them. This view is further supported by observed SHAP values (Figure 6.2b) which indicate that a majority of features have a negative impact on adherence. The influence of perceptions, such as denial, to negatively impact adherence has an implication for intervention design, which currently remains focused on the physical components of asthma and related conditions [152]. That said, limitations within our study still exist, and it important to acknowledge the constraints on generalization of present findings. While the list of features (and groups) included in our model is based on hypothesized predictors from the literature, this is not an exhaustive list; increased data set sizes would better serve this analysis; and a range models, for which RF-MCR is not available, are yet to be considered.

6.6 Conclusions of Study 4

This work allowed investigation of combined groups of features, which have been independently proposed as related to asthma non-adherence. Considering intricate interactions between different features, their group's utility was considered in predicting non-adherence, using a novel Grouped Variable approach to Model Class Reliance (Group-MCR). This method overcomes challenges such as masking issues due to multicollinearity and the non-linear nature of relationships occurring between variable groups. Group-MCR Analysis also highlighted the risk of assuming that feature importances derived from a single 'optimal' model of non-adherence are fully representative.

Results indicated that perceptual factors were the strongest predictor of non-adherence. Responses to asthma must not be considered a pharmacological challenge alone, but also one of managing patient beliefs. A more detailed examination of individual features revealed that both denial and perceived discrimination play a crucial role in non-adherence to medication regimes. These insights can be used to develop better markers for non-adherence in the future, and open potential routes to tailored services, that can target patients most at risk. The indispensability of perceptual factors produces a clear recommendation, that policy should attempt to reduce both stigma and discrimination surrounding the condition, focusing on allaying denial and patient fears.

Chapter 7

Discussion

> "Akira Kurosawa's 1950 film Rashomon presents four different accounts of a contested event—the murder of a Japanese nobleman and the rape of his wife. As the events are retold from four different points of view, the viewer is left wondering which of the four witnesses was telling the truth and whether a single "truth" really exists. The film makes it clear that there are different truths for these characters, for they are not simply lying to protect themselves (in fact, each main party's version of events implicates the teller for the murder); rather, they have deceived themselves into believing the version they have told. These same questions about truth might be asked about contested events in social research. When multiple sources relate different and sometimes conflicting accounts of an episode, how do we decide who is "right"? Is it possible that they all are right?"
>
> - *(Roth, 2002, p.131)*

This work undertook a combination of qualitative research, traditional statistics and machine learning in an investigative role in dealing with two overarching tasks of this thesis - 1) the examination of perceptions that exist about asthma (in Chapter 3 and Chapter 4); 2) the assessment of their impact on adherence (in in Chapter 5 and Chapter 6).

Qualitative research proved its exploratory value in Chapter 3, in the identification of internal and external perceptions, providing a rich context for exploring perceptions related to stigma (e.g. the unwillingness of patients to disclose they have asthma was only mentioned in interviews, and not in tweets). However, big data in the form of tweets in Chapter 4 demonstrated it was even more versatile in examining perceptions. Big data's unbiased representation of perceptions about people with asthma uncovered the prevalence of perceptions that are either less openly spoken about or unconscious - particularly, stigma about asthma that people choose to not discuss in interviews and the idea of a community for asthma patients that exists online. Finally, the state of the art predictive power of linguistic features and neural networks used to detect perception in Chapter 4, highlighted additional reasons why big data can have a valuable application in psychology and be a useful addition in a mixed-researchers' toolbox. However, big data analytics also proved to have their own limitations. Tweets contain a lot of noise, which makes uncovering perceptions harder. Additionally, using big data opened new points of discussion about the biggest pain point of this and similar research in the field: the challenge coined as data poverty. The benefits and limitations of each of these research approaches are summarized in Table 7.1.

The second part of the thesis, Chapter 5 and Chapter 6 focused on the assessment of the impact perceptions have on treatment adherence. Firstly, to fill in the gaps about negative perceptions about asthma, Chapter 5 showed that traditional statistics provide a valuable measure of linear associations - mediation analysis uncovered that stigma has a complex underlying set of mechanisms by which it affects different coping strategies that patients adopt. This work also highlighted that stigma-related features originating in the society could play a strong role in affecting adherence, as patients may want to 'fit in' to societal norms by denying they have asthma. To re-iterate the value of perceptions, the final study, Chapter 6 offered a unique holistic approach in uncovering which set of factors is the most important for non-adherent patients, by focusing on non-linear relationships in order to account for potential masking effects from Chapter 5. Results of this study demonstrate that patients' perceptions, which are in nature inseparable and arguably informed by the societal perceptions, are the group of factors of highest importance to non-adherence. Intentional treatment non-adherence does not only have roots in the minds of patients, but also in the wider societal context, which should be taken into consideration when fighting negative perceptions about asthma, particularly stigma. Furthermore, the effectiveness of machine learning adds value to qualitative and traditional quantitative techniques. In conclusion, mixed methods, albeit with a careful appreciation of each method's limitations, may be the optimal solution for examination of complex and multi-dimensional topics such as treatment adherence and perceptions.

7.1 Internal and external perceptions

At the start of this work, a careful consideration had to be made in relation to the choice of the approaches used. Even after taking into account the major benefits and limitations of each of the methods, examined in Chapter 2, the choice of the method used is strongly dependent on the nature of the topic that is being examined. As it was previously stated in Chapter 2, both perceptions and treatment adherence are unique topics and a special formulation of these topics needs to be made, since this dictates what kind of data is needed/available in the analysis (4.1). Perceptions were defined in this thesis as socio-cognitive factors that fall under the umbrella of social concepts, which can be explored using both qualitative and quantitative approaches. Adherence to asthma medication, as a topic, belongs to the groups of behaviours that need to be observed, in order to be measured, which usually involves practical difficulties. Adherence to medication is similar to other behaviours that are notoriously hard to measure - predominantly due to the privacy factor, but also due to a lack of techniques of directly observing adherence (which is why self-report is the most widely used technique). This topic, therefore, provides less flexibility in terms of research approaches.

Considering the particularity of perceptions as a topic, Chapter 3 approached their investigation using qualitative research, as the first method in this mixed method research. Being previously praised for its versatility in exploratory research, and providing more context than standard survey instruments, qualitative work in Chapter 3 was deemed as appropriate for investigating perceptions. Without a formal theoretical background, a range of interviews examined perceptions that asthma patients have about life with asthma. The resulting set of perceptions was categorised as either *internal* (perceptions about asthma, self-image, asthma management and impact of asthma on everyday life) or *external* (fear of disclosure, lack of understanding, others' perceptions and others' reactions).

The relevance of these results is reflected not only in the useful identification of the main perceptions held by asthma patients, but also in delineation between perceptions patients had about themselves and their condition (internal) and perceptions that were related to interactions (or anticipated interactions) with other people (external). The external perceptions are usually related to how patients *believe* others see them, in the light of their condition. The relevance of external perceptions is particularly noticeable when it comes to stigma, which was in this thesis suggested as a group of negative perceptions. To illustrate, patients were afraid to disclose they have asthma, out of the fear of being seen as less competent (at their workplace) or simply - weaker than their peers. Similar stigma-related sentiment could be noticed when patients spoke about the anticipation of discrimination due to others' perceived lack of understanding. This classification of perceptions to internal and external, in Chapter 3, also carries a signal that stigmatization has a complex nature - a notion that was further examined in Chapter 5.

Delineation between internal and external perceptions, a resulting finding from Chapter 3, can be interpreted though the attribution theory: external attribution is related to outside forces that motivate the event, whereas internal attribution assigned the causality within the individual [227]. However, a further distinction can be made in the field of external perceptions - between perceptions that patients think other people have about them (which are considered in interviews in Chapter 3) *and* perceptions *actually* held by other people - which are similar to public attitudes in previous work [16]. Public perceptions were the focus of qualitative analysis of tweets in Chapter 3 and Chapter 4 - and they have been, in the work about attitudes about mental health, defined as reflection of cultural conceptions that create the reality of people involved in that culture (that also consequently affects their behaviour) [16]. This work found some important similarities between perceptions held by patients and publicly expressed perceptions held by other people, through the convergence analysis in Chapter 3.

To illustrate, some interviewed patients from Chapter 3 claim that they see other people with asthma as ‘less competent’ and ‘weak’. A similar discriminatory notion was detected in assumed public opinion obtained from tweets in Chapter 4, where asthma patients were objects of jokes and called derogatory names and represented as ‘weaker’ than non-patients. This particular similarity emphasises the level of internalised stigma held by interviewed patients - and even though causality cannot be claimed, it can still be argued that some patients’ perceptions about other asthma patients may be *informed* by the public opinion.

Furthermore, the inclusion of individuals in the social environment might also be related to how much they align with the prevailing social norms. For example, even when stigma is present, some patients may not be susceptible to it due to their own perceptions and conversely, some patients may *feel* stigmatized even when there is no real proof stigmatization is happening. Additionally, some patients may not be directly subjected to negative perceptions if their medical condition is concealable (which is the case with asthma). It is, still, worth noting that some public perceptions could be completely disassociated from the personal experience of stigma - as interviews stated, many perceived stigmatizations are simply anticipated and are not based on real interactions. However, even these perceived perceptions have an origin, that could be examined in the future work.

Based on this analysis of patients' and 'public' perceptions highlights that it is very hard to distinguish between effects of external and internal perceptions, since a patient is a part of a much larger social circle (to which the individual, herself, may be contributing in terms of perception creation). This adds to the complex nature of perceptions and the idea that using a single approach in investigation of perceptions - may be insufficient and simultaneously justifies the approach of combining qualitative approach and big data analytics to get a fuller picture.

7.2 Qualitative work

The findings mentioned above are closely related to the methodological challenge that was one of the implications of the work in Chapter 3 and Chapter 4. As it was mentioned, regardless of the value of qualitative work in conducting an exploratory analysis, there were some perceptions that were hard to obtain through interviews. Namely, asthma patients - regardless of the level of stigma they experienced, initially did not disclose they face any stigmatization (when asked about it directly). Therefore, it was extremely hard to examine stigma using interviews from Chapter 3. This could be interpreted as a dual challenge - since patients were either not conscious of stigma or they did not want be seen as different from non-patients and, hence, judged or marginalized.

This particular issue can be a significant obstacle when qualitative research precedes quantitative. This is because, even though exploratory analysis may be appropriate in theory, some topics may be too hard to investigate using the tools of qualitative research. As a result, findings may wrongly inform the following steps of the analysis. For example, stigma was detectable in interviews only through examples patients made, relating to their everyday life experiences and participants shared no insights when asked directly about potential stigmatization in their lives. In Chapter 3, participants spoke about their unwillingness to disclose they have asthma at their workplace, reflecting anticipated stigma. Conversely, Chapter 4 implied that stigma about asthma is not only present, but also one of the dominant topics when asthma is discussed in a non-laboratory space (such as Twitter).

Similarly, the sense of community was detected as one of the leading groups of perceptions from Chapter 4, but was hardly mentioned in interviews - presumably because even if patients felt supported, this perception is likely to remain undetected simply because patients see it as part of their everyday life. Based on the research process from this thesis, it can be argued that in the case of vague, undesired, stigmatized, ill-defined or unknown topics, there is a strong chance that qualitative research would not provide complete results. Therefore, as an implication for the future mixed method research - the nature of the topic should be carefully considered when deciding on the *order of methods* in the mixed method approach. The order of methods in the mixed methods approach should be driven by the nature of the topic, as much as by other design factors. In this thesis, the order was 'developmental', which means that qualitative research was used to identify perceptions and create a conceptual framework - that was then used to lead the quantitative work focused on further theory testing. Therefore, on the example of this work, if stigma had not been reflected in rare examples, it would not have been investigated in the stages that followed, and a research opportunity could have been easily missed.

	Benefits	Limitations
Qualitative Research	- Exploratory analysis - Insights about the context	- Findings not generalisable - Impacts of features on the output cannot be quantified - If participants are unaware of something they cannot disclose it
Traditional Statistics	- Theory driven - Provides interpretable results	- Prone to biases (e.g. sociability bias; demand characteristic) - Usually small sample sizes - Many contradictory findings in the field - Possibility of overfitting - Only linear relationships taken into account
Big Data Analytics (Data Science)	- Naturalistic setting (not prone to bias) - Takes into account both Linear and Non-linear relationships - Can use secondary data (data collection process is less expensive)	- Does not provide a wider context - Incomplete view of one's life - Usually hard to link different big data sources (no impact can be measured) - Noise in data - 'Black box' models - no interpretation - Not easily available

Table 7.1: The summary of benefits and limitations of the methodological approaches used in this thesis

7.3 Big data in psychology

One of the greatest advantages that big data brings to research is related to its unobtrusiveness and less bias in data collection. This was particularly evident in Chapter 4 when the resulting four groups of perceptions uncovered topics that were not mentioned in interviews. As an example, there was a topic that describes the 'Frustration with patients' (expressed in tweets by presumed non-patients), in the group of active negative perceptions. Similarly, there was a topic that contains self-pity, named 'Sadness due to limitations', where people complained about the poor quality of life due to asthma. These examples illustrate that big data, such as tweets can be a place where people express their true perceptions about asthma and themselves more openly than in case when they are openly subjected to research participation. This is particularly the case with socially undesirable perceptions. For example, patients may choose to not portray themselves in interviews as someone who is riddled with self-pity due to their medical condition, or as someone who is stigmatized. Twitter is a more naturalistic setting where perceptions are freely expressed; there are less biases introduced by the researcher and there is a much larger quantity of data that contributes to avoiding the selection bias-distortion. These characteristics of tweets illustrate the value that big data may bring not only to examining perceptions, but also other, similar concepts from the field of psychology.

However, there are some limitations to big data that can be acknowledged, based on the work in this thesis. As said before, machine learning is focused on the prediction and detecting links between variables, however, this method rarely offers any explanation as to why connections between variables exist, which is a common complaint about the inductive approach [30]. Therefore, a challenge, and a significant limitation of using big data in the form of tweets in this thesis was related to the fact that big data did not necessarily provide the context for the topic of interest. To illustrate, when tweets were analysed, it was impossible to know whether the author has asthma or not, unless it was explicitly stated in the tweet, which significantly changes the overarching conclusions. As an example, a tweet: ‘Asthma sucks’ could signify an act of discrimination - if written by a non-patient or it could signal self-pity - if written by an asthma patient.

In the context of this work, this limitation meant that no definitive discussion about the relationship between patients and public perceptions could be made. We can only discuss similarities and differences between the two sources. Stigma was detected in tweets, however, a deeper understanding about *how* stigmatization affects patients and how it is represented in the everyday, off-line events, could only be elaborated on through interviews - and even then, only mainly through ‘accidental’ examples. For example, one theme from interviews - the issue of disclosure about asthma, was not mentioned on Twitter, presumably because mentioning the issue would, in itself, represent acknowledgement that one has asthma (which is the opposite of non-disclosure).

Another challenge related to big data use is the bias in terms of the incomplete view of someone's life. This is an issue similar to the lack of the context. It reflects the incomplete view of one's behaviour, since big data analytics usually does not take into account behaviours a particular person expresses in different contexts (for example, off-line or, in the case of transactional data, when purchasing items in another shop). In this vein, public perceptions in this thesis were considered as perceptions expressed on Twitter - however, there is a chance that public perceptions are expressed differently on another medium. Furthermore, even though big data is not affected by standard biases (e.g. desirability bias), there can still be biases related to the image that social media users *choose to portray* about themselves, in order to maintain their image on social media, which, needless to say, may be different to their true self.

Additionally, Twitter's audience is younger than the general public and this could potentially mean that (even when they are still valuable), perceptions expressed on Twitter may not represent the general public's views on asthma. This was recognised as one of the limitations of this thesis. This also means that some machine learning results could precede another round of qualitative work, that could bring more context about findings and elaboration on the nature of newly found predictions. This is also another illustration as to how using a single lens (a single method) could lead to incomplete or even skewed results - and demonstrates the synergy that mixed methods can bring to alleviate these issues.

In terms of the analysis in Chapter 4 it was also noted that big data also means 'big noise' in the data. The data used in this study consisted of informal, colloquial language, which, in itself, presents challenges to data analysis. This means that even though neural networks provided valuable accuracy results, it is still useful to recognise that each step has introduced another layer of noise in the results. Additionally, unlike linguistic features that were used Chapter 4, in the task of classifying tweets as perception tweets or non-perceptions tweets, neural networks provided no interpretation of results, which could have served as a valuable check of the results. Lastly and most importantly, one of the biggest limitations of big data, that has also been reflected on the work of this thesis, was - the general lack of it, or simply called, ‘data poverty’.

7.4 Data poverty

As mentioned, there are many benefits that big data analytics brings, including the variety of sources and forms that big data can be in. Only some of these include numerical, text, images, sounds, etc. In short, there are now many ways to quantify various areas of life and collect information about people and their behaviour in an easier and less expensive manner [265]. The Medical field is no different - big data provided the promise to transform the industry by raising awareness about the value of data in scientific and medical research [197]. This was particularly relevant with the appearance of e-health, telemedicine, m-health and other devices that were able to quantify our every day health-related behaviours [30].

However, more than anything else, the use of big data has the ability to provide a complete shift in the paradigm of medicine - it could not only make the treatment of patients more personalised, but also more proactive: patterns in big data could provide the opportunity to capture the warning signs and early detection of disease outbreaks [301] and help patients to take control over their condition management. This was recognised as a strong opportunity for the work in this thesis as well, which resulted in the use of big data in Chapter 4. However, there were other opportunities that were recognised, such as collecting treatment adherence data from e-health apps, that could have been a strong addition to this work. Unfortunately, this research had no access to medical data from NHS or medical health apps. As a result, the work conducted in this thesis relied exclusively on publicly available data (Tweets used in Chapter 3 and Chapter 4); primary data obtained through interviewing patients (in Chapter 3) and data obtained through a survey (in Chapter 5 and Chapter 6). This led to challenges when assessing the impact perceptions and other factors have on adherence, as it was extremely hard to make a direct link between perceptions expressed in 'the wild' (on Twitter) and one's level of adherence.

In many ways, the data is there, but not everyone can use it [167]. Obtaining data (including big data), for research reasons, has mostly been associated with challenges of high cost and the unavailability of resources for data sharing [251]. Many research areas are still described as data-poor, regardless of continuous initiatives to make secondary data available to researchers [202]. Some issues are ethical ones and related to integrity, trust, security, infrastructure, respect for patients' freedoms and right for privacy, reputations and regulations [30, 251]. However, some of the most prominent ethical considerations have been dedicated to privacy consequences and security risks, as well as the challenges of aggregating and processing data in the cloud [209].

