2020

Appraisal of free online symptom checkers and applications for self-diagnosis and triage: An Australian evaluation

Michella Gaye Hill
Edith Cowan University

Appraisal of free online symptom checkers and applications for self-diagnosis and triage: An Australian evaluation

This thesis is presented in partial fulfilment of the degree of

Master of Medical and Health Science by Research

Michella Gaye Hill

B.Sc. (Paramedicine)

Edith Cowan University

School of Medical and Health Science

2020

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

I also grant permission for the Library at Edith Cowan University to make duplicate copies of my thesis as required.

Student signature: ....[redacted]..

(Michella Hill)

Primary supervisor signature: [redacted] ....

(Professor Moira Sim)

Co-supervisor signature: ....[redacted]...

(Dr. Brennen Mills)

ACKNOWLEDGEMENTS

This thesis would not have been possible without the genuine support and guidance provided by my supervisors Professor Moira Sim and Dr. Brennen Mills. I am forever grateful for Professor Sim's time and assistance with this research project. She graciously accepted me as a Master by Research student at the 11th hour, following an unexpected leave of absence by my original Lead Supervisor and has given 100% to this project despite the unforeseen increase to her own workload. Professor Sim has been an ongoing resource for medical expertise and information throughout this entire process. Her keen insights and observations have made this project not only possible, but acceptable for publication (yet to be published) with the Medical Journal of Australia.

Dr. Mills has been my inspiration to follow a research degree. His enthusiasm for research and data has sparked a previously unknown joy for research in myself. Statistics provided me with a steep learning curve, however Dr. Mills has been immensely patient and supportive, guiding this project to completion. He was never too busy to listen to my problems and always had a viable solution. Dr. Mills has additionally provided opportunities for my personal learning and development beyond that which is necessary for my degree, and I cannot thank him enough. He has taught me how to disseminate and promote our research, a step often forgotten by many academics, yet communication of research outcomes is of utmost importance to the wider community. Whenever my proposal, thesis or publication crossed the desks of my supervisors, a superior product emerged from their insightful comments. It has been my privilege and pleasure to work with such outstanding supervisors.

Special thanks must be extended to Associate Professor Glenn Arendts and Dr. Eric Khong for their medical expertise, valued time and input into the clinical vignettes and feedback on our publication. This was not a simple process and their perceptiveness and observations were very much appreciated.

This research was supported by the Australian Government Research Training Program Scholarship and I am grateful of the opportunities awarded by this scholarship. We are fortunate to live in a country which promotes the pursuit of research and knowledge.

Last but not least, my children must be thanked for their unwavering support and good-hearted tolerance of my studies. These past two years have seen a massive upheaval in our family unit, yet they have cheered my modest successes and given me words of encouragement during challenging times. I am fortunate to parent four very gifted children. However, one child has been in and out of hospital with serious gastrointestinal issues, another diagnosed with multiple food allergies, and my youngest diagnosed with Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder. To say this period has been demanding is an understatement. Nonetheless, it has been incredibly rewarding as well. Family is the most important aspect of my life and I am thankful for such a wonderful 'team' at my side.

ABSTRACT

The internet has impacted society and changed the way companies and individuals operate on a daily basis. Seeking information online via computer or mobile device is common practice. The phrase 'Google it' is now part of modern vernacular and is a resource increasingly utilised by young and old alike. Around 80% of Australian's search health-related information online as it is convenient, cheap, and available 24/7. Symptom checkers are one tool used by consumers to investigate their health issues. Symptom checkers are automated online programs which use computerised algorithms, asking a series of questions to help determine a potential diagnosis and/or provide suitable triage advice. Recent evidence suggests symptom checkers may not work the way they are intended. Inferior or incorrect healthcare information can potentially have serious consequences on the consumer's wellbeing and may not have the desired effect of directing consumers to the appropriate point of care.

This research evaluated the clinical performance of 36 symptom checkers found on websites and smartphone applications that are freely available for use by the Australian general public. Symptom checkers were exposed to 48 clinical vignettes, generating 1858 symptom checker vignette tests (SCVT). Diagnosis was assessed on the inclusion of the correct diagnosis in the first, the top three or top ten differential diagnoses (n = 1,170 SCVT). Triage advice was assessed on whether the triage category recommended was concordant with our assessment (n = 688 SCVT).

The correct diagnosis was listed first in 36% (95% CI 31–42) of SCVT, within the top three in 52% (95% CI 47–59) and within the top ten in 58% (95% CI 53–65). Symptom checkers which claimed to utilise artificial intelligence (AI) outperformed non-AI with the first listed diagnosis being accurate in 46% (95% CI 40–57) versus 32% (95% CI 26–38) of SCVT. Individual symptom checker performance varied considerably, with the average rate of correct diagnosis provided first ranging between 12%–-61%. Triage advice provided was concordant with our assessment in 49% (95% CI 44–54) of SCVT. Appropriate triage advice was provided more frequently for emergency care SCVT at 63% (95% CI 52–71) than for non-urgent SCVT at 30% (95% CI 11–39).