Previous work also stated that using 'industry' data for academic health research was limited because of the absence of data donors [270]. A particularly challenging aspect of this is related to the fact that big data generators do not (or rarely) share their data with the public, or with other parties. Simply said, it is often not in the interest of a company to share their data with another source, including researchers, regardless of how impactful this research might be. This is an important issue, because insights are extremely powerful when created through the juxtaposition and combination of several sources. However, in addition to big data owners not sharing their data, these data sources are also often incompatible and or inaccessible due to ethical considerations [178]. Therefore, the challenge of data sharing remains (mostly) unsolved, even though the benefits of connecting several data traces of an individual would enable not only a more insightful analysis of patterns, but also more personal interventions. For example, in the case of asthma patients, it would be useful to connect data about their location, activities and asthma symptoms, in order to uncover useful insights about potential triggers and adherence behavior.

As greater convergence of different data sources becomes possible, a particular solution to previously mentioned data-poverty is expected in the form of data-sharing platforms, and particularly - data philanthropy [301]. This idea is reflected in partnerships between the public and industry or academic providers, that are created with the idea of sharing data for social good [301]. This kind of data sharing, depending on the form it takes, could benefit individuals who donate their data, as they could get insights about patterns in their behaviours that could improve their quality of life. In addition, their data could be used to advance research and lead to more profound insights about public health. The core advancement of this idea lies in the consented connection of several data sources that could synergically lead to more powerful insights than analysis based on a single lens.

In terms of opportunities for the current work, data sharing could provide a greater chance of capturing unbiased adherence behaviour, as well as a connection with the patients individual perceptions. This could aid examination of relationships between public and individual perceptions and their causality. In terms of public perceptions, the future of data sharing seems bright: according to Skatova (2014), 60% of study participants are willing to donate their data if the analysis is related to a public good [272]. In terms of the challenges, one of the greatest concerns is related to perceived risk of sharing particular types of data, most commonly financial data and location sharing data [271]. However, if clear rules related to privacy and codes of conducts are established, the future of data sharing could lead to more effective investigation of not only disease surveillance, but also how various perceptions affect behaviours (similarly to this work's topic).

7.5 A closer look at stigma

While the first part of the thesis (Chapter 5 and Chapter 6) examined perceptions that exist about asthma and asthma patients, the second part of this thesis (Chapter 5 and Chapter 6) was focused on quantifying the effect perceptions have on asthma treatment adherence. In order to accomplish this, Chapter 5 used a traditional statistical approach and Chapter 6 utilized a machine learning approach. Of particular interest was the effect that negative perceptions, specifically stigma about asthma, have on one's adherence to medication. Previous work has examined stigma in many areas and has usually focused heavily on psychological approaches, observing stigma as a unified concept. It has been claimed that stigma is so under-recognised because the research about stigma is usually focused on a specific outcome and in relation to one circumstance [179]. In order to address this gap, the work in Chapter 5 examined stigma-related features (denial, exogenous perceptions, discrimination, disparaging humour, internalised stigma and media) that were previously mentioned in literature and were found as themes of stigma in Chapter 4.

The mediation analysis in Chapter 5 signified that *ignoring asthma* was a maladaptive coping mechanism that patients adopt when exposed to disparaging humour, exogenous perceptions and their own denial, and this had a negative effect on adherence. On the other hand, *information seeking* had a positive impact on adherence, however, it was also positively associated with discrimination and disparaging humour. These results highlight how underlying factors of stigmatization have different mechanisms by which they impact patients. However, additionally, recognising that people potentially search for information when they are faced with different aspects of stigmatization is also an important input for the future work.

As discussed in Chapter 5, the reason behind the potentially dual nature of information seeking (which is usually observed as a positive coping mechanism) could be explained by the fact that people today usually use online sources to gather information - and this search may occur *at the same place* where other people openly express their negative views about asthma. Therefore, it is possible that asthma patients who use Twitter to find useful (or supportive) information about asthma, may come across disparaging humour about asthma, or even worse - discriminating tweets, hence being exposed to additional stigma. This is why the future interventions should carefully choose which platforms they use to communicate with asthma patients or where they suggest patients should search for information.

Linear Regression findings (from Chapter 5) implied that denial, exogenous perceptions and discrimination, were significantly correlated with adherence. However, only denial had a statistically significant beta value in the model. This result further adds to the idea that stigma is a complex issue and that different stigma features may have underlying links that cannot be observed through a simple model. However, considering the nature of the features used in the model, it is also useful to distinguish between patient and non-patient-related stigma features, since previous research has not distinguished between the object and the subject of stigmatization. This work evidenced the damaging effect of denial, however, it can be argued that denial itself represents a potential 'response' to negative perceptions. For example, if patients recognise they have no control over the image that exists about asthma in the public eye, they may exercise a sense of agency by holding control over whether they associate with the identity of an 'asthma patient'.

To further this point, stigma about asthma as a medical condition is unique, since it is concealable and not visible (as some other medical conditions are) unless patients decide to disclose they have asthma, use their medication in public or publicly experience an asthma exacerbation. This characteristic of asthma, was given more context in interviews from Chapter 3, when it was claimed that non-disclosure could serve as a 'shield', because patients can choose not to disclose they have asthma, even if they did not go far as to deny having asthma. Even more importantly, Lee et al. (2006) stated (on the example of patients with schizophrenia) that patients may see non-adherence to the treatment as their attempt to obtain a sense of relief - a particular relief that comes with the idea they are no longer a part of the group that is perceived by the society as devalued [170]. Future work may, however, be able to extend these findings, since denial, itself, can be further categorised as denial of asthma in front of other people (while acknowledging that a person does have asthma) and a denial followed by a complete separation from an asthma identity - even in the mind of patients.

7.6 The issue of masking

From the methodological point of view, it can be acknowledged that the use of traditional statistics in Chapter 5 enabled the quantification of the impact stigma-related factors have on adherence and explained the associations using mediation analysis. As it has been mentioned, traditional statistics have a long standing focus on explanation and it most commonly relies on theory (which serves as the basis of the research) [142, 318]. However, there are cases when the phenomena is too complex and cannot be explained though simple models that are based on theoretical background. In these situations, scientists that engage in quantitative work have no choice but to choose between building simple models that are theoretically elegant, but do not accurately predict actual behaviour and building more complex models that may not be congruent with the previously established theory [318].

This is related to previously mentioned issues of reproducability of the work, which may also lead to potentially wrong inputs for the method that supersedes quantitative work (and uses results of traditional statistics). It was previously highlighted that one of the main gaps in the current field of this research are the incompatible, contradictory results from previously conducted quantitative studies about the presence of stigma. To add to this, Chapter 5 and Chapter 6 state that stigma-related features have negative, statistically significant impact on adherence, while some previously conducted work states that stigma has no impact on adherence [267]. The reason behind the lack of generalisation of findings about stigma may be attributed to small sample sizes (with less than 100 participants) that can introduce biases. Additionally, there can be different ways of defining concepts (e.g. using different scales for stigma). To avoid this issue, this work conducted a study that relied on adjusted stigma scales (previously used for stigma related to other concepts).

However, there is another potential obstacle for mixed methods, when traditional statistics need to feed the results into another method. This relates to complex problems when there is a *significant level of masking* (or information sharing) between variables, and their connections with the output variable are non-linear. Even if there is a simple enough model that can be easily interpretable, if these complex relationships are not taken into account, the findings of the quantitative work that are then forwarded to the following method could input misleading results. This is often due to the presence of high multicollinearity in the data, a highly common occurrence in big data analytics, or any analysis that includes a large number of input features. This results in insights being obfuscated by extensive shared information and non-linear interactions occurring across variables.

To illustrate this, it is useful to reflect on findings from Chapter 5. Several stigma-related features have been assessed in the Multiple Linear Regression, with non-adherence as the output variable and the results indicate that *denial* is the only feature that has a statistically significant beta value ($p < 0.05$). This result is theoretically surprising, especially since other stigma-related features had a statistically significant (negative) correlation with adherence. These findings can be contrasted with the findings of Chapter 6 which, instead of using regression, treated adherence prediction as a classification task (the adherence scale was turned in a binary output variable). An additional difference was also that classification, several groups of factors were used to predict adherence, whereas in regression task only stigma-related features were used in the model. The results of these two tasks were slightly different - individual MCR results, following the classification task imply that not *only* denial, but also other stigma-related features, such as discrimination, were related to adherence.

It is, still, important to repeat that the task used in Chapter 5 was regression and the task used in Chapter 6 was classification. Even though both traditional statistics and machine learning work on both regression and classification tasks, machine learning has traditionally been more associated with classification problems [289]. On the other hand, traditional statistics focus on regression. Statisticians usually deal with smaller sample sizes and if an output variable is continuous in nature, there is a strong need to not lose information unnecessarily, which is why output variables are rarely dichotomized.[1] Additionally, as previously stated, traditional statistics, as a dominant psychological tools [183], are focused on explanation, which is easily accessible through traditional statistical outcomes such as level of

[1]A related challenge that traditional statistical methods could face is related to the arbitrary nature of 0.05 p value (the probability of data, given the null hypothesis). In 1996, Geoffrey Loftus illustrated the case when psychologists get divided between non-effects and real effects when the difference between results is 0.001 (p value of 0.051 is seen as non-effect) [183].

statistical significance of a model and relationships.

A regression algorithm can predict a discrete value which is in the form of an integer quantity A classification algorithm can predict a continuous value if it is in the form of a class label probability

However, even after taking into account the paradigm differences between regression and classification, the difference between the results generated by the two tasks is still surprising. Along with the challenges that both traditional statistics and machine learning methods face, it can be argued that the following is the reason behind this difference: Linear Regression focused on simple, linear relationships. However, as we said before, stigma mechanisms seem to be more complex. Therefore, it is possible that by taking into account both linear and non-linear relationships, RF-MCR model (based on the classification tasks) was able to highlight the underlying masking effects and more complex relationships, hence providing the result that there is more than one stigma-related feature that is useful is predicting adherence.

7.7 The importance of perceptions: revisited

The function of machine learning was to expand the qualitative work and predict prevalence of perceptions in Chapter 4, but also to complement previous work by making a full circle by re-assessing the value of perceptions in affecting adherence to medication (Chapter 6). The value of machine learning was particularly evident in Chapter 6. While various drivers for non-adherence have been considered in isolation, interactions between demographic, behavioural, perceptual and situational factors had never been modelled in concert mostly due to the limited ability of traditional methods to group such a large variety of features. In order to address this gap, this study used a non-linear modelling approach, introducing Group-MCR to unpack and quantify the importance of specific non-adherence feature sets in underpinning explanations.

The empirical value of this study lies in the finding that perceptions, as a group of features, have a crucial impact on non-adherence - and are more valuable in predicting adherence levels than any other group. As mentioned in Chapter 2, the current theories that analyse asthma treatment non-adherence, have usually focused on demographics, psychological traits or lifestyle choices (such as smoking) and the focus has rarely been on the importance of perceptions. This lack of investigation about perceptions is not limited to asthma research. In the field of schizophrenia research, it was mentioned that non-adherence is rarely viewed as a result of negative perceptions about the condition. Some theoretical models do focus on the value of perceptions, however, they usually examine asthma patients' perceptions about their condition and treatment, such as their views about the necessity of medication and their concerns about the medication and asthma itself [50, 49]. There is a very small number of publications that consider asthma patients' larger social context. Even when they exist, these considerations (in the topic of asthma treatment adherence) usually only tangentially address proximal socio-structural pressures, such as perceptions of other people in terms of their impact on patients asthma treatment adherence.

However, the findings from Chapter 6 signify the real value of patients' perceptions. The idea that one's belief drives their asthma medication adherence may be considered intuitive. However, having compared it to other groups that have traditionally been considered as dominant in their impact, it is notable that perceptions are not just important - they are essential. These findings open the door for the future work, and future interventions, which should focus more on making positive changes in patients' views, building more understanding about asthma and building a better 'image' of asthma patients. Another reason why the finding that perceptions are of crucial relevance is important is because unlike features like age, gender, habits or psychological traits perceptions can be not inherently linked to a single patient - their origin and influence may be outside of an individual's agency. In simple terms, it is currently not possible to delineate where the affect of a patients' individual perceptions end and where the influence of public perceptions begin (as such we cannot comment on the causality based on studies of this work).

We could further argue that it is possible that the mental representations that patients create in their minds about asthma (and themselves) may be a response to discrimination or disparaging humour about asthma. If this was the case, this could almost alleviate a level of responsibility for one's lack of proper asthma management, if they are (or perceive themselves to be) victims of stigmatization. Following this point of view lessons could be learnt from considering the Social Disability model. The social disability model emphasises that in the case of physical disabilities, stigma originated in the public opinion represents a significant burden in patients' everyday lives and may make patients feel oppressed and marginalised [264].

Apart from widely discussed negative perceptions about asthma, there is also an important positive finding, related to the perceived sense of community (that was detected from tweets). This perception about an alternative community is a valuable finding given that relationships that affect patients have so far been only considered to be relationships patients have with their medical professional and family. However, as our lives become more 'digitalised', it is important to recognise the value of the existing online community. Chapter 4 emphasised topics related to support, help and advice (presumed) asthma patients share. There is also a set of communication strategies that patients engage in when they seek the support from the online community, which is most commonly represented in help-seeking tweets. The perceived sense of community is just another implication for the future work, given that patients are not only a part of their offline social circle, their social media experiences can also be an invaluable source of support (even if it is a source of stigma as well).

7.8 Changing perceptions: The Future of Perceptions about Asthma

There were several implications created as a result of this work, that can be useful in the future interventions and campaigns aimed at changing negative perceptions about asthma and asthma patients.

The leitmotiv of this thesis was the idea that negative perceptions patients have may be informed by perceptions held by the public, and reinforced through social interactions of patients and non-patients, even though we cannot claim causality. Based on this idea, and as highlighted in the previous section, patients have their own, internal perceptions and external perceptions that are based on actions and reactions of others. Additionally, the work from this thesis also emphasised that some of the most prominent stigma features are largely associated with the interactions between patients and non-patients (i.e. discrimination was found to be correlated with non-adherence in Chapter 5). This means that future interventions should not only focus on patients, but also on changing perceptions of non-patients. This could even mean that future interventions could create separate campaigns to reduce stigma: one specifically created for patients and the other one created for non-patients.

In terms of the design of interventions focused on patients, this thesis provided several insights. Firstly, Chapter 4 provided evidence that social media channels represent a significant source of support for asthma patients. This means that in future, changing perceptions could focus on development of a platform that would have a form of a social media (or an app). The caveat for this was discovered in Chapter 5 where findings imply the possibility that people may be exposed to even more stigma when they visit social media. Therefore, the recommendation is that the channel where the intervention can be organised is either chosen carefully, or created separately from currently existing social media channels.

Another implication for future interventions was obtained in Chapters 5 and 6, where denial was recognised as the most important stigma-related feature. This finding means that in future, clinicians and health professionals should focus on addressing denial. This could be done by developing a lingua franca between patients and doctors, where doctors could directly ask patients if they 'accept' their diagnosis. Additionally, patients' views and denial of their conditions could be gently challenged, by providing new information and de-stigmatizing asthma - therefore, by addressing both cognitive and emotional component of denial [227].

Lastly, some inspiration for perception change could be adopted from previously conducted work related to other stigmatised medical conditions. For example, the social disability model introduced a bright example of potential change, since the 1995 Disability Discrimination Act has led to a social reform - general adoption of practical changes in terms of accessibility. This change led to changes in public perceptions, as having a disability started being seen as more normalised and less stigmatised. Importantly, as a result, this change led to a significant increase in self-esteem of people with disabilities [264], which could be another reason why focus on changing public opinion can lead to changes in individual, internal perceptions and hence - increased adherence. Simply said, it can be argued that a positive change can be made if patients that perceive and experience stigma about their medical condition, including asthma, realise that it is not only them who need to change - the society should change as well.

7.9 Rashomon effect: a parallel to mixed methods

The current research combines the fields of psychology and data science. Historically, the main focus of psychology was on the explanation of mechanisms that drive one's behaviour and as such it was most commonly associated with the use of traditional statistics [318]. This use of statistical methods in social sciences is almost exclusively reserved for testing causal theory, as it was previously stated - traditional statistics rely on the theory to generate the hypothesis. The reason why there is such a focus on explanation in statistical models may be explained by this prevalent idea: it is assumed that a high explanatory power of a model will also mean a high predictive power [266].

Conversely, data science, and in particular, machine learning techniques are focused on predictive modelling. Predictive modelling is praised for its application value, but it is not used for theory building or testing [266]. One part of the reason may be that even when the resulting prediction model has a great accuracy, it often lacks an explicit interpretation that relates to existing knowledge [142]. This may be why machine learning is more commonly associated with practical research applications - and is more widely used in the industry than traditional statistics. In short, statistics requires the use of a model that supports our knowledge of the system, while machine learning uses predictive modelling by relying on empirical capabilities [142].

These two goals and research strategies seem entirely opposite and exclusive, especially when researchers from either field need to make a trade off between theoretical accuracy and improved empirical precision. For example, neural networks used in Chapter 4 have a state of the art prediction accuracy, in classifying between perception and non-perception tweets. However, neural networks are most commonly described as 'black box' models - since it is impossible for a human to interpret which patterns were recognised as significant in distinguishing between two groups. However, since the goal of predicting perceptions and non-perceptions was to correctly classify tweets, the transparency of the rules that were used was secondary. Conversely, in Chapter 5, Linear Regression was used along with the traditional statistical modeling. The resulting variance explained using this model was relatively low, however, the model was deemed as statistically significant, and the focus was on the interpretation: resulting beta coefficients of each of the features that was used to predict adherence (out of which only one had a statistically significant role in the model).

However, after a long discussion about the differences between the machine learning and traditional statistics approach, Chapter 6 introduced a method that may be seen as a conciliation between the two paradigms. Namely, some authors have lately begun to develop methods that are capable of combining both prediction and explanation [276]. In simple terms, the transfer of knowledge is needed - *why would we explain, without being able to generalise*, and *why would we predict, if we cannot explain*. Therefore, it was recognised that these two goals do not need to be necessarily exclusive [256].

As mentioned, an example of such a methodology is Model Class Reliance (MCR) which was used in this thesis in Chapter 6. MCR can be interpreted as a bridge between two opposing sides, given that it crosses the boundaries between prediction and inference. In a way, MCR can be interpreted as a combined method, in itself, and not only because it combines prediction and inference. As a parallel to qualitative research that asserts there can be several interpretations of a single reality and each one of them is correct - MCR acknowledges that several competing (equally performing) models may exist and that each is valid and that each brings value to the interpretation [276]. This is why MCR is most commonly associated with Roshomon Sets, a concept based on a Japanese film that provided the basis for an explanatory paradigm for complexity [14].

Development of novel methods that combine both prediction and explanation, such as MCR highlights the usefulness of bridging long-existing and even outdated gaps between methodological and even philosophical approaches, by providing a more complete and multi-layered understanding of a topic. This latter characteristic especially enables the assessment of the distance between the theory and practical application - and may serve as a reality check [266]. Even more importantly, it can provide guidelines for new theory generation or at least, suggest improvements to existing theoretical knowledge. Additionally, the relevance of methods that combine both prediction and explanation is reflected in the fact that these two tasks are most frequently compatible. Interpretation of the results helps us understand the hidden aspects of the construct in question. For example, in Chapter 6 we were able to successfully predict non-adherence based on the groups of factors, however, the MCR model provided a deeper insight by revealing that it was the group of perceptions that played the main role (which in itself is a useful implication). This means that interpretation was able to further the goals of a prediction. A study conducted by Roth and Mehta (2002), which used a similar combination of methods, claims there is certainly an added value in asking for both interpretation and prediction from the same set of data, given there may be compatibility as answering one question may help answer the other [256].

The novel methodological approaches that are able to combine prediction and explanation can, in nature, be seen as a parallel to mixed methods. They both raise the notion that when examining complex phenomena, such as perceptions, prediction and explanation are not only both valuable - they are often inseparable. Specifically, in this thesis, qualitative analysis uncovered the sets of social meanings behind particular perceptions, such as the perceived lack of understanding about asthma. Quantitative work in the form of Twitter analysis complimented this work by uncovering previously unavailable social structures, thanks to its naturalistic setting and quantified the prevalence of negative perceptions. Next, quantitative work in the form of traditional statistics then used the relevant input based on patients experiences and assessed the impact relevant negative perceptions have on adherence. Finally, the machine learning model was able to deal with the complex relationships, jointly increasing the models' predictive power and using new methods to interpret the model by quantifying the value of perceptions as a group (Chapter 6).

Therefore, similarly to Rashomon effect that acknowledges the existence of several (competing) realities at the same time, mixed methods used in this thesis facilitate different avenues of exploration and utilise the synergy of using several (competing) methods, casting light on phenomena from different points of view to potentially reveal different 'truths'. The resulting findings emphasised that mixing methods should be considered an essential precondition in the exploration of asthma patients behavioural drivers, given that patients' perceptions and non-adherence are just like many other features of human nature - more complex than what it may seem if only a single-lens approach is followed. Simply said, sometimes 'the whole is greater than the sum of its parts.'

Chapter 8

Conclusion

This thesis used a synergy of methodological approaches to investigate the otherwise hardly observable perceptions regarding asthma and asthma patients; and more importantly - their impact on asthma treatment adherence. One of the main motivators for this thesis was that adherence to asthma medication is still a serious issue. To date, there is a large gap between the optimal and measured levels of adherence to medication. This thesis addressed this problem, by investigating the role socio-cognitive factors have on one's decision to regularly use their medication. This has, so far, been an under-investigated topic, with the majority of previous work being focused on patients' characteristics and practical limitations. Additionally, previously conducted research in the field of asthma perceptions relied on theories based on views and beliefs patients held about asthma, perceived asthma consequences and the necessity of asthma medication. Apart from this, another challenge introduced the fact that perceptions, as a topic, have a highly elusive nature. While previous work suffered from general data poverty, high cost to acquire an adequate number of participants and general limitations of qualitative research and traditional statistics, the current work learned from all these lessons. As a result, this work used a combination of various approaches, including novel methodological processes, in order to capture perceptions, not only through conversations and surveys, but also 'in the wild', using publicly available 'big data'.

Qualitative research has been a crucial part of this investigation, because it enabled the discovery of perceptions that exist about asthma, based on conversations with patients, but also from social media. The main power of the qualitative approach in this work is that it provides context. As a result, this was the first signal that perceptions about asthma have a more nuanced nature, since they were identified as internal - aligning with the previous work that focused mostly on patients perceptions, but also as external - perceptions that originate from the public opinion. Through a comparison between publicly available tweets and perceptions that arose in interviews, we could detect the similarities between the two sources - mostly reflecting the general lack of understanding and perceived 'weakness' that patients believed would be assigned to them once they disclose they have asthma to other people. However, there was also a significant disconnect between 'public' and 'patients' perceptions. For example, stigmatization reflected in disparaging humour, did not arise as a theme in interviews. This introduced the first significant characteristic of stigma, later defined as a group of negative perceptions about asthma - stigma was established as a topic that cannot be widely discussed in interviews. Namely, biases can be introduced in qualitative research of perceptions since patients often choose what kind of views they disclose in conversations with researchers, from the fear of being labelled as socially undesirable, but also due to demand characteristics.

A unique methodological opportunity arose with the increased accessibility of publicly available social media data, where people express their opinions and beliefs more freely and without biases. Twitter content, in particular, was recognised as a rich source of public perceptions and was used in the process of perception extraction and analysis, as an addition to traditional interviews. Additionally, since perceptions currently lack a formalized definition of perceptions in big data

- a novel approach was introduced to perception detection and analysis, in which neural networks demonstrated a supreme performance (even though it was followed by a lack of interpretability). The overall concept enabled measurement of prevalence of different groups of perceptions, where stigma was found to be one of the most dominant overarching groups, followed by a perceived sense of community that exists online. However, Twitter often lacks context - for example, it does not enable delineation between people who do and do not have asthma (unless it is explicitly stated in the tweet). Additionally, there was no information about patients' levels of adherence, therefore, it was not possible to measure the actual effect perceptions have on adherence.

Following these gaps, the second part of the thesis was more concerned with measuring the extent to which uncovered perceptions affect treatment adherence, using traditional statistics which is praised for its focus on interpretation; and machine learning techniques which demonstrated ability to deal with complex relationships. Unlocking the stigma mechanisms provided an insight that even though it is considered to be an atomic concept, stigmatization can be further dissected into several underlying mechanisms. This is particularly relevant considering that the literature mostly does not delineate between the source and the target of stigmatization (hence the role of non-patients is often disregarded). Findings from this thesis indicate that different stigma-related features are associated with different coping mechanisms that patients undertake when faced with negative perceptions. For example, while exposure to disparaging humour about asthma was found to be associated with ignoring asthma, perceived discrimination was associated with information seeking. However, traditional statistics are not without their own limitations, therefore, the final study utilized a more nuanced approach to dissecting which features affect the adherence the most.

The final study of the work did not only make a full circle in terms of the theoretical contribution, but also represents the exponential increase in the power of methodological approaches - considering that the novel method used in the final study combines both prediction and interpretation *and* takes into account the potential effects of masking that were signaled in the preceding studies. The findings illustrate that patients perceptions are not only predictive of non-adherence, but also *essential* to one's adherence behaviour. As mentioned at several points in this thesis, the finding that perceptions are, as a group of factors, the most predictive of non-adherence, and that stigma, as the leading negative set of perceptions impacts adherence, demonstrate the need for future interventions to observe patients as a part of their ever-changing social environment. Patients are inevitably intertwined with their social environment and one's personal *and* perceptions of other people have an impact on one's health decisions. This is particularly important today, as (negative) perceptions of others are commonly in the digitalised reality, which is unfiltered and readily available to all of us. This works demonstrated that stigma can have damaging impact on adherence, contributing to previous idea that stigma can cause a cascade of adverse outcomes, non-adherence and social exclusion being only some of them [170].

However, the underlying sentiment of this thesis advocates the idea that asthma patients should have no guilt nor embarrassment about having a health condition, therefore, the urgency of perception change should be focused on addressing negative public perceptions, at least as much as patients' perceptions. Additionally, this thesis provides support for future development of asthma care in the digital age, since the findings have illustrated that patients turn to their online sources in the search for alternative support. Technology is increasingly integrated in the way patients manage their asthma, but, based on the findings of this thesis, it could not only provide us with ways to adopt more effective interventions, but also with means to collect more complete and much needed data about asthma patients.

Finally this thesis took readers on a journey into the areas of psychological and machine learning approaches. While critically discussing and uncovering the weaknesses and strengths of each of the methods applied, some lessons were learned about the application and value of mixed methods. It can be concluded that for wicked topics, such as perceptions, mixed methods provide the synergy that not only overrides weaknesses of individual approaches, but also enables one to understand the big picture with clarity. A significant addition to qualitative approach and traditional statistics was recognised in machine learning approaches, particularly the use of big data analytics that provides uninterrupted and (almost) unbiased insights. Regardless of the distinct challenges and hurdles of combining different paradigms, this work is another advocate of creating the space for fast-evolving fields to aid the research of the social world with the depth and scope it requires and, ultimately - deserves.

Bibliography

[1] Anna-Mari Aalto, Kristiina Härkäpää, Arja R Aro, and Pekka Rissanen. Ways of coping with asthma in everyday life: validation of the asthma specific coping scale. *Journal of psychosomatic research*, 53(6):1061–1069, 2002.

[2] Melissa M Abo, Michael D Slater, and Parul Jain. Using health conditions for laughs and health policy support: The case of food allergies. *Health communication*, 32(7):803–811, 2017.

[3] Dominic Abrams. Social identity on a national scale: Optimal distinctiveness and young people's self-expression through musical preference. *Group Processes & Intergroup Relations*, 12(3):303–317, 2009.

[4] Robert J Adams, David Wilson, Brian J Smith, and Richard E Ruffin. Impact of coping and socioeconomic factors on quality of life in adults with asthma. *Respirology*, 9(1):87–95, 2004.

[5] Stephanie Adams, Roisin Pill, and Alan Jones. Medication, chronic illness and identity: the perspective of people with asthma. *Social science & medicine*, 45(2):189–201, 1997.

[6] SOHAIL Ahmad and NAHLAH ELKUDSSIAH Ismail. A qualitative study exploring the impact of stigma in the lives of adult asthma patients in selangor malaysia. *Int J Pharm Pharm Sci*, 7:373–5, 2015.

[7] Sohail Ahmad and NAHLAH ELKUDSSIAH Ismail. Stigma in the lives

of asthma patients: a review from the literature. *International Journal of Pharmacy and Pharmaceutical Sciences*, 7(7):40–46, 2015.

[8] Harith Kh Al-Qazaz, Mohamed A Hassali, Asrul A Shafie, Syed A Sulaiman, Shameni Sundram, and Donald E Morisky. The eight-item morisky medication adherence scale mmas: translation and validation of the malaysian version. *Diabetes research and clinical practice*, 90(2):216–221, 2010.

[9] Angel B Algarin, Diana M Sheehan, Nelson Varas-Diaz, Kristopher P Fennie, Zhi Zhou, Emma C Spencer, Robert L Cook, Jamie P Morano, and Gladys E Ibanez. Health care-specific enacted hiv-related stigma's association with antiretroviral therapy adherence and viral suppression among people living with hiv in florida. *AIDS Patient Care and STDs*, 34(7):316–326, 2020.

[10] L Alison Phillips, Howard Leventhal, and Elaine A Leventhal. Assessing theoretical predictors of long-term medication adherence: Patients' treatment-related beliefs, experiential feedback and habit development. *Psychology & Health*, 28(10):1135–1151, 2013.

[11] Paul Allison. When can you safely ignore multicollinearity. *Statistical horizons*, 5(1):1–2, 2012.

[12] Peter Allmark and Katarzyna Machaczek. Realism and pragmatism in a mixed methods study. *Journal of advanced nursing*, 74(6):1301–1309, 2018.

[13] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.

[14] Robert Anderson. The rashomon effect and communication. *Canadian Journal of Communication*, 41(2), 2016.

[15] Kelly L Andrews, Sandra C Jones, and Judy Mullan. Stigma: Still an impor-

tant issue for adults with asthma. *Journal of Asthma & Allergy Educators*, 4(4):165–171, 2013.

[16] Matthias C Angermeyer, Herbert Matschinger, Bruce G Link, and Georg Schomerus. Public attitudes regarding individual and structural discrimination: Two sides of the same coin? *Social Science & Medicine*, 103:60–66, 2014.

[17] Andrea J Apter, Ray C Boston, Maureen George, A Lorraine Norfleet, Thomas Tenhave, James C Coyne, Kathleen Birck, Susan T Reisine, Andrew J Cucchiara, and Harold I Feldman. Modifiable barriers to adherence to inhaled steroids among adults with asthma: it's not just black and white. *Journal of Allergy and Clinical Immunology*, 111(6):1219–1226, 2003.

[18] Malin Axelsson, Eva Brink, and Jan Lötvall. A personality and gender perspective on adherence and health-related quality of life in people with asthma and/or allergic rhinitis. *Journal of the American Association of Nurse Practitioners*, 26(1):32–39, 2014.

[19] Malin Axelsson, Eva Brink, Jesper Lundgren, and Jan Lötvall. The influence of personality traits on reported adherence to medication in individuals with chronic disease: an epidemiological study in west sweden. *PloS one*, 6(3):e18241, 2011.

[20] Malin Axelsson, Maria Emilsson, Eva Brink, J Lundgren, K Torén, and JJRM Lötvall. Personality, adherence, asthma control and health-related quality of life in young adult asthmatics. *Respiratory medicine*, 103(7):1033–1040, 2009.

[21] Malin Axelsson, Jan Lötvall, Jesper Lundgren, and Eva Brink. Motivational foci and asthma medication tactics directed towards a functional day. *BMC public health*, 11(1):1–9, 2011.

[22] Alan P Baptist, Michael Thompson, Karla Stoermer Grossman, Layla Mohammed, Annie Sy, and Georgiana M Sanders. Social media, text messaging, and email—preferences of asthma patients between 12 and 40 years old. *Journal of Asthma*, 48(8):824–830, 2011.

[23] Meredith A Barrett, Olivier Humblet, Robert A Hiatt, and Nancy E Adler. Big data and disease prevention: from quantified self to quantified communities. *Big data*, 1(3):168–175, 2013.

[24] Christopher Barton, David Clarke, Nabil Sulaiman, and Michael Abramson. Coping as a mediator of psychosocial impediments to optimal management and control of asthma. *Respiratory medicine*, 97(7):747–761, 2003.

[25] Roy F Baumeister and Mark Muraven. Identity as adaptation to social, cultural, and historical context. *Journal of adolescence*, 19(5):405–416, 1996.

[26] Marshall H Becker, Susan M Radius, Irwin M Rosenstock, Robert H Drachman, Kenneth C Schuberth, and Katherine C Teets. Compliance with a medical regimen for asthma: a test of the health belief model. *Public health reports*, 93(3):268, 1978.

[27] Saul Becker, Alan Bryman, and Harry Ferguson. *Understanding research for social policy and social work 2E: themes, methods and approaches*. policy press, 2012.

[28] Emma Bell, Alan Bryman, and Bill Harley. *Business research methods*. Oxford university press, 2018.

[29] Bruce Bender, Henry Milgrom, and Andrea Apter. Adherence intervention research: what have we learned and what do we do next? *Journal of Allergy and Clinical Immunology*, 112(3):489–494, 2003.

[30] Jérôme Béranger. *Big data and ethics: the medical datasphere*. Elsevier, 2016.

[31] Barbara E Berger, Mary C Kapella, and Janet L Larson. The experience of stigma in chronic obstructive pulmonary disease. *Western journal of nursing research*, 33(7):916–932, 2011.

[32] Kenneth C Bessant and Eric D MacPherson. Thoughts on the origins, concepts, and pedagogy of statistics as a "separate discipline". *The American Statistician*, 56(1):22–28, 2002.

[33] Jiang Bian, Kenji Yoshigoe, Amanda Hicks, Jiawei Yuan, Zhe He, Mengjun Xie, Yi Guo, Mattia Prosperi, Ramzi Salloum, and François Modave. Mining twitter to assess the public perception of the "internet of things". *PloS one*, 11(7):e0158450, 2016.

[34] Natalie Bidad, Neil Barnes, Chris Griffiths, and Rob Horne. Understanding patients' perceptions of asthma control: a qualitative study. *European Respiratory Journal*, 51(6), 2018.

[35] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[36] Prakhar Biyani, Cornelia Caragea, Amit Singh, and Prasenjit Mitra. I want what i need! analyzing subjectivity of online forum threads. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 2495–2498, 2012.

[37] Susan W Blaakman, Alyssa Cohen, Maria Fagnano, and Jill S Halterman. Asthma medication adherence among urban teens: a qualitative analysis of barriers, facilitators and experiences with school-based care. *Journal of Asthma*, 51(5):522–529, 2014.

[38] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

[39] Catherine Bolman, Titia G Arwert, and Trijntje Völlink. Adherence to prophylactic asthma medication: Habit strength and cognitions. *Heart & Lung*, 40(1):63–75, 2011.

[40] Louis-Philippe Boulet. Perception of the role and potential side effects of inhaled corticosteroids among asthmatic patients. *Chest*, 113(3):587–592, 1998.

[41] Louis-Philippe Boulet, Daniel Vervloet, Yves Magar, and Juliet M Foster. Adherence: the goal to control asthma. *Clinics in chest medicine*, 33(3):405–417, 2012.

[42] Constantinos Boulis and Mari Ostendorf. Text classification by augmenting the bag-of-words representation with redundancy-compensated bigrams. In *Proc. of the International Workshop in Feature Selection in Data Mining*, pages 9–16. Citeseer, 2005.

[43] Fulvio Braido. Failure in asthma control: reasons and consequences. *Scientifica*, 2013, 2013.

[44] Elizabeth Broadbent, Keith J Petrie, Jodie Main, and John Weinman. The brief illness perception questionnaire. *Journal of psychosomatic research*, 60(6):631–637, 2006.

[45] C Michael Brooks, James M Richards, Connie L Kohler, Seng-Jaw Soong, Beverly Martin, Richard A Windsor, and William C Bailey. Assessing adherence to asthma medication and inhaler regimens: a psychometric analysis of adult self-report scales. *Medical care*, pages 298–307, 1994.

[46] Emery N Brown and Robert E Kass. What is statistics? *The American Statistician*, 63(2):105–110, 2009.

[47] Marie T Brown and Jennifer K Bussell. Medication adherence: Who cares? In *Mayo clinic proceedings*, volume 86, pages 304–314. Elsevier, 2011.

[48] Alan Bryman and Emma Bell. Business research methods‖, chapter 11, 2008.

[49] Patricia Vernal Burkhart and Mary Kay Rayens. Self-concept and health locus of control: factors related to children's adherence to recommended asthma regimen. *Pediatric nursing*, 31(5), 2005.

[50] B Byer and Lynn B Myers. Psychological correlates of adherence to medication in asthma. *Psychology, Health & Medicine*, 5(4):389–393, 2000.

[51] Erik Cambria, Andrew Livingstone, and Amir Hussain. The hourglass of emotions. In *Cognitive behavioural systems*, pages 144–157. Springer, 2012.

[52] DA Campbell, Peter Mackinlay Yellowlees, G McLennan, JR Coates, PA Frith, PA Gluyas, KM Latimer, CG Luke, AJ Martin, and RE Ruffin. Psychiatric and medical features of near fatal asthma. *Thorax*, 50(3):254–259, 1995.