Symptom checker performance varied considerably in relation to diagnosis. Triage advice was risk-averse, typically recommending more urgent care pathways than necessary. Given this, symptom checkers may not be working to alleviate demand for health services (particularly emergency services) within Australia—counter to marketing materials of some organisations' symptom checkers. It is important that symptom checkers do not further burden the healthcare system with inappropriate referrals or incorrect care advice. Although, a balance must be struck as avoiding unsuitable triage advice could potentially result in life-threatening consequences for consumers. Nonetheless, the results of this research make clear that the accuracy of diagnosis and triage advice provided from readily available symptom checkers for the Australian public require improvements before everyday consumers can rely entirely on health information provided via these mediums.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE	CONTENT	PAGE

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
AARP	American Association of Retired Persons
AI	Artificial intelligence
AMA	Australian Medical Association
ANOVA	Analysis of variance
App	Application
APWUHP	American Postal Workers Union Health Plan
BEACH	Bettering the Evaluation and Care of Health
CDC	Centre for Disease Control and Prevention
CHW	Children's' Hospital of Wisconsin
CI	Confidence interval
ED	Emergency department
FDA	United States Food and Drug Administration

GP	General Practitioner
HON	Health on the Net
Hopkinsallchildrens	John Hopkins Children's Hospital
NHS	National Health Service
SCVT	Symptom checker vignette test
STI	Sexually transmitted infection
TGA	Therapeutic Goods Administration
TIA	Transient Ischemic Attack
UK	United Kingdom
UofMHealth	University of Michigan Health
USA	United States of America

GLOSSARY

Artificial Intelligence (AI)

There are a multitude of definitions for AI, and a 'pure' definition is unlikely to be found in such a rapidly changing field of computer science. Broadly speaking, AI definitions are concerned with machines or computers that can engage in human-like thought processes such as learning, analysis and self-correction. For the purposes of this research, AI is taken as a computer program with some capability to imitate intelligent human behaviour (Kok, Boers, Kosters, Van der Putten, & Poel, 2009).

Acuity

In the context of this research, 'acuity' means how quickly the consumer needs to seek care. The levels of triage category this research uses are 'emergency care' being the most urgent or acute care, 'urgent care' followed by 'non-urgent care' and lastly 'self-care' which is the least urgent or acute situation. When talking about the 'level of acuity' a vignette may have, this is referring to one of these four categories and accordingly, how urgently the consumer should seek care for these scenarios. This is loosely based on the Emergency Severity Index, a classification system used in emergency departments to stratify patients from 'most urgent' to 'least urgent' based on acuity of presentation and resource needs (Gilboy, Tanabe, Travers, & Rosenau, 2012).

eHealth

The use of information and communication technologies to facilitate health and health-related activity (Cowie et al., 2016; Mars & Scott, 2010).

mHealth

According to the World Health Organization, there is no standard definition for mHealth. For the purposes of this research, mHealth is considered an aspect of eHealth which deals with mobile health products. Examples are mobile applications, health call centres, patient monitoring devices and other wireless devices supporting healthcare (Cowie et al., 2016; Kay, Santos, & Takane, 2011).

Symptom checker

A website or mobile app that is an automated online system. Symptom checkers use computerised algorithms to help formulate a medical differential diagnosis or provide triage advice (Semigran, Linder, Gidengil, & Mehrotra, 2015).

CHAPTER 1

INTRODUCTION

Thesis Map
➡ Chapter One – Introduction
Chapter Two – Literature Review
Chapter Three – Methodology
Chapter Four – Results
Chapter Five – Discussion
Chapter Six – Conclusion

1.1 The influence of the world wide web

The internet is a vast global interconnected system of networks using standardised communication protocols, which allows one computer (with permission) to 'talk' to another computer anywhere in the world. The world wide web is a collection of webpages found on these computer networks. The world wide web was invented and established initially by Sir Tim Berners-Lee as a means for scientists to share data from their research, however, this 'web' not only changed how scientists shared their knowledge, but how society operates and conducts business (BBC, 2019; Burn & Loch, 2003). Computers have led an information technology revolution allowing a global economy to exist without geographical boundaries or time limitations (Burn & Loch, 2003). For example, the banking industry has been transformed by internet banking where banks can collect, authenticate and disseminate information and services to people anywhere, anytime. The internet has also had a major impact on how individuals live on a day-to-day basis. The world wide web has become a prime source for information on any subject and internet shopping is a common occurrence (Burn & Loch, 2003). The fields of health and medicine are not immune to this internet revolution. eHealth is defined as information and communication technology which facilitates health and health-related activity including medicine (Cowie et al., 2016; Mars & Scott, 2010). There are several domains which fall under the umbrella term of eHealth including (but not limited to) telemedicine, clinical information systems and mobile health (mHealth). mHealth incorporates mobile applications (apps) and other mobile technologies that help to deliver health information or are devices used to screen patients or monitor physiological signs (Cowie et al., 2016). Medical students and junior doctors routinely utilise diagnostic apps to support their clinical practice and education (Payne, Wharrad, & Watts, 2012).