[53] Marisa Casale, Mark Boyes, Marija Pantelic, Elona Toska, and Lucie Cluver. Suicidal thoughts and behaviour among south african adolescents living with hiv: Can social support buffer the impact of stigma? *J. of Affective Disorders*, 245:82–90, 2019.

[54] Davide Castelvecchi. Can we open the black box of ai? *Nature News*, 538(7623):20, 2016.

[55] Marianne Celano, Robert J Geller, Keith M Phillips, and Robin Ziman. Treatment adherence among low-income children with asthma. *Journal of Pediatric Psychology*, 23(6):345–349, 1998.

[56] Victoria L Champion, Celette Sugg Skinner, et al. The health belief model. *Health behavior and health education: Theory, research, and practice*, 4:45–65, 2008.

[57] José Miguel Chatkin, Daniela Cavalet-Blanco, Nóris Coimbra Scaglia, Roberto Guidotti Tonietto, Mário B Wagner, and Carlos Cezar Fritscher. Compliance with maintenance treatment of asthma (adere study). *Jornal brasileiro de pneumologia*, 32:277–283, 2006.

[58] Iti Chaturvedi, Edoardo Ragusa, Paolo Gastaldo, Rodolfo Zunino, and Erik Cambria. Bayesian network based extreme learning machine for subjectivity detection. *Journal of The Franklin Institute*, 355(4):1780–1797, 2018.

[59] Cleo H Cherryholmes. Notes on pragmatism and scientific realism. *Educational researcher*, 21(6):13–17, 1992.

[60] Cynthia Chew and Gunther Eysenbach. Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak. *PloS one*, 5(11):e14118, 2010.

[61] Janne Chung and Gary S Monroe. Exploring social desirability bias. *Journal of Business Ethics*, 44(4):291–302, 2003.

[62] Cindy Dell Clark. Asthma episodes: stigma, children, and hollywood films. *Medical anthropology quarterly*, 26(1):92–115, 2012.

[63] Eric Clark. Towards a copenhagen interpretation of gentrification. *Urban studies*, 31(7):1033–1042, 1994.

[64] Sarah Clifford, Nick Barber, and Rob Horne. Understanding different beliefs held by adherers, unintentional nonadherers, and intentional nonadherers: application of the necessity–concerns framework. *Journal of psychosomatic research*, 64(1):41–46, 2008.

[65] Anne Cocos, Alexander G Fiks, and Aaron J Masino. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in twitter posts. *Journal of the American Medical Informatics Association*, 24(4):813–821, 2017.

[66] Robyn Cohen, Karen Franco, Ferrell Motlow, Marina Reznik, and Philip O Ozuah. Perceptions and attitudes of adolescents with asthma. *Journal of Asthma*, 40(2):207–211, 2003.

[67] Nigel Collier and Son Doan. Syndromic classification of twitter messages. In *International Conference on Electronic Healthcare*, pages 186–195. Springer, 2011.

[68] Kelly M Conn, Jill S Halterman, Susan G Fisher, H Lorrie Yoos, Nancy P Chin, and Peter G Szilagyi. Parental beliefs about medications and medication adherence among urban children with asthma. *Ambulatory Pediatrics*, 5(5):306–310, 2005.

[69] Kelly M Conn, Jill S Halterman, Kathleen Lynch, and Michael D Cabana. The impact of parents' medication beliefs on asthma management. *Pediatrics*, 120(3):e521–e526, 2007.

[70] Lucy Cooke, Lynn B Myers, and Naz Derakshan. Lung function, adherence and denial in asthma patients who exhibit a repressive coping style. *Psychology, health & medicine*, 8(1):35–44, 2003.

[71] Patrick W Corrigan, Benjamin G Druss, and Deborah A Perlick. The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*, 15(2):37–70, 2014.

[72] Patrick W Corrigan, Kristin A Kosyluk, J Konadu Fokuo, and Jin Hee Park. How does direct to consumer advertising affect the stigma of mental illness? *Community Mental Health Journal*, 50(7):792–799, 2014.

[73] Angelo G Corsico, Lucia Cazzoletti, Roberto de Marco, Christer Janson, Deborah Jarvis, Maria C Zoia, Massimiliano Bugiani, Simone Accordini, Simona Villani, Alessandra Marinoni, et al. Factors affecting adherence

to asthma treatment in an international cohort of young and middle-aged adults. *Respiratory medicine*, 101(6):1363–1367, 2007.

[74] Paul T Costa and Robert R McCrae. *The NEO personality inventory*. Psychological Assessment Resources Odessa, FL, 1985.

[75] Neil S Coulson. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *CyberPsychology & Behavior*, 8(6):580–584, 2005.

[76] Jeremy W Crampton, Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W Wilson, and Matthew Zook. Beyond the geotag: situating 'big data'and leveraging the potential of the geoweb. *Cartography and geographic information science*, 40(2):130–139, 2013.

[77] Hugh Alistair Cross, Miriam Heijnders, Ajit Dalal, Silatham Sermrittirong, and Stephanie Mak. Interventions for stigma reduction–part 1: theoretical considerations. *Disability, CBR & Inclusive Development*, 22(3):62–70, 2011.

[78] Aron Culotta and Jennifer Cutler. Mining brand perceptions from twitter social networks. *Marketing science*, 35(3):343–362, 2016.

[79] G D'Amato, L Cecchi, G Liccardi, F Pellegrino, M D'Amato, and M Sofia. Social networks: A new source of psychological stress or a way to enhance self-esteem? negative and positive implications in bronchial asthma. *International Journal on Immunorehabilitation*, 15(1):8–11, 2013.

[80] Emma Davidson, Rosalind Edwards, Lynn Jamieson, and Susie Weller. Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality & Quantity*, 53(1):363–376, 2019.

[81] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz.

Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.

[82] Anna De Simoni, Robert Horne, Louise Fleming, Andrew Bush, and Chris Griffiths. What do adolescents with asthma really think about adherence to inhalers? insights from a qualitative analysis of a uk online forum. *BMJ open*, 7(6):e015245, 2017.

[83] TA Deenen and EC Klip. Coping with asthma. *Respiratory medicine*, 87:67–70, 1993.

[84] Martyn Denscombe. Communities of practice: A research paradigm for the mixed methods approach. *Journal of mixed methods research*, 2(3):270–283, 2008.

[85] Laurence Devillers, Laurence Vidrascu, and Lori Lamel. Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4):407–422, 2005.

[86] Amitabha Dey, Rafsan Zani Rafi, Shahriar Hasan Parash, Sauvik Kundu Arko, and Amitabha Chakrabarty. Fake news pattern recognition using linguistic analysis. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 305–309. IEEE, 2018.

[87] Vasant Dhar. Data science and prediction. *Communications of the ACM*, 56(12):64–73, 2013.

[88] Ralph J DiClemente, Jennifer L Brown, and Teaniese Latham Davis. Determinants of health-related behaviors in adolescence. In *Handbook of Adolescent Health Psychology*, pages 107–127. Springer, 2013.

[89] M Robin DiMatteo. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Medical care*, pages 200–209, 2004.

[90] Pedro Domingos. The role of occam's razor in knowledge discovery. *Data mining and knowledge discovery*, 3(4):409–425, 1999.

[91] Sarjoun Doumit and Ali Minai. Online news media bias analysis using an lda-nlp approach. In *International Conference on Complex Systems*, 2011.

[92] Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5-6):352–359, 2002.

[93] Valerie A Earnshaw and Diane M Quinn. The impact of stigma in healthcare on people living with chronic illnesses. *Journal of health psychology*, 17(2):157–168, 2012.

[94] Geoff Easton. Critical realism in case study research. *Industrial marketing management*, 39(1):118–128, 2010.

[95] Kate Edgecombe, Sue Latter, Sheila Peters, and Graham Roberts. Health experiences of adolescents with uncontrolled severe asthma. *Archives of disease in childhood*, 95(12):985–991, 2010.

[96] Paul Ekman. Are there basic emotions? *Psychological Review*, 1992.

[97] Össur Ingi Emilsson, Christer Janson, Bryndís Benediktsdóttir, Sigurdur Júlíusson, and Thórarinn Gíslason. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respiratory medicine*, 106(3):459–466, 2012.

[98] Neil Eriksen. Popular culture and revolutionary theory: Understanding punk rock. *Theoretical Review*, 18:13–35, 1980.

[99] Gunther Eysenbach and James E Till. Ethical issues in qualitative research on internet communities. *Bmj*, 323(7321):1103–1105, 2001.

[100] Golnoosh Farnadi, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie-Francine Moens, and Martine De Cock. Computational personality recognition in social media. *User modeling and user-adapted interaction*, 26(2):109–142, 2016.

[101] Gilles Fauconnier. Pragmatic scales and logical structure. *Linguistic inquiry*, 6(3):353–375, 1975.

[102] David Feldman, Phyllis A Gordon, Michael J White, and Christopher Weber. The effects of people-first language and demographic variables on beliefs, attitudes and behavioral intentions toward people with disabilities. *Journal of Applied Rehabilitation Counseling*, 33(3):18–25, 2002.

[103] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[104] Alfio Ferrara, Stefano Montanelli, and Georgios Petasis. Unsupervised detection of argumentative units though topic modeling techniques. In *Proceedings of the 4th Workshop on Argument Mining*, pages 97–107, 2017.

[105] Thiemo Fetzer and Thomas Graeber. Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. *Proceedings of the National Academy of Sciences*, 118(33), 2021.

[106] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. of Machine Learning Research*, 20(177):1–81, 2019.

[107] Frances M Ford, M Hunter, MJ Hensley, A Gillies, S Carney, AJ Smith,

J Bamford, M Lenzer, G Lister, S Ravazdy, et al. Hypertension and asthma: psychological aspects. *Social Science & Medicine*, 29(1):79–84, 1989.

[108] Thomas E Ford and Mark A Ferguson. Social consequences of disparagement humor: A prejudiced norm theory. *Personality and social psychology review*, 8(1):79–94, 2004.

[109] Jerome H Friedman. Data mining and statistics: What's the connection? *Computing science and statistics*, 29(1):3–9, 1998.

[110] Joseph N Fry and John D Claxton. Semantic differential and nonmetric multidimensional scaling descriptions of brand images. *Journal of Marketing Research*, 8(2):238–240, 1971.

[111] Katsuya Futagami, Yusuke Fukazawa, Nakul Kapoor, and Tomomi Kito. Pairwise acquisition prediction with shap value interpretation. *The Journal of Finance and Data Science*, 2021.

[112] Mariana Gaytan Camarillo, Eamonn Ferguson, Vanja Ljevar, and Alexa Spence. Big changes start with small talk: Twitter and climate change in times of coronavirus pandemic. *Frontiers in Psychology*, 12:2308, 2021.

[113] Manoochehr Ghiassi, James Skinner, and David Zimbra. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16):6266–6282, 2013.

[114] Amirata Ghorbani, Abubakar Abid, and James Zou. Interpretation of neural networks is fragile. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3681–3688, 2019.

[115] Maria Giatsoglou, Despoina Chatzakou, Neil Shah, Christos Faloutsos, and Athena Vakali. Retweeting activity on twitter: Signs of deception. In *Pacific-*

Asia Conference on Knowledge Discovery and Data Mining, pages 122–134. Springer, 2015.

[116] Maria Giatsoglou, Manolis G Vozalis, Konstantinos Diamantaras, Athena Vakali, George Sarigiannidis, and Konstantinos Ch Chatzisavvas. Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, 69:214–224, 2017.

[117] Karen Glanz and Donald B Bishop. The role of behavioral science theory in development and implementation of public health interventions. *Annual review of public health*, 31:399–418, 2010.

[118] Erving Goffman. The moral career of the mental patient. *Psychiatry*, 22(2):123–142, 1959.

[119] Erving Goffman et al. *The presentation of self in everyday life*, volume 21. Harmondsworth London, 1978.

[120] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 149–156. IEEE, 2011.

[121] John A Grace. Method and apparatus for automatic table selection and generation of structured query language instructions, May 21 1996. US Patent 5,519,859.

[122] Bradi Granger, Inger Ekman, Christopher Granger, Jan Ostergren, Bertil Olofsson, Eric Michelson, John McMurray, Salim Yusuf, Marc Pfeffer, and Karl Swedberg. Adherence to medication according to sex and age in the charm programme. *Euro. J. of Heart Failure*, 2009.

[123] Wendy N Gray, Mallory Netz, Andrew McConville, David Fedele, Scott T Wagoner, and Megan R Schaefer. Medication adherence in pediatric asthma:

a systematic review of the literature. *Pediatric Pulmonology*, 53(5):668–684, 2018.

[124] Paul E Green and Vithala R Rao. Rating scales and information recovery—how many scales and response categories to use? *Journal of Marketing*, 34(3):33–39, 1970.

[125] Baptiste Gregorutti, Bertrand Michel, and Philippe Saint-Pierre. Grouped variable importance with random forests and application to multiple functional data analysis. *Computational Statistics & Data Analysis*, 90:15–35, 2015.

[126] Kathrin Grosse, Carlos Iván Chesñevar, and Ana Gabriela Maguitman. An argument-based approach to mining opinions from twitter. *AT*, 918:408–422, 2012.

[127] Naomi J Gryfe Saperia. The relative influences of knowledge, beliefs and preferences on adherence to asthma medication. 2012.

[128] Ethan A Halm, Pablo Mora, and Howard Leventhal. No symptoms, no asthma: the acute episodic disease belief is associated with poor self-management among inner-city adults with persistent asthma. *Chest*, 129(3):573–580, 2006.

[129] David J Hand. Data mining: statistics and more? *The American Statistician*, 52(2):112–118, 1998.

[130] David J Hand. Statistics and data mining: intersecting disciplines. *Acm Sigkdd Explorations Newsletter*, 1(1):16–19, 1999.

[131] Guy S Handelman, Hong Kuan Kok, Ronil V Chandra, Amir H Razavi, Shiwei Huang, Mark Brooks, Michael J Lee, and Hamed Asadi. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *American Journal of Roentgenology*, 212(1):38–43, 2019.

[132] AF Hayes. Model templates for process for spss and sas. hayes, a. and the guilford press, 2013.

[133] Aldo Hernandez-Suarez, Gabriel Sanchez-Perez, Karina Toscano-Medina, Hector Perez-Meana, Jose Portillo-Portillo, Victor Sanchez, and Luis Javier García Villalba. Using twitter data to monitor natural disaster social dynamics: a recurrent neural network approach with word embeddings and kernel density estimation. *Sensors*, 19(7):1746, 2019.

[134] Sameer Hinduja and Justin W Patchin. Bullying, cyberbullying, and suicide. *Archives of suicide research*, 14(3):206–221, 2010.

[135] Rob Horne. Compliance, adherence, and concordance: implications for asthma treatment. *Chest*, 130(1):65S–72S, 2006.

[136] Rob Horne. Improving adherence with asthma therapies. Future Medicine, 2012.

[137] Robert Horne and John Weinman. Self-regulation and self-management in asthma: exploring the role of illness perceptions and treatment beliefs in explaining non-adherence to preventer medication. *Psychology and Health*, 17(1):17–32, 2002.

[138] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

[139] Bingquan Huang, Mohand Tahar Kechadi, and Brian Buckley. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1):1414–1425, 2012.

[140] Shelby D Hunt. Positivism and paradigm dominance in consumer research: toward critical pluralism and rapprochement. *Journal of consumer research*, 18(1):32–44, 1991.

[141] Nasir Hussain and Tom Ritchey. Wicked problems. *Eur J Ind Pharm*, 31:4–7, 2011.

[142] H IJ. Statistics versus machine learning. *Nature methods*, 15(4):233, 2018.

[143] Maree Inder, Marie Crowe, Stephanie Moor, Janet Carter, Sue Luty, and Peter Joyce. 'it wouldn't be me if i didn't have bipolar disorder': managing the shift in self-identity with bipolar disorder. *Journal of Nursing and Healthcare of Chronic Illness*, 3(4):427–435, 2011.

[144] Jasmina Ivanova, Howard G Birnbaum, Matthew Hsieh, Andrew Yu, Brian Seal, Thys van der Molen, Srinivas Emani, Richard Rosiello, and Gene Colice. Adherence to inhaled corticosteroid use and local adverse events in persistent asthma. *American J. of Managed Care*, 14(12), 2008.

[145] Lia Jahedi, Sue R Downie, Bandana Saini, Hak-Kim Chan, and Sinthia Bosnic-Anticevich. Inhaler technique in asthma: how does it relate to patients' preferences and attitudes toward their inhalers? *Journal of aerosol medicine and pulmonary drug delivery*, 30(1):42–52, 2017.

[146] Adriana Jakovcevic, Linda Steg, Nadia Mazzeo, Romina Caballero, Paul Franco, Natalia Putrino, and Jesica Favara. Charges for plastic bags: Motivational and behavioral effects. *Journal of Environmental Psychology*, 40:372–380, 2014.

[147] Leslie M Janes and James M Olson. Jeer pressure: The behavioral effects of observing ridicule of others. *Personality and Social Psychology Bulletin*, 26(4):474–485, 2000.

[148] Ana Janežič, Igor Locatelli, and Mitja Kos. Criterion validity of 8-item morisky medication adherence scale in patients with asthma. *PloS one*, 12(11):e0187835, 2017.

[149] Bernard J Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188, 2009.

[150] Donna C Jessop and Derek R Rutter. Adherence to asthma medication: the role of illness representations. *Psychology and Health*, 18(5):595–612, 2003.

[151] Bridget Johnson, Jessica Henderson, Peggy Pedersen, and Linda Stonecipher. Framing asthma: A content analysis in us newspapers. *Journal of Asthma & Allergy Educators*, 2(3):135–142, 2011.

[152] Grace J Johnson and Paul J Ambrose. Neo-tribes: The power and potential of online communities in health care. *Communications of the ACM*, 49(1):107–113, 2006.

[153] R Burke Johnson and Anthony J Onwuegbuzie. Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7):14–26, 2004.

[154] Marina Jonsson, Marja Schuster, Jennifer LP Protudjer, Anna Bergström, Ann-Charlotte Egmar, and Inger Kull. Experiences of daily life among adolescents with asthma–a struggle with ambivalence. *Journal of pediatric nursing*, 35:23–29, 2017.

[155] Ulrik S Kesmodel. Cross-sectional studies–what are they good for? *Acta obstetricia et gynecologica Scandinavica*, 97(4):388–393, 2018.

[156] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. Big data: survey, technologies, opportunities, and challenges. *The scientific world journal*, 2014, 2014.

[157] Vishal Kharde, Prof Sonawane, et al. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*, 2016.

[158] Dong Jin Kim, Woo Gon Kim, and Jin Soo Han. A perceptual mapping of online travel agencies and preference attributes. *Tourism management*, 28(2):591–603, 2007.

[159] Youllee Kim, James Price Dillard, and Rachel A Smith. Communicating antibiotic stewardship: emotional responses and their impact on adherence. *Health communication*, 2019.