Internet access and improved connection speeds have changed how people search for information and health information. Australian internet use for health information is a growing trend, with the prevalence of those using the internet and seeking online information inversely related to age for consumers over 15 years (Wong, Harrison, Britt, & Henderson, 2014). Recent data indicates approximately 80% of Australians search health information online, with nearly 40% seeking self-treatment advice via the internet, higher than estimates in the United States of America (USA) (Cheng & Dunn, 2015; healthdirect Australia, 2016). Those who are socioeconomically disadvantaged are less likely to seek health information online (Wong et al., 2014). Cocco et al. (2018) advised that 34% of people attending two Melbourne emergency departments (EDs) had already searched the internet concerning their condition prior to presentation. Being younger and having greater health literacy increased the likelihood of using the internet to research their current health problem. Researching the condition ahead of time was felt to improve physician–patient interactions in 77% of instances for the consumer (Cocco et al., 2018). Similarly, patients presenting at a United States ED seeking urgent care were asked to share their Google search histories and medical records. From the 103 participating patients, it was found 6% of Google search queries were health related. In the seven days prior to ED presentation, the number of health-related searches rose from the baseline of 6% to 15%, and during those seven days, 56% of participants were researching symptoms, 53% were locating information pertaining to a hospital and 23% were investigating the treatment or management of a disease. In this study, 53% of participants had sought information relating to their chief complaint on Google prior to presenting at ED (Asch et al., 2019). Research has shown that consumers may experience increased apprehension following web-based searches for health information (White & Horvitz, 2009). In one study, one in five participants experienced an escalation of concerns and two in five reported increased anxiety following online interactions. Conversely, half of participants claimed to have their anxiety reduced following their online search. Individual traits and predispositions may be contributing factors to such responses (White & Horvitz, 2009). However online information influenced consumer behaviour in 40% of participants surveyed.

1.2 The rapid growth of eHealth and mobile products

Australians have embraced online technology with 86% of households having internet access and approximately 70% of the Australian population using a smartphone in 2018 (Australian Bureau of Statistics, 2018; Statista, 2018).

eHealth is a burgeoning industry with numerous health-related websites available on the world wide web. As a result, an abundance of information has become available to the general public that was formally the almost exclusive knowledge of subject specialists. Medicine is one field where the shift towards patient self-empowerment and the effortless availability of information on the internet has had a huge impact for consumers, health practitioners and the greater health system alike (Lupton & Jutel, 2015). According to the Pew Research Internet Project, 35% of Americans have gone online for health information, with 80% of inquiries starting at a search engine such as Bing, Yahoo or Google (Fox & Duggan, 2013). Google alone reports approximately one billion health questions are searched online daily—this amounts to 70,000 health related searches per minute. Google consumers are seeking information and advice regarding a range of topics, but especially pertaining to symptoms, conditions, medication, and insurance (Drees, 2019; Murphy, 2019). Furthermore, there are now over 318,000 health apps (a component of mHealth) available to smartphone consumers from leading app stores. Digital health is accelerating, with the number of health apps available practically doubling since 2015, with an additional 200 health apps being uploaded daily (Liquid State, 2018). With such a proliferation of websites and apps, determining quality and veracity of information can be difficult for the average consumer. Consumers can search Google for a wide range of medical and health concerns and follow links to websites dealing with specific health queries. These sites can contain views and opinions not necessarily supported by Google. eHealth has become such a concern for Google that it has employed the former Chief Executive of Geisinger Health, Dr. David Feinberg, in an attempt to correct their 'fake news' problem and build a stronger health foundation (Murphy, 2019). According to Rege (2019):

> Google has come under fire… for providing misinformation on some health topics, including vaccinations and certain health issues. Dr. Feinberg said the company has "teams" of physicians and nurses analyzing search results, and that Google has worked to correct the flow of misinformation (Rege, 2019).

Regulation of health information and products is discussed in greater depth later in this thesis, yet it is troubling a highly resourced and monitored search engine such as Google can still direct consumers to misleading health information. As an example of the extent of unregulated data on the internet, Wikipedia declares itself as a free encyclopedia that allows any person to edit any document—of which there is nearly six million in their electronic

library. Wikipedia is often a starting point to research health information (Beck, 2014), and in 2014, Wikipedia had 25,000 health related topics in the English directory. Collectively across all languages, these health topics received 4.8 billion pageviews in 2013. This makes Wikipedia one of the most consulted health resources worldwide (Wikipedia, 2019). Given the unique dynamic editing practices of the site, researchers have investigated Wikipedia's accuracy and reliability of information. One 2008 study determined that the online encyclopedia had an accuracy of 80% and verifiability of 90.7%, lower than Britannica's 96% accuracy and 99.6% verifiability on the same topics (Rector, 2008). A more recent study into drug information favourably compared Wikipedia to textbooks, with accuracy of drug information found to be 99.7% (± 0.2%) and the completeness of information was 83.8% (± 1.5%) when compared to textbook data. This research suggested Wikipedia was a suitable and comprehensive source for drug-related information for undergraduate students (Kräenbring et al., 2014). Due to the open and anonymous nature of editing articles in Wikipedia, malicious modifications can occur (called vandalism), however, vandalism and vulgar edits are frequently repaired within 2 to 3 minutes, such is the 'self-healing' abilities of Wikipedia (Viégas, Wattenberg, & Dave, 2004). Wikipedia has a dispute resolution protocol, which usually involves both editors discussing conflicting opinions. There is a third-party intervention process when required; options include posting the dispute on various noticeboards or requesting community opinion. 'Crowd-sourcing', or the opinion of multiple people is a defining feature of Wikipedia, and this collaborative writing process has provided more positive evaluations concerning reliability and accuracy than negative in one systemic review of 110 peer-reviewed publications on Wikipedia (Mesgari, Okoli, Mehdi, Nielsen, & Lanamäki, 2015).

1.3 Using eHealth and mHealth

Online websites and smartphone apps now allow simple mobile access to a vast array of health and medical information. Some sites merely provide healthcare information whilst others are far more specific, providing potential diagnosis and triage advice (Mueller et al., 2017). The impact of digital healthcare is ubiquitous; according to Rainie (2012), health apps are one of the most popular applications on smartphones, with recent reports suggesting 50% of smartphone owners have downloaded a health-related app (Liquid State, 2018). Research by Fiksdal et al. (2014) indicates the general public are frequently motivated to go online regarding their health for three reasons. Firstly, to seek information regarding a symptom (called 'symptom troubleshooting'); secondly to

seek information prior to visiting a health clinic; or finally to search information on behalf of someone else. The information available on the internet is regarded by consumers as quick, simple to use, anonymous and practical in terms of time and money (Fiksdal et al., 2014). Symptoms tended to be searched more than medical conditions, particularly when they were long term; considered embarrassing; were perceived as trivial; or had been presented to a healthcare provider previously (Mueller et al., 2017). An interesting phenomenon called 'cyberchondria' has occurred with the emergence of online health information. Cyberchondria occurs when an individual excessively searches the internet regarding health care information (Taylor, 1999). This practice has been found to increase negative emotions in individuals with high illness-related anxiety. The more time an individual spent searching symptoms on the internet resulted in the participant recalling higher anxiety levels and increased functional impairment during and after online searches. Conversely, those individuals with low illness-related anxiety expressed relief after searching their symptoms on the internet. It was suggested these results may adversely impact on physician workloads with increased visits from hypochondriacal patients (Doherty-Torstrick, Walton, & Fallon, 2016).

Both online health websites and health call centres enable a consumer to gather information, become educated about healthcare and gain advice on potential treatment options from the comfort of their own home, with minimal cost involved (Lupton & Jutel, 2015). Health call centres permit the general public to become more proactive with their health concerns and 'shop around' for advice on medical issues or guidance on where to seek assistance (North, Varkey, Laing, Cha, & Tulledge-Scheitel, 2011). Frequently, call centres are created to reduce demand on the overburdened healthcare system, specifically designed to lessen the workload of EDs and direct or 'triage' healthcare traffic to a more appropriate healthcare provider (Armstrong, 2018; Baum & News, 2017). Many of these health call centre companies also invest in an online equivalent, eliminating the need for human interaction. This potentially expands the customer base for the organisation whilst allowing the consumer flexibility in their search for health information. These online applications which diagnose and/or triage a set of symptoms are commonly referred to as 'symptom checkers' and are readily available on the internet or come as a downloadable software product app for use on smartphones (Kratzke & Cox, 2012). Symptom checkers have been touted as a feasible option to direct consumers to an appropriate health service, thereby reducing the demand on primary and emergency healthcare organisations (Armstrong, 2018; NHS England, 2017). Generally, symptom checker

websites and apps are free to the consumer, although some are only available to paying members (e.g. the MyClinic app which costs AU$2.99 to download (MyClinic Healthsite Apps, 2019)) while symptom checkers utilised by physicians tend to be expensive applications and touted as higher quality products (Boruff & Storie, 2014; Buijink, Visser, & Marshall, 2013; Payne et al., 2012). From a business perspective, symptom checkers typically cost less to operate than health call centres and are viewed as a viable (if not preferable) alternative (Skowronski, Kaanapu, Oakkar, & Mostafa, 2014).

1.4 About symptom checkers

Symptom checkers are automated online systems which use computerised algorithms to ask a series of questions to help determine a differential medical diagnosis and provide appropriate triage advice. Some sites avoid providing a diagnosis but offer suggestions regarding triage from the symptoms described (Middleton et al., 2016; Semigran et al., 2015). Generally, differential diagnoses are listed in order of likelihood, whilst triage advice recommends where care should be sought (e.g. ED vs. medical clinic vs. self-care) and the degree of urgency of action for the potential condition. For example, self-care may be suggested for a simple headache, but this may be upgraded to emergency care if the headache has other associated symptoms such as photophobia or a stiff neck (Sanders, Lewis, & Quick, 2012, p. 868). The algorithmic logic used for symptom checkers is frequently based on those utilised in health call centres, such as the Schmitt and Thomson telephone nursing triage guidelines (Armstrong, 2018; Gardiner, Cullen, Karabatsos, & St George, 2008; Semigran et al., 2015).