[160] Michael King, Sokratis Dinos, Jenifer Shaw, Robert Watson, Scott Stevens, Filippo Passetti, Scott Weich, and Marc Serfaty. The stigma scale: development of a standardised measure of the stigma of mental illness. *The British Journal of Psychiatry*, 190(3):248–254, 2007.

[161] Rob Kitchin. Big data, new epistemologies and paradigm shifts. *Big data & society*, 1(1):2053951714528481, 2014.

[162] Praveen K Kopalle and Donald R Lehmann. Alpha inflation? the impact of eliminating scale items on cronbach's alpha. *Org. Behavior & Human Decision Proc.*, 70(3), 1997.

[163] Mandy Korpusik, Shigeyuki Sakaki, Francine Chen, and Yan-Ying Chen. Recurrent neural networks for customer purchase prediction on twitter. *CBRecSys@ RecSys*, 1673:47–50, 2016.

[164] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15):5802–5805, 2013.

[165] Ellen S Koster, Daphne Philbert, Tjalling W de Vries, Liset van Dijk, and Marcel L Bouvy. "i just forget to take it": asthma self-management needs and preferences in adolescents. *Journal of Asthma*, 52(8):831–837, 2015.

[166] Fajri Koto and Mirna Adriani. A comparative study on twitter sentiment analysis: Which features are good? In *International Conference on Applica-*

tions of Natural Language to Information Systems, pages 453–457. Springer, 2015.

[167] Rosa Lavelle-Hill. *Big data psychology.* PhD thesis, University of Nottingham, 2020.

[168] Jayeon Lee and Young-shin Lim. Gendered campaign tweets: the cases of hillary clinton and donald trump. *Public Relations Review*, 42(5):849–855, 2016.

[169] Ji Young Lee, Michael D Slater, and John Tchernev. Self-deprecating humor versus other-deprecating humor in health messages. *Journal of health communication*, 20(10):1185–1195, 2015.

[170] Sing Lee, Marcus YL Chiu, Adley Tsang, Helena Chui, and Arthur Kleinman. Stigmatizing experience and structural discrimination associated with the treatment of schizophrenia in hong kong. *Social Science & Medicine*, 62(7):1685–1696, 2006.

[171] Christine M Lehane, Tine Nielsen, Walter Wittich, Shelby Langer, and Jesper Dammeyer. Couples coping with sensory loss: A dyadic study of the roles of self-and perceived partner acceptance. *British journal of health psychology*, 23(3):646–664, 2018.

[172] Helen-Maria Lekas, Karolynn Siegel, and Jason Leider. Felt and enacted stigma among hiv/hcv-coinfected adults: the impact of stigma layering. *Qualitative Health Research*, 21(9):1205–1219, 2011.

[173] Howard Leventhal, Michael Diefenbach, and Elaine A Leventhal. Illness cognition: Using common sense to understand treatment adherence and affect cognition interactions. *Cognitive therapy and research*, 16(2):143–163, 1992.

[174] Howard Leventhal, David R Nerenz, and David J Steele. *Illness representations and coping with health threats*. Routledge, 2020.

[175] Baoli Li, Yandong Liu, Ashwin Ram, Ernest V Garcia, and Eugene Agichtein. Exploring question subjectivity prediction in community qa. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 735–736, 2008.

[176] Maria H Lindberg, Lena Wettergren, Maria Wiklander, Veronica Svedhem-Johansson, and Lars E Eriksson. Psychometric evaluation of the hiv stigma scale in a swedish context. *PLoS One*, 9(12):e114867, 2014.

[177] John T Lindsay and Liam G Heaney. Nonadherence in difficult asthma–facts, myths, and a time to act. *Patient preference and adherence*, 7:329, 2013.

[178] Martin Lindstrom. *Small data: the tiny clues that uncover huge trends*. St. Martin's Press, 2016.

[179] Bruce G Link and Jo C Phelan. Stigma and its public health implications. *The Lancet*, 367(9509):528–529, 2006.

[180] Shiyu Liu, Ming Lun Ong, Kar Kin Mun, Jia Yao, and Mehul Motani. Early prediction of sepsis via smote upsampling and mutual information based downsampling. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.

[181] Vanja Ljevar, James Goulding, Gavin Smith, and Alexa Spence. Using model class reliance to measure group effects on non-adherence to asthma medication. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1699–1708. IEEE, 2021.

[182] Vanja Ljevar, James Goulding, Alexa Spence, and Gavin Smith. Perception

detection using twitter. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 4250–4256. IEEE, 2020.

[183] Geoffrey R Loftus. Psychology will be a much better science when we change the way we analyze data. *Current directions in psychological science*, 5(6):161–171, 1996.

[184] Christine Loignon, Christophe Bedos, Robert Sévigny, and Nicole Leduc. Understanding the self-care strategies of patients with asthma. *Patient education and counseling*, 75(2):256–262, 2009.

[185] Yingjie Lu, Pengzhu Zhang, Jingfang Liu, Jia Li, and Shasha Deng. Health-related hot topic detection in online communities using text clustering. *Plos one*, 8(2):e56221, 2013.

[186] Harry Luna-Aveiga, José Medina-Moreira, Katty Lagos-Ortiz, Oscar Apolinario, Mario Andrés Paredes-Valverde, María del Pilar Salas-Zárate, and Rafael Valencia-García. Sentiment polarity detection in social networks: an approach for asthma disease management. In *International conference on computer science, applied mathematics and applications*, pages 141–152. Springer, 2017.

[187] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 2017.

[188] Helen Lycett, Emilie Wildman, Eva M Raebel, Jon-Paul Sherlock, Tom Kenny, and Amy Hai Yan Chan. Treatment perceptions in patients with asthma: synthesis of factors influencing adherence. *Respiratory medicine*, 141:180–189, 2018.

[189] Noella Mackenzie and Sally Knipe. Research dilemmas: Paradigms, methods and methodology. *Issues in educational research*, 16(2):193–205, 2006.

[190] Jean Macq, Alejandro Solis, and Guillermo Martinez. Assessing the stigma of tuberculosis. *Psychology, health & medicine*, 11(3):346–352, 2006.

[191] Pruthikrai Mahatanankoon, H Joseph Wen, and Billy Lim. Consumer-based m-commerce: exploring consumer perception of mobile applications. *Computer standards & interfaces*, 27(4):347–357, 2005.

[192] Satu MÄkinen, Tarja Suominen, and Sirkka Lauri. Self-care in adults with asthma: how they cope. *Journal of Clinical Nursing*, 9(4):557–565, 2000.

[193] Jane Mallett. Use of humour and laughter in patient care. *British Journal of Nursing*, 2(3):172–175, 1993.

[194] Aaron Manson. Language concordance as a determinant of patient compliance and emergency room use in patients with asthma. *Medical care*, pages 1119–1128, 1988.

[195] Riham Mansour, Mohamed Farouk Abdel Hady, Eman Hosam, Hani Amr, and Ahmed Ashour. Feature selection for twitter sentiment analysis: An experimental study. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 92–103. Springer, 2015.

[196] Annette Markham and Elizabeth Buchanan. Ethical decision-making and internet research: Version 2.0. recommendations from the aoir ethics working committee. *Available online: aoir. org/reports/ethics2. pdf*, 2012.

[197] Fernando Martin-Sanchez and Karin Verspoor. Big data in medicine is driving big changes. *Yearbook of medical informatics*, 23(01):14–20, 2014.

[198] Theresa Vera Masuch, Myriam Bea, Barbara Alm, Peter Deibler, and Esther Sobanski. Internalized stigma, anticipated discrimination and perceived public stigma in adults with adhd. *ADHD attention deficit and hyperactivity disorders*, 11(2):211–220, 2019.

[199] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[200] Rita Gunther McGrath, Ian C MacMillan, and Sari Scheinberg. Elitists, risk-takers, and rugged individualists? an exploratory analysis of cultural differences between entrepreneurs and non-entrepreneurs. *Journal of business venturing*, 7(2):115–135, 1992.

[201] Amanda Meier, Rick Csiernik, Laura Warner, and Cheryl Forchuk. The stigma scale: A canadian perspective. *Social Work Research*, 39(4):213–222, 2015.

[202] Kathy A Mills. What are the threats and potentials of big data for qualitative research? *Qualitative Research*, 18(6):591–603, 2018.

[203] John Mingers, Alistair Mutch, and Leslie Willcocks. Critical realism in information systems research. *MIS quarterly*, 37(3):795–802, 2013.

[204] Tanushree Mitra, Graham P Wright, and Eric Gilbert. A parsimonious language model of social media credibility across disparate events. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 126–145, 2017.

[205] Lee F Monaghan and Jonathan Gabe. Embodying health identities: A study of young people with asthma. *Social Science & Medicine*, 160:1–8, 2016.

[206] Marcela D Monti and Rose M Nealis. Indicators of non-adherence to asthma treatment in pediatric primary care. *Journal of pediatric nursing*, 36:7–12, 2017.

[207] Andrés Montoyo, Patricio MartíNez-Barco, and Alexandra Balahur. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, 2012.

[208] Donald E Morisky. Predictive validity of a medication adherence measure for hypertension control. *Journal of clinical hypertension*, 10:348–354, 2008.

[209] Richard Mortier, Jianxin Zhao, Jon Crowcroft, Liang Wang, Qi Li, Hamed Haddadi, Yousef Amar, Andy Crabtree, James Colley, Tom Lodge, et al. Personal data management with the databox: What's inside the box? In *Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking*, pages 49–54, 2016.

[210] Mark Myslín, Shu-Hong Zhu, Wendy Chapman, and Mike Conway. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of medical Internet research*, 15(8):e174, 2013.

[211] Dawn Nafus and Jamie Sherman. Big data, big questions| this one does not go up to 11: the quantified self movement as an alternative big data practice. *International journal of communication*, 8:11, 2014.

[212] David R Naimi, Tovia G Freedman, Kenneth R Ginsburg, Daniel Bogen, Cynthia S Rand, and Andrea J Apter. Adolescents and asthma: why bother with our meds? *Journal of Allergy and Clinical Immunology*, 123(6):1335–1341, 2009.

[213] Hugo Neffen, Carlos Fritscher, Francisco Cuevas Schacht, Gur Levy, Pascual Chiarella, Joan B Soriano, and Daniel Mechali. Asthma control in latin america: the asthma insights and reality in latin america (airla) survey. *Revista Panamericana de Salud Pública*, 17:191–197, 2005.

[214] Andrew Nickels and Vesselin Dimov. Innovations in technology: social media and mobile technology in the care of adolescents with asthma. *Current allergy and asthma reports*, 12(6):607–612, 2012.

[215] Ipek Kıvılcım Oğuzülgen, Nurdan Köktürk, and Zeynep Işikdoğan. Turkish validation study of morisky 8-item medication adherence questionnaire

(mmas-8) in patients with asthma and chronic obstructive pulmonary disease. *Tuberkuloz ve toraks*, 62(2):101–107, 2014.

[216] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, et al. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, 2011.

[217] Magdalena Olszanecka-Glinianowicz and Agnieszka Almgren-Rachtan. The adherence and illness perception of patients diagnosed with asthma or chronic obstructive pulmonary disease treated with polytherapy using new generation cyclohaler. *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii*, 31(4):235, 2014.

[218] Damilola T Olufemi-Yusuf, Sophie Beaudoin Gabriel, Tatiana Makhinova, and Lisa M Guirguis. "being in control of my asthma myself" patient experience of asthma management: A qualitative interpretive description. *Pharmacy*, 6(4):121, 2018.

[219] Liesl Marten Osman. Psychological factors in asthma control and attack risk, 2002.

[220] Sankar K Pal. Soft data mining, computational theory of perceptions, and rough-fuzzy approach. *Information Sciences*, 163(1-3):5–12, 2004.

[221] Zizi Papacharissi. Without you, i'm nothing: Performances of the self on twitter. *International journal of communication*, 6:18, 2012.

[222] Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934, 2015.

[223] Yoon Soo Park, Lars Konge, and Anthony R Artino. The positivism paradigm of research. *Academic Medicine*, 95(5):690–694, 2020.

[224] Michael J Paul and Mark Dredze. A model for mining public health topics from twitter. *Health*, 11(16-16):1, 2012.

[225] Bernice A Pescosolido. The public stigma of mental illness: What do we think; what do we know; what can we prove? *Journal of Health and Social behavior*, 54(1):1–21, 2013.

[226] Keith J Petrie and John Weinman. Patients' perceptions of their illness: The dynamo of volition in health care. *Current directions in psychological science*, 21(1):60–65, 2012.

[227] Jeffrey Pickens. Attitudes and perceptions. *Organizational behavior in health care*, 4(7):43–76, 2005.

[228] Georgios K Pitsilis, Heri Ramampiaro, and Helge Langseth. Detecting offensive language in tweets using deep learning. *arXiv preprint arXiv:1801.04433*, 2018.

[229] Vicente Plaza, Concepción Fernández-Rodríguez, Carlos Melero, Borja G Cosío, Luís Manuel Entrenas, Luis Pérez De Llano, Fernando Gutiérrez-Pereyra, Eduard Tarragona, Rosa Palomino, Antolín López-Viña, et al. Validation of the 'test of the adherence to inhalers'(tai) for asthma and copd patients. *Journal of aerosol medicine and pulmonary drug delivery*, 29(2):142–152, 2016.

[230] David Price, Monica Fletcher, and Thys Van Der Molen. Asthma control and management in 8,000 european patients: the recognise asthma and link to symptoms and experience (realise) survey. *NPJ primary care respiratory medicine*, 24(1):1–10, 2014.

[231] Jennifer LP Protudjer, Anita L Kozyrskyj, Allan B Becker, and Gail Marchessault. Normalization strategies of children with asthma. *Qualitative Health Research*, 19(1):94–104, 2009.

[232] Steven Pryjmachuk and David A Richards. Predicting stress in pre-registration nursing students. *British Journal of health psychology*, 12(1):125–144, 2007.

[233] Matthew Purver and Stuart Battersby. Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–491, 2012.

[234] Claudia Put, Omer Van den Bergh, Valentine Lemaigre, Maurice Demedts, and Geert Verleden. Evaluation of an individualised asthma programme directed at behavioural change. *European Respiratory Journal*, 21(1):109–115, 2003.

[235] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 180–185. IEEE, 2011.

[236] Daniele Quercia, Renaud Lambiotte, David Stillwell, Michal Kosinski, and Jon Crowcroft. The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 955–964, 2012.

[237] Diane M Quinn and Valerie A Earnshaw. Concealable stigmatized identities and psychological well-being. *Social and personality psychology compass*, 7(1):40–51, 2013.

[238] Peter Ragaert, Wim Verbeke, Frank Devlieghere, and Johan Debevere. Consumer perception and choice of minimally processed vegetables and packaged fruits. *Food quality and preference*, 15(3):259–270, 2004.

[239] Sudha Ram, Wenli Zhang, Max Williams, and Yolande Pengetnze. Predicting asthma-related emergency department visits using big data. *IEEE journal of biomedical and health informatics*, 19(4):1216–1223, 2015.

[240] Cynthia S Rand. Adherence to asthma therapy in the preschool child. *Allergy*, 57:48–57, 2002.

[241] Cynthia S Rand, Mitchell Nides, Mary Kathryn Cowles, Robert A Wise, and John Connett. Long-term metered-dose inhaler adherence in a clinical trial. the lung health study research group. *American journal of respiratory and critical care medicine*, 152(2):580–588, 1995.

[242] Eshrag Refaee and Verena Rieser. Evaluating distant supervision for subjectivity and sentiment analysis on arabic twitter feeds. In *Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP)*, pages 174–179, 2014.

[243] W Jack Rejeski and Jason Fanning. Models and theories of health behavior and clinical interventions in aging: a contemporary, integrative approach. *Clinical interventions in aging*, 14:1007, 2019.

[244] Ruben D Restrepo and Jay Peters. Near-fatal asthma: recognition and management. *Current opinion in pulmonary medicine*, 14(1):13–23, 2008.

[245] Hyekyun Rhee, Michael J Belyea, and Judith Brasch. Family support and asthma outcomes in adolescents: barriers to adherence as a mediator. *Journal of adolescent health*, 47(5):472–478, 2010.

[246] Hyekyun Rhee, Michael J Belyea, Susan Ciurzynski, and Judith Brasch. Barriers to asthma self-management in adolescents: Relationships to psychosocial factors. *Pediatric pulmonology*, 44(2):183–191, 2009.

[247] Hyekyun Rhee, Mona N Wicks, Jennifer S Dolgoff, Tanzy M Love, and Donald Harrington. Cognitive factors predict medication adherence and

asthma control in urban adolescents with asthma. *Patient preference and adherence*, 12:929, 2018.

[248] Nizar Rifaat, Elham Abdel-Hady, and Ali A Hasan. The golden factor in adherence to inhaled corticosteroid in asthma patients. *Egyptian journal of chest diseases and tuberculosis*, 62(3):371–376, 2013.

[249] Ellen Riloff, Janyce Wiebe, and William Phillips. Exploiting subjectivity classification to improve information extraction. In *AAAI*, pages 1106–1111, 2005.

[250] Graham Roberts. Predicting the long-term outcome of preschool wheeze: are we there yet? *Journal of Allergy and Clinical Immunology*, 124(5):911–912, 2009.

[251] Frank Rockhold, Perry Nisen, and Andrew Freeman. Data sharing at a crossroads. *New England Journal of Medicine*, 375(12):1115–1117, 2016.

[252] Simona Romani, Silvia Grappi, and Daniele Dalli. Emotions that drive consumers away from brands: Measuring negative emotions toward brands and their behavioral effects. *International Journal of Research in marketing*, 29(1):55–67, 2012.

[253] Jenni Romaniuk and Byron Sharp. Measuring brand perceptions: Testing quantity and quality. *Journal of targeting, measurement and analysis for marketing*, 11(3):218–229, 2003.

[254] Shiho Rose, Christine Paul, Allison Boyes, Brian Kelly, and Della Roach. Stigma-related experiences in non-communicable respiratory diseases: a systematic review. *Chronic respiratory disease*, 14(3):199–216, 2017.

[255] Catherine E Ross and John Mirowsky. Age and the balance of emotions. *Social Science & Medicine*, 66(12):2391–2400, 2008.

[256] Wendy D Roth and Jal D Mehta. The rashomon effect: Combining positivist and interpretivist approaches in the analysis of contested events. *Sociological methods & research*, 31(2):131–173, 2002.

[257] Seref Sagiroglu and Duygu Sinanc. Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)*, pages 42–47. IEEE, 2013.

[258] Graham Scambler and Anthony Hopkins. Being epileptic: coming to terms with stigma. *Sociology of health & illness*, 8(1):26–43, 1986.

[259] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control*, 38(3):182–188, 2010.