Many symptom checkers are in the first generation for such programs, but recent innovations have changed the algorithmic logic used in some symptom checkers, with improved questioning techniques to fine-tune diagnostic accuracy (Baum & News, 2017; Semigran et al., 2015). Some companies such as Buoy Health have commenced using artificial intelligence (AI) algorithms—where the symptom checker program adds information to its database with each transaction which is meant to improve the diagnostic accuracy of the system over time (Baum & News, 2017; Buoy Health, 2018). Isabel Healthcare advertise they use 'multiple AI and enhanced machine learning algorithms' designed to handle complex medical presentations (Isabel Healthcare, 2018). Joshi (2019) describes AI as having four commonly used classifications:

- reactive, where the program simply responds to a stimulus
- limited memory, where the program uses historical data and memory to learn and improve its responses
- theory of mind, where AI would understand the needs of other intelligent beings and
- self-aware, where AI would be similar to a human in intelligence and self-awareness.

The theory of mind classification is currently under development, while the self-aware AI is simply a theoretical concept at this point in time. So contemporary use of the term AI would either refer to a reactive or limited memory program (Joshi, 2019; Reynoso, 2019).

Symptom checkers can be stand-alone programs or be embedded within a larger website. These might provide a library of information pertaining to symptoms and a variety of diseases and conditions, and can provide treatment options, disease management information, and guide the consumer on what to do next. For example, WebMD's website includes information pertaining to 'Drugs & Supplements', 'Family & Pregnancy' and 'Health A–Z' as well as having a symptom checker and 'Find a Dentist' and 'Find a Doctor' function (WebMD, 2017). Healthdirect Australia's home page not only has the symptom checker link but provides links to 'Health topics A–Z', 'Medicines', 'Service Finder and 'Pregnancy and parenting'. This practice provides a comprehensive solution to finding healthcare information. Healthdirect Australia is a not-for-profit organisation established by the Council of Australian Governments in 2006 and delivers a national service providing 24/7 health and wellbeing advice and information through multiple avenues (healthdirect Australia, 2017; Ng et al., 2012). Healthdirect Australia (2016) does not provide a differential diagnosis, as it is solely a triage symptom checker. They operate the only symptom checker designed specifically for the Australian general public, funded by the Australian Federal Government, Australian Capital Territory, Northern Territory, and the State Governments of New South Wales, South Australia, Tasmania and Western Australia, (hereafter referred to as 'Government funded') and is therefore used within this research as the 'gold standard' for other triage symptom checkers.

1.5 Potential advantages of symptom checkers

There are multiple advantages to a well-constructed symptom checker. The consumer can become more educated and empowered about their own health, which in turn can aid patient-physician relationships (Lupton & Jutel, 2015).

Skowronski et al. (2014) advised symptom checkers can be linked to voluminous online libraries providing extensive up-to-date evidence-based health information. An interesting concept is the possibility of using symptom checkers in countries where access to healthcare services are limited. Morita, Rahman, Hasegawa, Ozaki, & Tanimoto (2017) suggest symptom checkers may assist with triaging consumers in a country such as Bangladesh, where the physician-to-population ratio is 1:12,690 and the physician-to-nurse ratio is 2.5:1. Whilst online symptom checkers are not yet as accurate as a physician's diagnosis, over time such databases can accumulate vast amounts of information to improve clinical diagnosis and may become a useful tool to assist clinical practice in such resource poor nations (Morita et al., 2017). Furthermore, consumers in well-resourced countries can save time by obtaining pertinent information online, and potentially save money by avoiding an unnecessary visit to a healthcare service (Mueller et al., 2017) which in turn alleviates the strain on the public health system and improves health outcomes for the community as a whole (Skowronski et al., 2014). Another advantage of symptom checkers is that in an emergency situation the triage advice provided could encourage the patient to seek life-saving assistance (Semigran et al., 2015). Of particular interest, online websites have a higher utilisation rate than triage call centres do (Skowronski et al., 2014), which suggests consumers prefer online health information sources. Computers and apps are a convenient 24/7 resource which can be referred to at any point in time allowing consumers to not incur wasted time in a call centre queue. Nevertheless, not all websites and apps are created equally, with quality evidence-based information difficult to differentiate from inferior information for the layperson.

1.6 Potential disadvantages of symptom checkers

Both healthcare organisations and private companies operate symptom checkers (Lupton & Jutel, 2015; Semigran et al., 2015), but not all symptom checkers are affiliated with 'reputable' organisations. This has raised concerns regarding the regulation of these sites. Data security, privacy, quality and accuracy of health information are a few of the key concerns raised by the health industry, with many developers supplying limited information pertaining to the credentials of the authors or editors (Boudreaux et al., 2014; Jutel & Lupton, 2015).

The Australian Therapeutic Goods Administration (TGA) oversees the regulation of medical software products and medical devices in Australia. Many mobile apps do not, however, fall within the definition of a medical device.

According to the Therapeutic Goods Act 1989, section 41BD, a medical device is "any instrument, apparatus, appliance, material or other article … to be used by human beings for the purpose of one or more of the following: 1. Diagnosis, prevention, monitoring, treatment or alleviation of disease; …" (Australian Government, 1989, Medical devices, para. 5). Software that fulfils this description may include smartphone apps that diagnose cardiac dysrhythmias or devices that monitor blood glucose levels for example. However, a mobile app which simply provides information (as opposed to definitive diagnosis for example) falls outside the scope of the TGA (Department of Health and Ageing, 2011, 2013). Most symptom checkers place caveats on their website or app stating that caution should be used by the layperson when attempting self-diagnosis and that the app/website should not replace the services of a physician (Jutel & Lupton, 2015). Some sites, such as Healthdirect Australia, explicitly state their symptom checker is not a diagnostic tool, nor a substitute for seeking professional health care (Healthdirect Australia, 2016). Lack of regulation raises concerns about the validity of information provided on unregulated sites with little to no quality control (Boudreaux et al., 2014). In early 2019 the United States Food and Drug Administration (FDA) launched a voluntary software precertification program aimed at providing assurance of effectiveness, transparency of product performance and safety (US Food and Drug Administration, 2019). The FDA evaluates organisational excellence in relation to five principles; product quality, patient safety, clinical responsibility, cybersecurity responsibility and a proactive culture. The key question is whether voluntary precertification is going far enough on cyber-misinformation (Lim & Shorey, 2019), and whether there should be a more unified approach globally (Mars & Scott, 2010).

Further disadvantages of symptom checkers are that the information provided is not reviewed by a physician unless the consumer chooses to disclose this information to their general practitioner (GP). Additionally, the information provided may convince the patient they have a concerning illness which could send them to their GP with heightened anxiety (Mueller et al., 2017; Skowronski et al., 2014). Research conducted by Semigran et al. (2015) found some symptom checkers such as iTriage always triaged consumers to the emergency department regardless of their symptoms, which does nothing to alleviate the burden on emergency services and provides no value to the consumer. Conversely, if a symptom checker provides incorrect under-triage advice, this can delay definitive treatment and diagnosis with potentially serious consequences (Boudreaux et al., 2014). Another limitation of symptom checkers is the validity of the diagnosis or triage advice can be severely affected by the data

input by the consumer into the program, and many users experience difficulty with the degree of health literacy and computer literacy required to navigate the system (Boudreaux et al., 2014; Luger, Houston, & Suls, 2014; Ninh, 2014). There is a strong association between poor health literacy and the inability to accurately evaluate online health information (Diviani, van den Putte, Giani, & van Weert, 2015). South Australia Health recommend the reading age regarding health information be targeted at an Eighth-Grade level. However, Cheng and Dunn (2015) found that health information is typically two to four grades higher than Grade Eight level. This has potentially serious implications for how consumers understand and manage their health concerns. The extent that diagnostic and triage advice given from symptom checkers is accurate, relevant and appropriate for the Australian community is presently unclear.

1.7 Research aims

Purpose

The aim of this study was to ascertain the diagnostic and triage clinical performance of free online symptom checkers (both websites and smartphone apps) most frequently available to the Australian public. As Australians can easily access the worldwide web, the symptom checkers appearing foremost in five popular search engines or two online application stores were assessed to determine their diagnostic and triage accuracy. Extending previous research investigating the same in the USA conducted by Semigran et al. (2015), 36 different websites and smartphone symptom checkers were evaluated (17 re-reviewed, 19 newly appraised), utilising layperson's terminology when inputting signs and symptoms. Given Healthdirect Australia is the Australian gold standard, as it has been explicitly designed for the unique Australian conditions and healthcare system, Healthdirect Australia's performance was then compared to the accuracy of triage advice given from international symptom checker sites.

Research Questions

1) Are free online English language symptom checkers available to the Australian general public clinically accurate? Is the correct diagnosis listed as the first diagnosis, within the top three diagnoses, or within the top ten diagnoses?

2) Are free online English language symptom checkers which are currently available to the Australian general public providing accurate triage advice?
3) How do international English websites and apps most readily available to the Australian public perform in comparison to the Australian Government funded Healthdirect Australia symptom checker?

1.8 Significance of research

The trend toward self-diagnosis is accelerating with modern technology and easy access to information via the internet (Robertson, Polonsky, & McQuilken, 2014). The consumer's desire to become more educated and proactive in their own healthcare must be tempered by the potential for misinformation from internet 'experts' and the real possibility for self-misdiagnosis. One case reported in New Zealand had the consumer so convinced of his own diagnosis of renal calculi that his GP's diagnostic process was detrimentally influenced. Two weeks following his GP appointment the gentleman presented to ED with a perforated appendix (Avery, Ghandi, & Keating, 2012). This case demonstrates that the traditional model of seeking information from the medical professional first may no longer hold true, and that consumers now often present to their GPs with preconceived ideas regarding their symptoms which can influence the information provided to physicians, making the diagnostic process of analysis and interpretation more challenging than it was previously (Lupton & Jutel, 2015). The context that many physicians trained in has changed and a new paradigm is emerging, which can either be aided or hindered by modern technology. The information sourced from the internet has a direct impact on the relationship between a patient and their physician, potentially shifting the locus of authority and redistributing the power in the relationship (Jutel & Lupton, 2015).