[260] Marianne Hansson Scherman and Olle Löwhagen. Drug compliance and identity: reasons for non-compliance: experiences of medication from persons with asthma/allergy. *Patient Education and Counseling*, 54(1):3–9, 2004.

[261] Friederike Schultz, Sonja Utz, and Anja Göritz. Is the medium the message? perceptions of and reactions to crisis communication via twitter, blogs and traditional media. *Public relations review*, 37(1):20–27, 2011.

[262] Andreas Schwab and Zhu Zhang. A new methodological frontier in entrepreneurship research: Big data studies, 2019.

[263] Yoko Setoyama, Yoshihiko Yamazaki, and Kazuhiro Namayama. Benefits of peer support in online japanese breast cancer communities: differences between lurkers and posters. *Journal of medical Internet research*, 13(4):e122, 2011.

[264] Tom Shakespeare et al. The social model of disability. *The disability studies reader*, 2:197–204, 2006.

[265] Tamar Sharon and Federica Lucivero. Introduction to the special theme: The expansion of the health data ecosystem–rethinking data ethics and governance, 2019.

[266] Galit Shmueli. To explain or to predict? *Statistical science*, 25(3):289–310, 2010.

[267] Bonnie Sibbald. Patient self care in acute asthma. *Thorax*, 44(2):97–101, 1989.

[268] Lauren Sinnenberg, Alison M Buttenheim, Kevin Padrez, Christina Mancheno, Lyle Ungar, and Raina M Merchant. Twitter as a tool for health research: a systematic review. *American journal of public health*, 107(1):e1–e8, 2017.

[269] Juan Sixto, Aitor Almeida, and Diego López-de Ipiña. Analysis of the structured information for subjectivity detection in twitter. In *Transactions on Computational Collective Intelligence XXIX*, pages 163–181. Springer, 2018.

[270] Anya Skatova and James Goulding. Psychology of personal data donation. *PloS one*, 14(11):e0224240, 2019.

[271] Anya Skatova, Jaspreet Johal, Robert Houghton, Richard Mortier, Neelam Bhandari, Tom Lodge, Christian Wagner, James Goulding, Jon Crowcroft, and Anil Madhavapeddy. Perceived risks of personal data sharing. *Proc. Digital Economy: Open Digital (Nov. 2013)*, 2013.

[272] Anya Skatova, Esther Ng, and James Goulding. Data donation: Sharing personal data for public good. *Application of Digital Innovation. London, England: N-Lab*, 2014.

[273] Tamara A Small. What the hashtag? a content analysis of canadian politics on twitter. *Information, communication & society*, 14(6):872–895, 2011.

[274] Andrew Smith, Leigh Sparks, and James Goulding. Using commercial big data to inform social policy: Possibilities, ethics, methods and obstacles. In *Journal of Macromarketing*, volume 35, pages 141–141. SAGE PUBLICATIONS INC 2455 TELLER RD, THOUSAND OAKS, CA 91320 USA, 2015.

[275] Craig A Smith and Phoebe C Ellsworth. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48(4):813, 1985.

[276] Gavin Smith, Roberto Mansilla, and James Goulding. Model class reliance for random forests. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[277] DAVID SNADDEN and JUDITH BELLE BROWN. Asthma and stigma. *Family practice*, 8(4):329–335, 1991.

[278] David Snadden and Judith Belle Brown. The experience of asthma. *Social science & medicine*, 34(12):1351–1361, 1992.

[279] Beverley A Sparks and Victoria Browning. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, 32(6):1310–1323, 2011.

[280] Henry Pierce Stapp. The copenhagen interpretation. *American Journal of Physics*, 40(8):1098–1116, 1972.

[281] Cheryl A Steiman, Ves Dimov, and Frank J Eidelman. Twitter as a new medium for public health advocacy: Asthma, food allergy and allergic rhinitis. *Journal of Allergy and Clinical Immunology*, 135(2):AB69, 2015.

[282] Carol S Stilley, Susan Sereika, Matthew F Muldoon, Christopher M Ryan, and Jacqueline Dunbar-Jacob. Psychological and cognitive function: predictors of adherence with cholesterol lowering treatment. *Annals of Behavioral Medicine*, 27(2):117–124, 2004.

[283] David Stillwell. Modern psychometrics: The science of psychological assessment. 2020.

[284] Suphat Sukamolson. Fundamentals of quantitative research. *Language Institute Chulalongkorn University*, 1:2–3, 2007.

[285] Gail M Sullivan and Richard Feinn. Using effect size—or why the p value is not enough. *Journal of graduate medical education*, 4(3):279–282, 2012.

[286] Tiffany H Taft, Sarah Ballou, and Laurie Keefer. A preliminary evaluation of internalized stigma and stigma resistance in inflammatory bowel disease. *Journal of Health Psychology*, 18(4):451–460, 2013.

[287] Abbas Tashakkori and Charles Teddlie. Issues and dilemmas in teaching research methods courses in social and behavioural sciences: Us perspective. *International journal of social research methodology*, 6(1):61–77, 2003.

[288] Mike Thomas, Kevin Gruffydd-Jones, Carol Stonham, Sabbi Ward, and Tatiana Macfarlane. Assessing asthma control in routine clinical practice: use of the royal college of physicians '3 questions'. *Primary Care Respiratory Journal*, 18(2):83–88, 2009.

[289] Luís Torgo and João Gama. Regression by classification. In *Brazilian symposium on artificial intelligence*, pages 51–60. Springer, 1996.

[290] Olivier Toubia and Andrew T Stephen. Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science*, 32(3):368–392, 2013.

[291] Declan J Troy and JP Kerry. Consumer perception and the role of science in the meat industry. *Meat science*, 86(1):214–226, 2010.

[292] Daniel W Turner III. Qualitative interview design: A practical guide for novice investigators. *The qualitative report*, 15(3):754, 2010.

[293] Marina Vamos and John Kolbe. Psychological factors in severe chronic asthma. *Australian & New Zealand Journal of Psychiatry*, 33(4):538–544, 1999.

[294] Wim H Van Brakel. Measuring health-related stigma—a literature review. *Psychology, health & medicine*, 11(3):307–334, 2006.

[295] Monique OM Van De Ven and Rutger CME Engels. Quality of life of adolescents with asthma: the role of personality, coping strategies, and symptom reporting. *Journal of Psychosomatic Research*, 71(3):166–173, 2011.

[296] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10(66-71):13, 2009.

[297] José Van Dijck. *The culture of connectivity: A critical history of social media*. Oxford University Press, 2013.

[298] Sandra van Dulmen, Emmy Sluijs, Liset Van Dijk, Denise de Ridder, Rob Heerdink, and Jozien Bensing. Patient adherence to medical treatment: a review of reviews. *BMC health services research*, 7(1):1–13, 2007.

[299] Ryan J Van Lieshout and Glenda MacQueen. Psychological factors in asthma. *Allergy, Asthma & Clinical Immunology*, 4(1):1–17, 2008.

[300] Ryan J Van Lieshout and Glenda M MacQueen. Relations between asthma and psychological distress: an old idea revisited. *Allergy and the nervous system*, 98:1–13, 2012.

[301] Effy Vayena, Marcel Salathé, Lawrence C Madoff, and John S Brownstein. Ethical challenges of big data in public health, 2015.

[302] Maria Paz Garcia Villalba and Patrick Saint-Dizier. Some facets of argument mining for opinion analysis. *COMMA*, 245:23–34, 2012.

[303] Svitlana Volkova, Yoram Bachrach, and Benjamin Van Durme. Mining user interests to predict perceived psycho-demographic traits on twitter. In *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 36–43. IEEE, 2016.

[304] Ngoc Thang Vu, Pankaj Gupta, Heike Adel, and Hinrich Schütze. Bi-directional recurrent neural network with ranking loss for spoken language understanding. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6060–6064. IEEE, 2016.

[305] Richard D Waters and Jia Y Jamal. Tweet, tweet, tweet: A content analysis of nonprofit organizations' twitter updates. *Public Relations Review*, 37(3):321–324, 2011.

[306] David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988.

[307] Roger Watson. Quantitative research. *Nursing Standard (2014+)*, 29(31):44, 2015.

[308] CM Whissell. The dictionary of affect in language, emotion: Theory, research and experience. the measurement of emotions, r. plutchik and h. kellerman, eds., vol. 4, 1989.

[309] Sue Widdicombe and Rob Wooffitt. 'being'versus' doing'punk: On achieving authenticity as a member. *Journal of language and social psychology*, 9(4):257–277, 1990.

[310] Alicja A Wieczorkowska. Towards extracting emotions from music. In *Intelligent Media Technology for Communicative Intelligence*, pages 228–238. Springer, 2004.

[311] Rand R Wilcox. Robust generalizations of classical test reliability and cronbach's alpha. *British J. of Mathematical and Statistical Psychology*, 45(2):239–254, 1992.

[312] Matthew L Williams, Pete Burnap, and Luke Sloan. Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51(6):1149–1168, 2017.

[313] Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. Opinionfinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP 2005 Interactive Demonstrations*, pages 34–35, 2005.

[314] Ian Wood and Sebastian Ruder. Emoji as emotion tags for tweets. In *Proceedings of the Emotion and Sentiment Analysis Workshop LREC2016, Portorož, Slovenia*, pages 76–79, 2016.

[315] H Yang, X Li, B Stanton, X Fang, D Lin, and S Naar-King. Hiv-related knowledge, stigma, and willingness to disclose: A mediation analysis. *AIDS care*, 18(7):717–724, 2006.

[316] TienYu Owen Yang, Kathy Sylva, and Ingrid Lunt. Parent support, peer support, and peer acceptance in healthy lifestyle for asthma management among early adolescents. *Journal for Specialists in Pediatric Nursing*, 15(4):272–281, 2010.

[317] Tal Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373, 2010.

[318] Tal Yarkoni and Jacob Westfall. Choosing prediction over explanation in

psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6):1100–1122, 2017.

[319] Sierk Ybema, Tom Keenoy, Cliff Oswick, Armin Beverungen, Nick Ellis, and Ida Sabelis. Articulating identities. *Human relations*, 62(3):299–322, 2009.

[320] Markos Zachariadis, Susan Scott, and Michael Barrett. Methodological implications of critical realism for mixed-methods research. *MIS quarterly*, pages 855–879, 2013.

[321] Sandra E Zaeh, Monica A Lu, Kathryn V Blake, Elizabeth Ruvalcaba, Christabelle Ayensu-Asiedu, Robert A Wise, Janet T Holbrook, and Michelle N Eakin. "it is kind of like a responsibility thing": transitional challenges in asthma medication adherence among adolescents and young adults. *Journal of Asthma*, pages 1–11, 2021.

[322] Bouziane Zaid, Don Dongshee Shin, Sarah Waled Kteish, Jana Fedtke, and Mohammed Ibahrine. Gendered self-representation and empowerment on social media in the united arab emirates. *The Communication Review*, pages 1–19, 2021.

[323] Chun-Mei Zhao and Jing Luan. Data mining: Going beyond traditional statistics. *New Directions for Institutional Research*, 131:7–16, 2006.

[324] Michael Zimmer. "but the data is already public": on the ethics of research in facebook. In *The Ethics of Information Technologies*, pages 229–241. Routledge, 2020.

Appendices

Appendix A

Semi-structured interview protocol

Information for Research Participants

Thank you for agreeing to participate in the research project. Your participation in this research is voluntary, and you may change your mind about being involved in the research at any time, and without giving a reason. This information sheet is designed to give you full details of the research project, its goals, the research team, the research funder, and what you will be asked to do as part of the research. If you have any questions that are not answered by this information sheet, please ask. This research has been reviewed and given favourable opinion by the Nottingham University Business School Research Ethics Committee.

Title

Exploring perceptions about asthma

Research aims

This study seeks to reveal what perceptions people have about asthma. This is the first part of a PhD that explores patient-related factors that affect adherence to asthma medication, namely perceptions (social-cognition), situational and behavioural factors.

Participants

You have been asked to be a part of this research as you have been diagnosed with asthma. Since this study is all about asthma perceptions, we are recruiting a range of participants with varied ages, backgrounds, professions etc.

Interview content

You will be asked to engage in a conversation with the researcher, in the form of an interview, in which the researcher will ask you a range of questions. This interview does not have a strict structure, in order to allow for participants to express their perceptions about asthma as freely as possible. You will be asked about your opinions, experiences and feelings about asthma. Interview will take approximately 1 hour and will be recorded on a mobile phone, with your consent.

Data collection and handling

Your data will be stored in accordance with the Data Protection Act 1998 on a password-protected server at the University of Nottingham. Access will be restricted to include only those who are directly involved with the research. Any personal identifying information will be deleted after transcription and participants' names will be changed for pseudonyms. All the interviews will then be transcribed. Audio data as well as transcribed interviews will be used for the analysis, but will not be shown or made public under any circumstances or re-distributed to other researchers who are not included in this particular study. You have the right to withdraw your consent at any point, before, during or after the interview, for any reason, without penalty up until the point of anonymization at which point we will not be able to identify your data. In the event of consent withdrawal, all of participants' identifiable personal data (and related derived data) will be erased.

Research outputs

Thematic analysis will be used to extract the relevant themes and explore what perceptions emerged in conversations. The project aims to publish its findings in academic journals and conferences. All your data (audio, personal information and answers to questionnaires) will be stored not by using your name but with pseudonyms that cannot be traced back to you.

Complaint procedure

If you wish to complain about the way in which the research is being conducted or have any concerns about the research then in the first instance please contact the (Principal Investigator or supervisor) or contact the School's Research Ethics Officer, Chris Carter (christopher.carter@nottingham.ac.uk).

Interview guide

Prior to commencing the interview, participants were asked to sign the Consent Form, GDPR form and Information form for participants (described above). Participants were asked whether they are comfortable with interview being recorded (and transcribed). Following their positive response, the interview could start.

General questions

The first set of questions was used as an icebreaker, to establish the rapport with participants. There were questions about participants *age*, *city*, *profession or education.*

Questions about asthma duration and treatment adherence

The following set of questions was related to general characteristics of participants' asthma. To illustrate, these were questions such as *'What do you do to keep your asthma in check?'* and *'Do you know what triggers your asthma?'.*

Questions about participants' perceptions about asthma

This block of questions was focused on uncovering participants' general experiences about life with asthma (e.g. 'How does it feel to have asthma?'); their own perceptions about asthma (e.g. 'What is the first word that comes to your mind when someone mentions asthma?'.

Questions about perceptions *others* have about asthma

Questions in this section aimed to uncover which perceptions participants think others have about them - in the light of their condition (e.g. *'What do you think other people think when they hear you have asthma?'*) and *'How do you feel about sharing your experiences related to asthma in front of others?'.*

Questions about the impact asthma has on one's life

These were questions about the practical and pscyhological impact asthma has on ones life - including questions about coping mechanisms such as humour (e.g. *'Do you ever joke about asthma?'*.

Questions about the inhaler

The final set of questions was dedicated to questions about inhaler attachment and inhaler use. For example: *'How do you feel about your inhaler?'* and *'Do you use your inhaler in public?'*.

The interview would finish with the question: *'Do you have any other comments to make in general about this topic?'*.

Appendix B

Topic Modelling tables from Study 2

The following graphs for each of the four groups of perceptions were generated using Topic Modelling process (LDA algorithm), as discussed in more detail in Chapter 4.

B.0.1 Active positive perceptions

Four topics were chosen for the active positive group of perceptions as demonstrated in Figure B.1.

In terms of the most relevant words for each topic in Active positive perceptions, the following graphs contain more information, as presented in Figures: B.2, B.3, B.4 and B.5.

B.0.2 Passive negative perceptions perceptions

There were also four topics that were chosen for the passive negative group of perceptions, as demonstrated in Figure B.6.

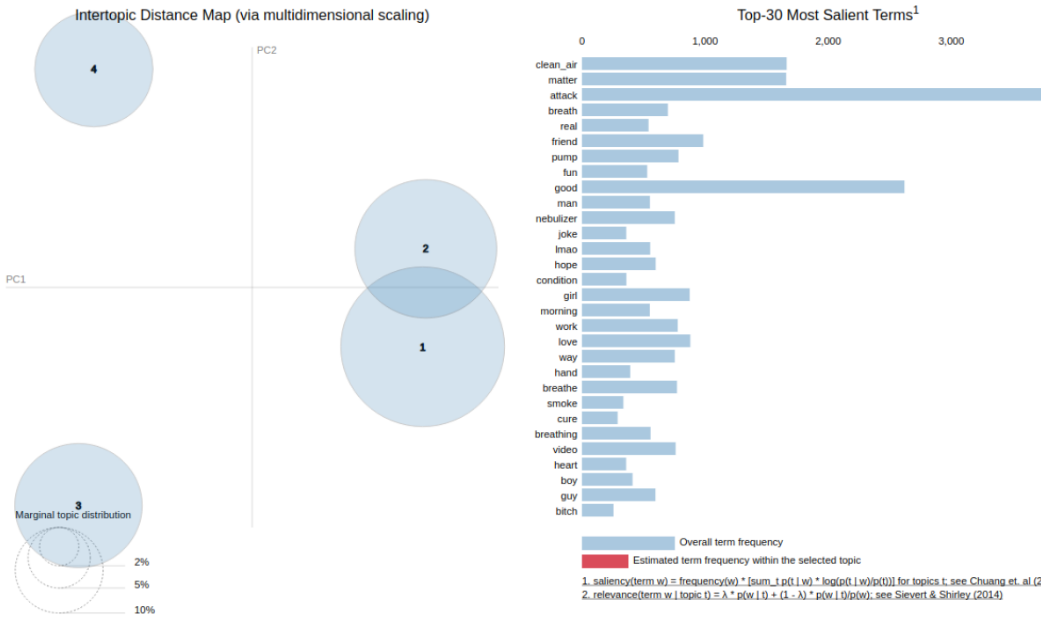

Figure B.1: Topic Modelling results for Active Positive Perceptions

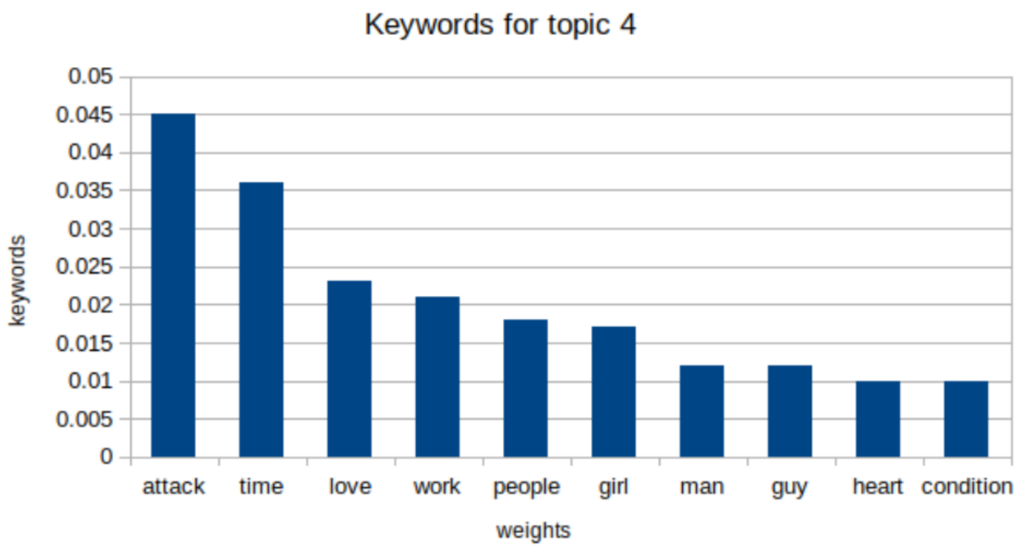

Figure B.2: Key words in the topic named: Empowering support / Disparaging humour