Many Australians research their chief complaint online prior to presentation, and e-literate and younger patients were more inclined to research health information online (Asch et al., 2019; Cocco et al., 2018). With four out of five Australian's turning to the internet for health information, this has serious implications for where and when Australians seek care (Cheng & Dunn, 2015; Healthdirect Australia, 2016). With an already overburdened healthcare system and EDs frequently running at capacity, it is vital tools such as symptom checkers direct

consumers to an appropriate point of care in a suitable timeframe, and information is applicable to the Australian community.

1.9 Structure of the thesis

Chapter one of this thesis introduces the concept of eHealth and current health-information seeking behaviours. It explains exactly what symptom checkers are and what they are designed to do. The positive and negative aspects to these programs are also explored. The structure of the thesis is explained, and a thesis map is provided to help guide the reader.

Chapter two presents a thorough review of the literature relevant to this research. Previous research into the performance of symptom checkers, current trends in the use of symptom checkers, and the health system response to community need are reviewed.

Chapter three explains the conceptual framework for this thesis and methodology employed. How appraised symptom checkers were located online and selected for evaluation, the clinical vignettes used for appraisal, and how triage categories were defined are all described in this section. The methods used to enter data into the symptom checkers are justified, as it was important there was consistency in the provision of information between programs evaluated. The planned analysis for diagnostic and triage accuracy is elucidated.

Chapter four provides the outcomes from this research. Results are presented for diagnostic accuracy and for triage advice overall, as well as for individual symptom checkers and are compared to the performance of Healthdirect Australia's symptom checker.

Chapter five is devoted to the discussion of these outcomes and explores implications from these findings. Issues faced with using the symptom checker interfaces are discussed, and local and international eHealth policy and regulation are reviewed. Recommendations are made for regulation of online health information including symptom

checkers, as well as ensuring products have consumer input prior to release on the market. Alerts for concerning symptoms and transparency of how symptom checkers are developed and informed are also recommended.

Chapter six reviews the research questions and explains how these have been addressed by this investigation, and provides conclusions based from the findings in chapter four and discussion section.

Thesis Map

➡ Chapter One – Introduction
Chapter Two – Literature Review
Chapter Three – Methodology
Chapter Four – Results
Chapter Five – Discussion
Chapter Six – Conclusion

A 'Thesis Map' (above) has been provided to guide the reader through this research thesis. It appears at the beginning of each chapter, with the specific chapter indicated by an arrow. The methodology and discussion chapters have additional sub-headings at the beginning of those chapters, as these sections are considerably longer and require some additional guidance. References have been provided at the end of each chapter, for ease of reference.

1.10 Summary

The internet has changed the face of society, and how information is sourced. Along with many other industries, medicine and health have been affected by this information technology boom, with millions of health-related searches occurring on Google alone each day. eHealth products are a thriving industry, with a multitude of new websites and apps being uploaded to the internet daily. Symptom checkers are one tool individuals are using online (via computer or mobile app) to seek assistance with their symptoms, searching for diagnostic and triage advice. Previous research conducted overseas indicates these tools may not work as well as intended. This study seeks to evaluate the clinical accuracy and appropriateness of triage advice, for free symptom checkers readily available to the Australian general public.

1.11 References

Armstrong, S. (2018). The apps attempting to transfer NHS 111 online. *BMJ (clinical research ed.), 360,* 156. doi:10.1136/bmj.k156. Retrieved from http://www.bmj.com/content/360/bmj.k156.abstract

Asch, J. M., Asch, D. A., Klinger, E. V, Marks, J., Sadek, N., & Merchant, R. M. (2019). Google search histories of patients presenting to an emergency department: an observational study. *BMJ Open, 9*(2), e024791. doi:10.1136/bmjopen-2018-024791

Australian Bureau of Statistics. (2018). *8146.0 - Household use of information technology, Australia, 2014-2015.* Retrieved April 14, 2018, from http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0#

Australian Government. (1989). *Therapeutic Goods Act 1989.* Retrieved May 2, 2018, from https://www.legislation.gov.au/Details/C2017C00226

Avery, N., Ghandi, J., & Keating, J. (2012). The 'Dr Google'phenomenon–missed appendicitis. *NZ Med J, 125*(1367), 135–137.