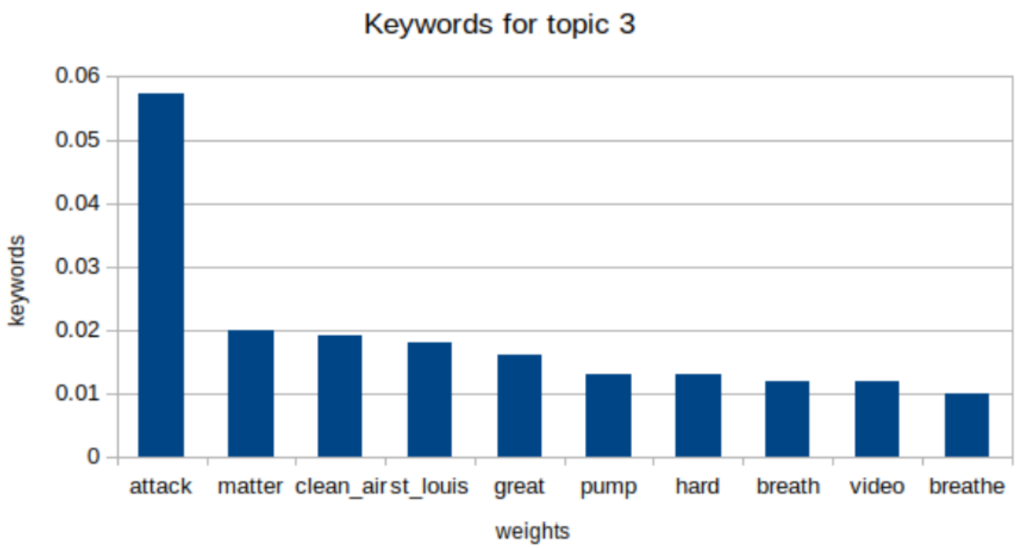

Figure B.3: Key words in the topic named: Support by raising awareness / Positive humour

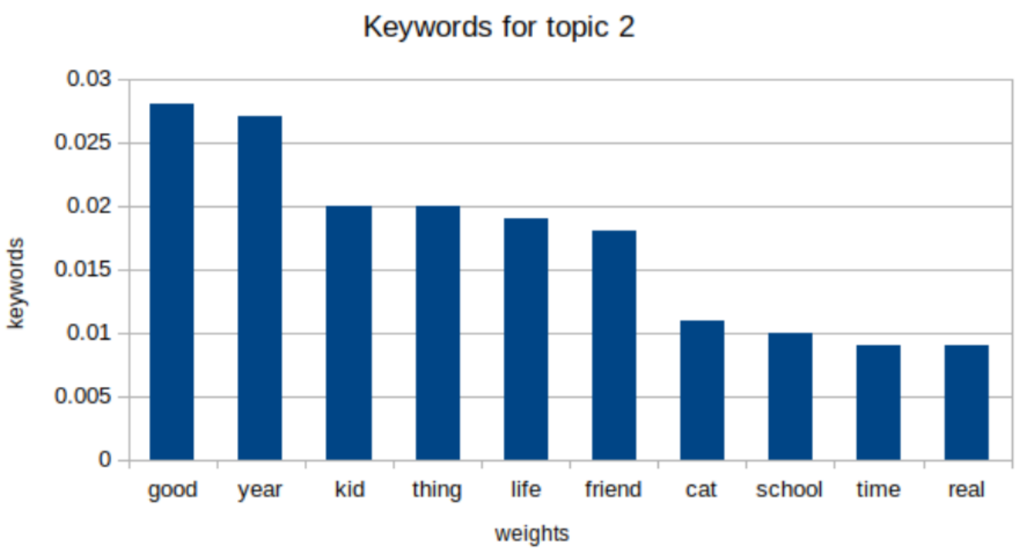

Figure B.4: Key words in the topic named: Support by sharing experiences

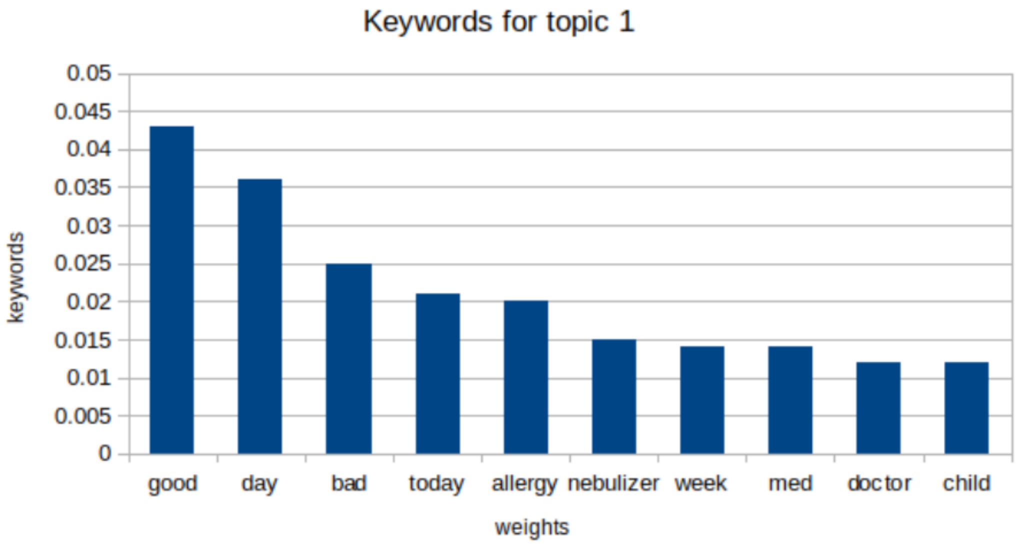

Figure B.5: Key words in the topic named: Practical tips

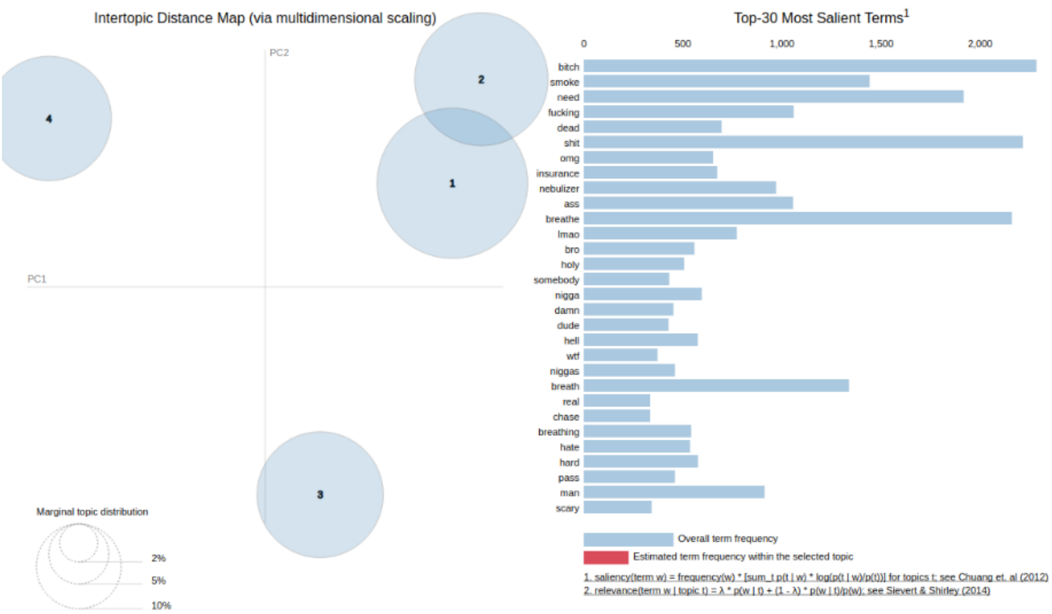

Figure B.6: Topic Modelling results for Passive Negative Perceptions

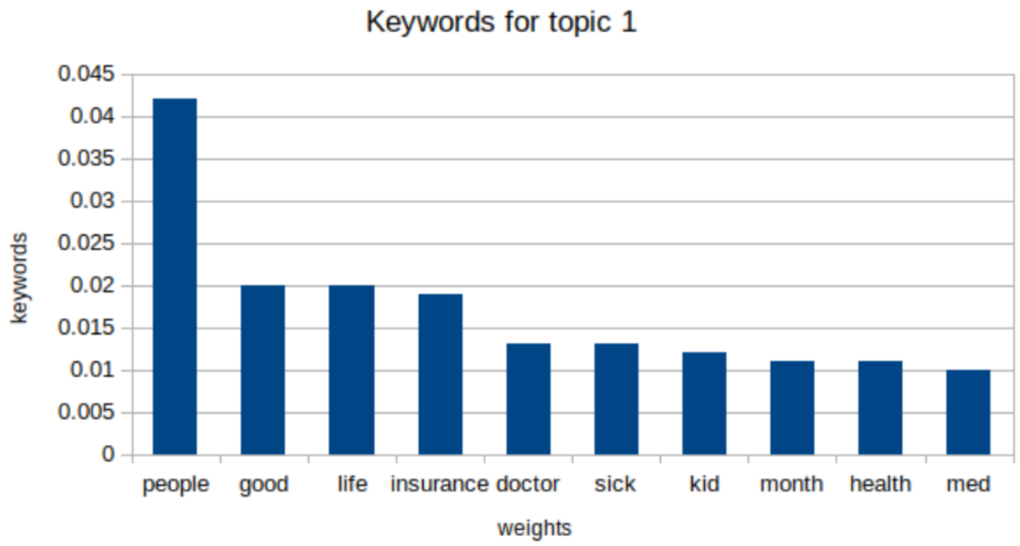

Figure B.7: Key words in the topic named: Politically charged disappointment

Keywords for each topic in passive negative perceptions are presented in Figures: B.7, B.8, B.9 and B.10.

B.0.3 Passive positive perceptions

This group of perceptions also had four topics, as demonstrated in Figure B.11.

The representative keywords for each topic in passive positive perceptions are presented in Figures: B.12, B.13, B.14 and B.15.

B.0.4 Active negative perceptions

The final group of perceptions had three topics, as demonstrated in Figure B.16.

The relevant keywords for each topic in this group of perceptions are in Figures: B.17, B.18, B.19.

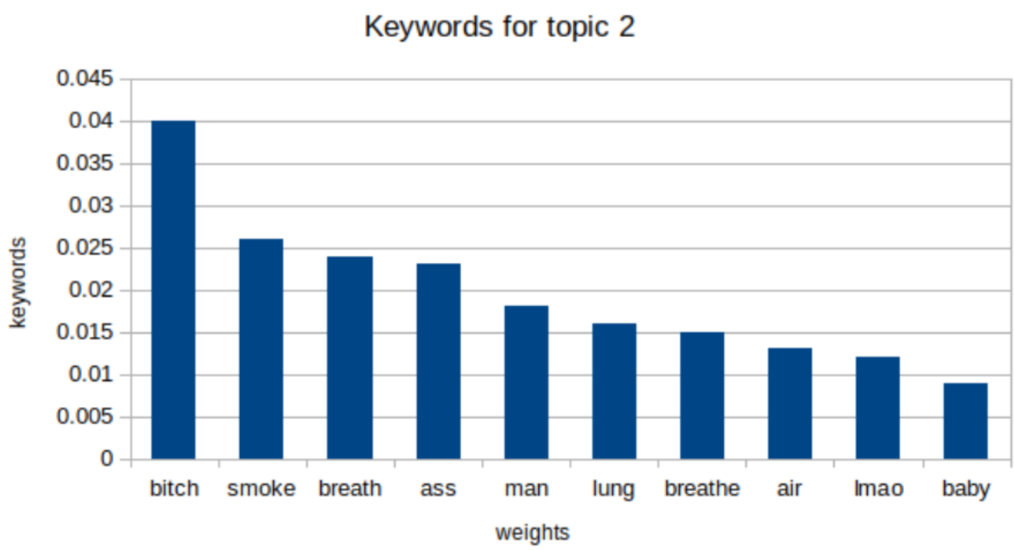

Figure B.8: Key words in the topic named: Complaints about inconsiderate others

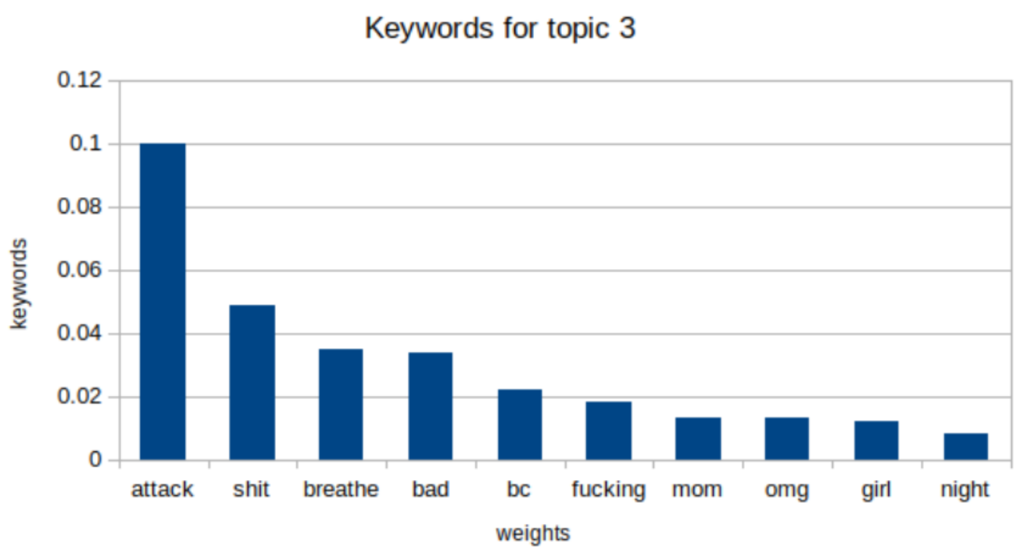

Figure B.9: Key words in the topic named: Sadness due to limitations

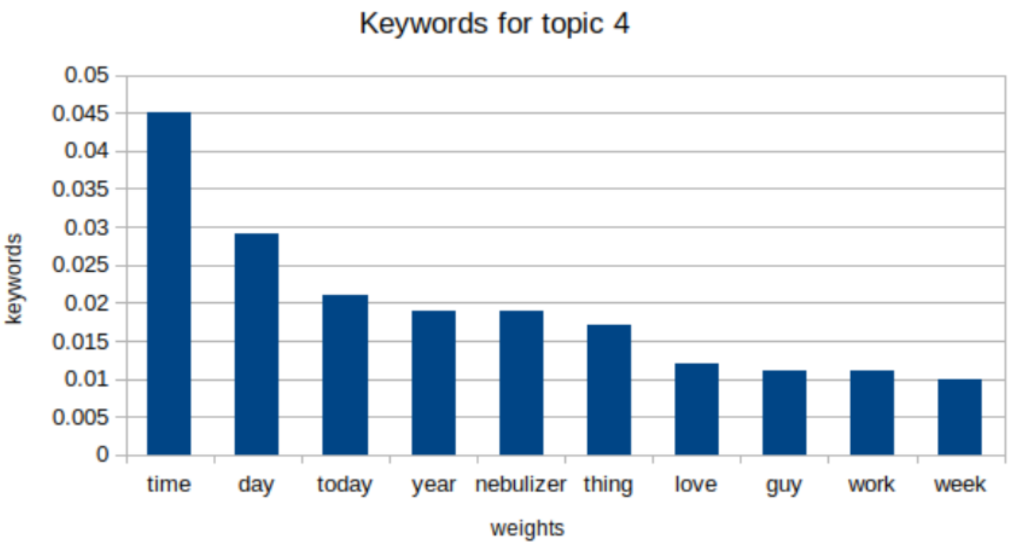

Figure B.10: Key words in the topic named: Sadness due to asthma

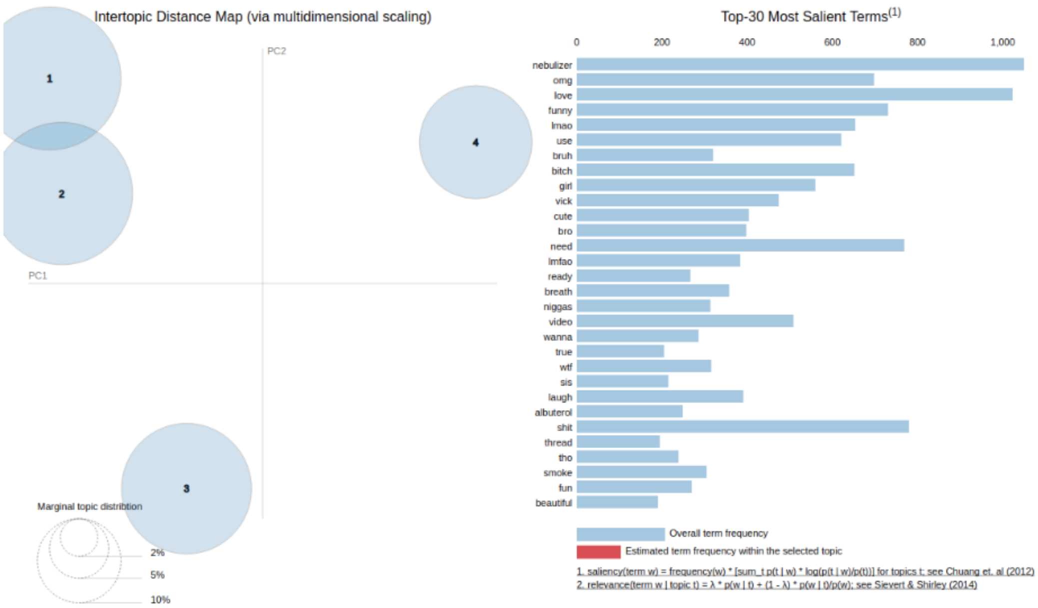

Figure B.11: Topic Modelling results for Passive Positive Perceptions

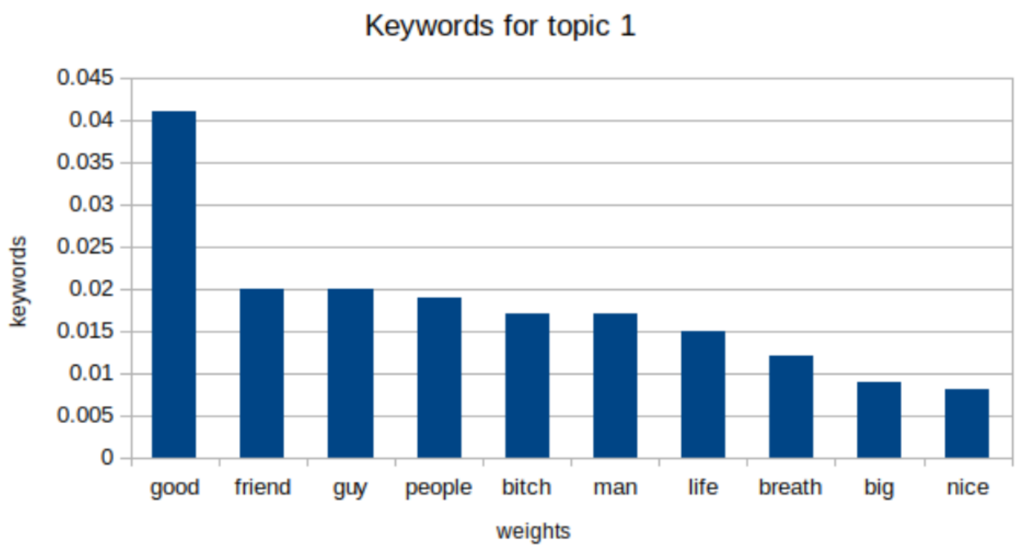

Figure B.12: Key words in the topic named: Gratefulness for other people's actions

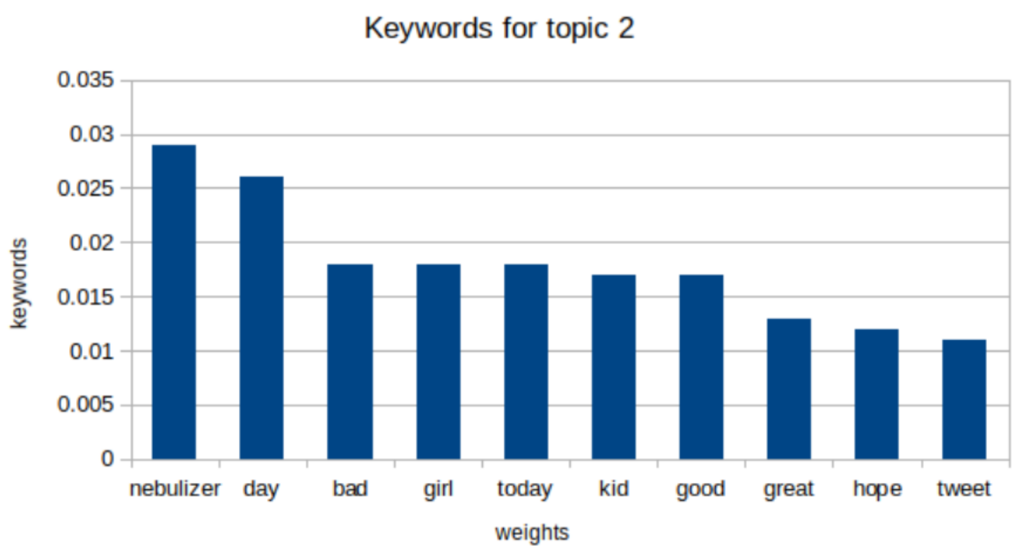

Figure B.13: Key words in the topic named: Content related to animals

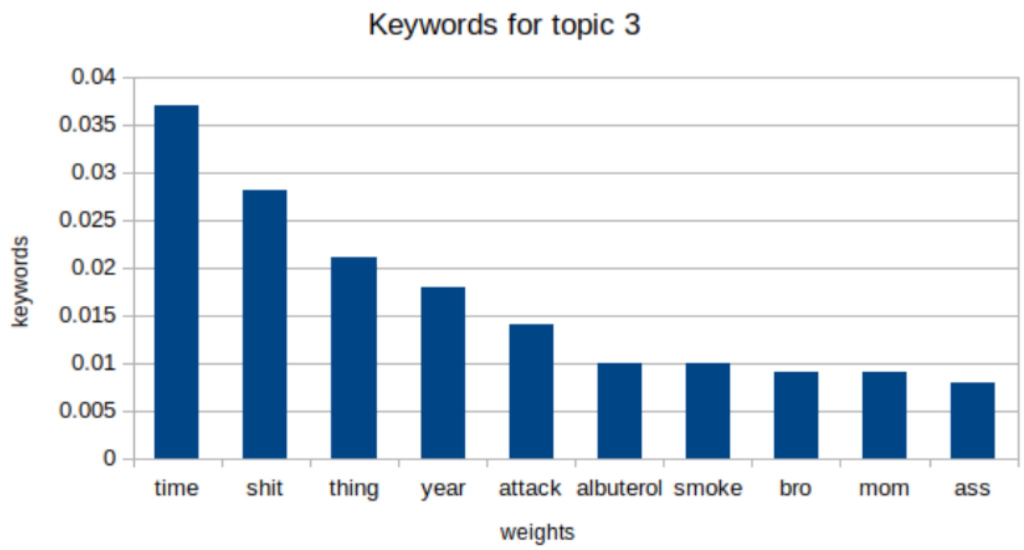

Figure B.14: Key words in the topic named: Feelings of empowerment

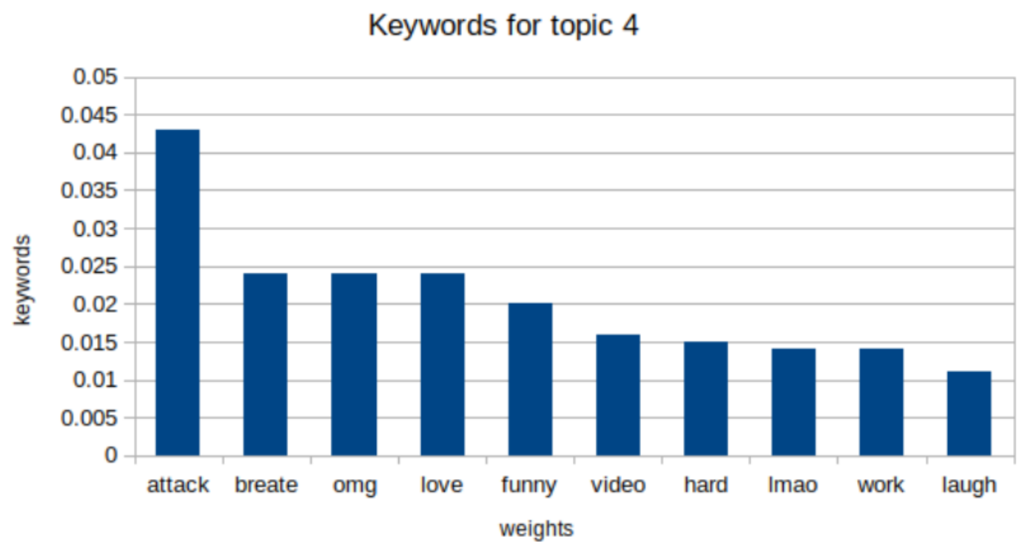

Figure B.15: Key words in the topic named: Positive humour

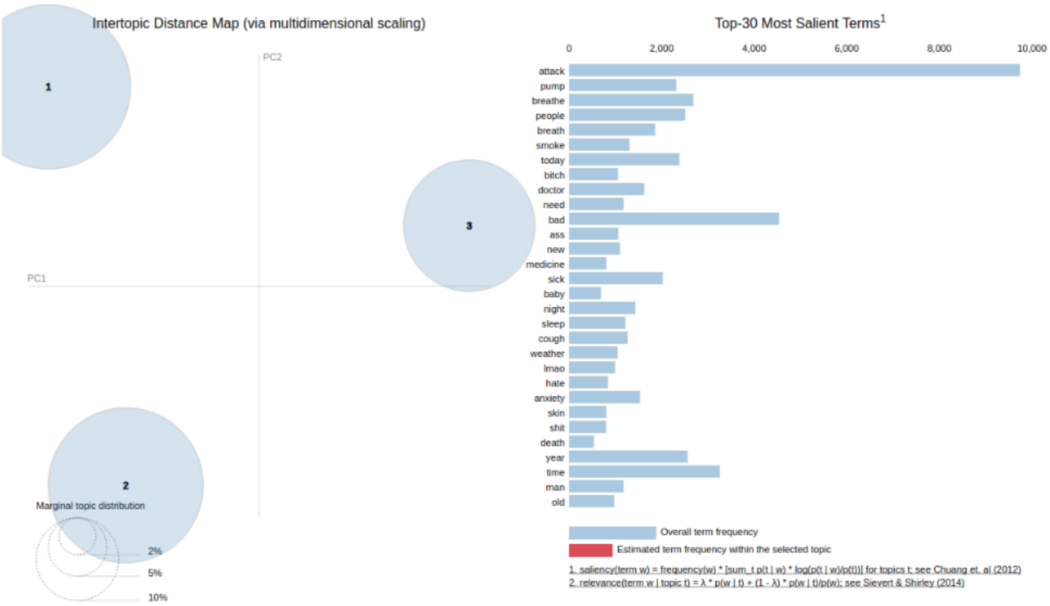

Figure B.16: Topic Modelling results for Active Negative Perceptions

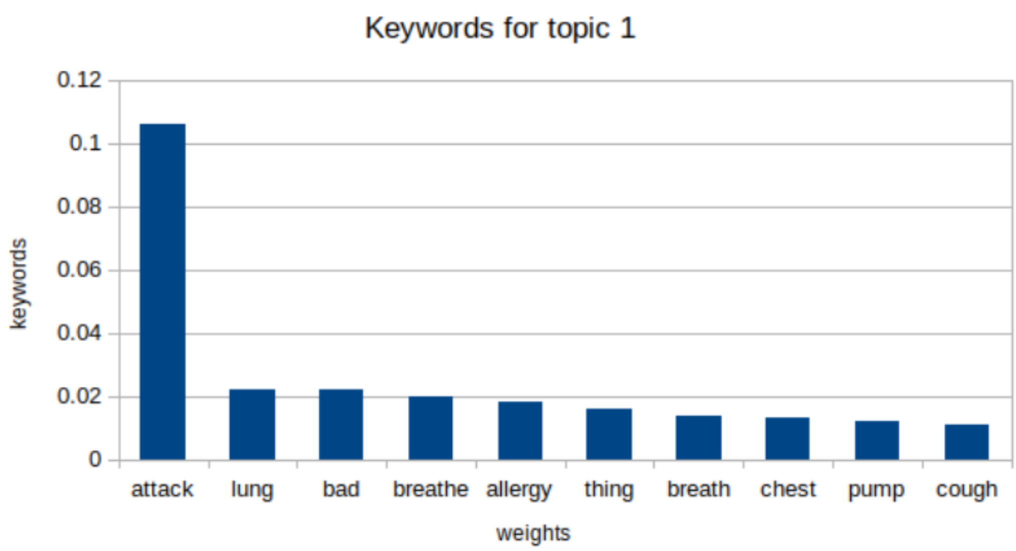

Figure B.17: Key words in the topic named: Frustration with non-patients

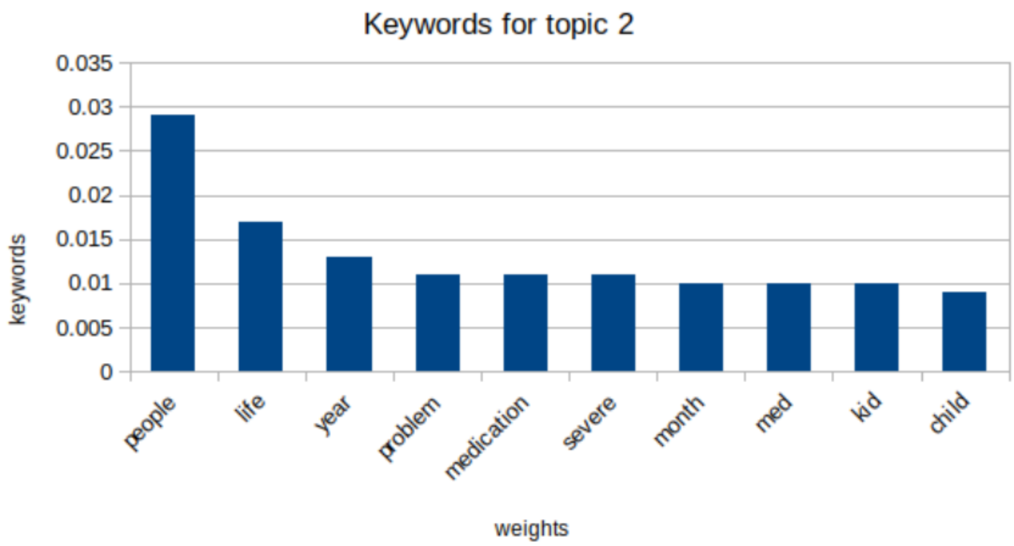

Figure B.18: Key words in the topic named: Frustration with patients

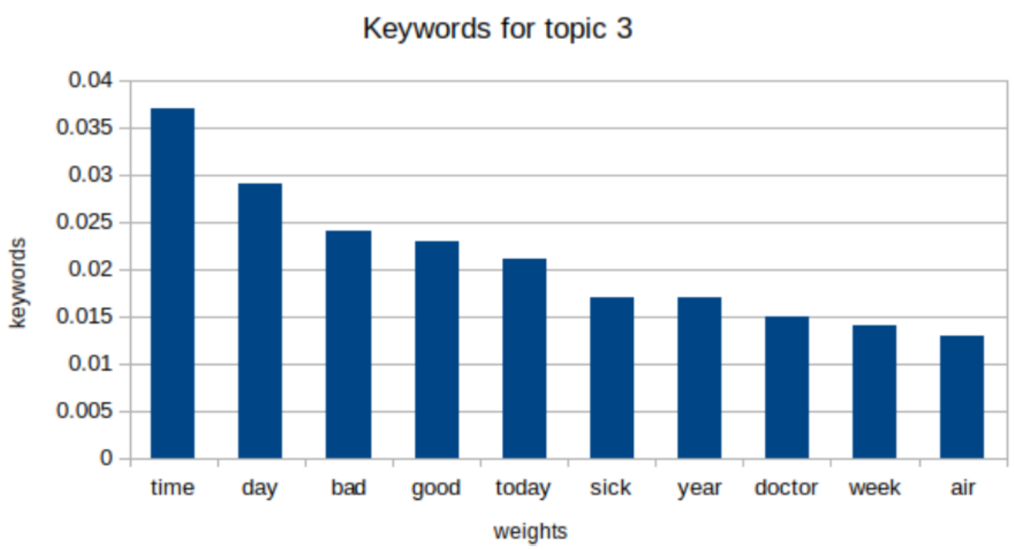

Figure B.19: Key words in the topic named: Anger about asthma

Appendix C

Survey About Asthma Question List

Survey conducted as part of PhD research by Vanja Ljevar. Vanja is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (UKRI Grant No. EP/L015463/1).

List of Survey Questions

Group	**Question**
Adherence score	Do you sometimes forget to take/use your medication?
	Thinking over the past two weeks, were there any days when you forgot to take your medication?
	Have you ever cut back or stopped taking your medications without telling your doctor, because you felt worse when you took it?
	Did you take your medications yesterday?
	When you feel like your health condition is under control, do you sometimes stop taking your medications?
	Do you ever feel hassled about sticking to your treatment plan?

	Prior to COVID-19, how often did you find it difficult to take all your medications simply due to daily life? How often do you have your inhaler with you when you...(travel abroad, go to work, socialise)
Asthma traits	I am diagnosed with mild; severe asthma How long have you had asthma? Which type of inhaler are you prescribed with?
Demographics	What is your age? What is your gender? What is the highest level of education you have completed? What is your current work status? What is your total annual household income before tax? Do you have any allergies?
Perceptions	People sometimes joke about asthma to put me down. People would make jokes about asthma in front of me. I do not use my inhaler in public because people might make fun of me. Media (films, TV shows, etc.) generally portray asthma in a positive light. Society perceives people with asthma as strong. People would treat me differently if they knew I had asthma. I have been discriminated against at work because of my asthma. Sometimes I feel that I am being talked down to because of my asthma. I worry about telling people I use an inhaler. Some people with asthma are weak. I do not think people understand what asthma really is.

I think people with asthma are as reliable co-workers as anybody else.

I worry about how people might react if they found out about my asthma.

I would have had better chances in life if I had not had asthma.

I do not tell people at my workplace that I have asthma.

I prefer if people did not see me using my inhaler.

I am angry with the way some people have reacted about my asthma.

I have not had any trouble with people because of my asthma.

Even though I am diagnosed, I think I may not have asthma.

My asthma is not as serious as my doctor and my diagnosis say it is.

Using an inhaler means you are not coping well with your asthma.

I would not ask other people to change their behaviour when it irritates my asthma, because it is my own problem.

I get valuable information on how to cope with asthma from people online (social media, asthma online groups, etc.).

I feel more understood about my asthma problems by people online than other people in my life.

I don't think that support groups for people with asthma are of any use to me.

I am concerned that I might get incorrect information about asthma from people I talk to online.

Realizing that my experience is not unique helped me cope better with my asthma.

I would rather suffer from cigarette smoke than explain to others that I have asthma.

Coping mechanisms	If asthma gets me down, I think about those who are even more seriously ill than I am. I avoid exertion. I do not try to live cautiously to avoid shortness of breath. I like talking about asthma. I try to hide my asthma. I try to think about having asthma in a positive light. I try to mature as a person through asthma-related experiences. I often try to find out more about asthma. I take note of new updates in the media concerning asthma. I try to forget that I have asthma. I pretend that asthma does not bother me at all. I am afraid that my asthma will get worse. I take my medication in advance if I think I may be in a situation that causes me shortness of breath.
Emotions	How I feel about having asthma: Sad Strong Guilty Scared Hostile Ashamed Nervous Determined Times like this make me feel guilty for not using my medication more regularly.
Lifestyle	Do you smoke? How often do you exercise? COVID-19 has affected how I use my asthma medication.

Personality traits	I see myself as: Extraverted, enthusiastic
	Critical, quarrelsome
	Dependable, self-disciplined
	Anxious, easily upset
	Open to new experiences, complex
	Reserved, quiet
	Sympathetic, warm
	Disorganised, careless
	Calm, emotionally stable
	Conventional, uncreative