Baum, S., & News, M. (2017*). Buoy health raises $6.7M for smarter triage tool aimed at hospitals and payers.* Retrieved April 14, 2018, from https://medcitynews.com/2017/08/buoy-health-raises-6-7m-smarter-triage-tool-aimed-hospitals-payers/

BBC. (2019). *What is the world wide web?* Retrieved December 26, 2019, from https://www.bbc.co.uk/bitesize/topics/zkcqn39/articles/z2nbgk7

Beck, J. (2014). *Doctors' #1 Source for Healthcare Information: Wikipedia*. Retrieved November 24, 2019, from https://www.theatlantic.com/health/archive/2014/03/doctors-1-source-for-healthcare-information-wikipedia/284206/

Boruff, J. T., & Storie, D. (2014). Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information. *Journal of the Medical Library Association: JMLA, 102*(1), 22–30. doi:10.3163/1536-5050.102.1.006

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014). Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine: Practice, Policy, Research, 4*(4), 363–371. https://doi.org/10.1007/s13142-014-0293-9

Buijink, A. W. G., Visser, B. J., & Marshall, L. (2013). Medical apps for smartphones: lack of evidence undermines quality and safety. *BMJ Evidence-Based Medicine, 18*(3), 90–92.

Buoy Health. (2018). *Take the guesswork out of health care*. Retrieved November 3, 2018, from https://www.buoyhealth.com

Burn, J. M., & Loch, K. D. (2003). *The societal impact of the World Wide Web-Key challenges for the 21st century.* Advanced Topics in Information Resources Management, Volume 2 (pp. 32–51). IGI Global.

Cheng, C., & Dunn, M. (2015). Health literacy and the Internet: a study on the readability of Australian online health information. *Australian and New Zealand Journal of Public Health, 39*(4), 309–314. https://doi.org/10.1111/1753-6405.12341

Cocco, A. M., Zordan, R., Taylor, D. M., Weiland, T. J., Dilley, S. J., Kant, J., … Hutton, J. (2018). Dr Google in the ED: searching for online health information by adult emergency department patients. *Medical Journal of Australia, 209*(8), 342–347.

Cowie, M. R., Bax, J., Bruining, N., Cleland, J. G. F., Koehler, F., Malik, M., … Vardas, P. (2016). e-Health: a position statement of the European Society of Cardiology. *European Heart Journal, 37*(1), 63–66. doi:10.1093/eurheartj/ehv416

Department of Health and Ageing. (2011). *Australian regulatory guidelines for medical devices (ARGMD).* Retrieved April 13, 2018, from https://www.tga.gov.au/sites/default/files/devices-argmd-01.pdf

Department of Health and Ageing. (2013). *Regulation of medical software and mobile medical 'apps'.* Retrieved April 13, 2018, from https://www.tga.gov.au/regulation-medical-software-and-mobile-medical-apps

Diviani, N., van den Putte, B., Giani, S., & van Weert, J. C. M. (2015). Low Health Literacy and Evaluation of Online Health Information: A Systematic Review of the Literature. *Journal of Medical Internet Research, 17*(5), e112. https://doi.org/10.2196/jmir.4018

Doherty-Torstrick, E. R., Walton, K. E., & Fallon, B. A. (2016). Cyberchondria: parsing health anxiety from online behavior. *Psychosomatics, 57*(4), 390–400. doi: 10.1016/j.psym.2016.02.002

Drees, J. (2019). *Google receives more than 1 billion health questions every day.* Retrieved September 9, 2019, from https://www.beckershospitalreview.com/healthcare-information-technology/google-receives-more-than-1-billion-health-questions-every-day.html

Fiksdal, A. S., Kumbamu, A., Jadhav, A. S., Cocos, C., Nelsen, L. A., Pathak, J., & McCormick, J. B. (2014). Evaluating the Process of Online Health Information Searching: A Qualitative Approach to Exploring Consumer Perspectives. *Journal of Medical Internet Research, 16*(10), e224. https://doi.org/10.2196/jmir.3341

Fox, S., & Duggan, M. (2013). Health Online. Pew Research Internet Project. *Health, 2013, 1*-55. Retrieved April 10, 2018, from http://www.pewinternet.org/2013/01/15/health-online-2013/

Gardiner, L., Cullen, M., Karabatsos, G., & St George, I. (2008). Universal telenursing triage in Australia and New Zealand: A new primary health service in Australia and New Zealand. *Australian Family Physician, 37*(6), 476-479.

Gilboy, N., Tanabe, P., Travers, D. A., & Rosenau, A. M. (2012). *Emergency Severity Index (ESI): A Traige Tool for Emergency Department Care, Version 4.* Implementation Handbook 2012 Edition. AHRQ Publication, (No. 12-0014). Retrieved from https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/esi/esihandbk.pdf

Healthdirect Australia. (2016). *About Healthdirect symptom checker.* Retrieved April 14, 2018, from https://www.healthdirect.gov.au/about-healthdirect-symptom-checker

Healthdirect Australia. (2017). *Annual Report 2016 – 2017 Healthdirect Australia.* Retrieved April 14, 2018, from https://media.healthdirect.org.au/publications/Annual_Reports_Business_Highlights_2016-2017.pdf

Isabel Healthcare. (2018). *Isabel.* Retrieved November 1, 2018, from https://symptomchecker.isabelhealthcare.com/the-symptom-checker/how-it-works

Joshi, N. (2019). *7 Types Of Artificial Intelligence.* Retrieved September 16, 2019, from https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/#a54cea7233ee

Jutel, A., & Lupton, D. (2015). Digitizing diagnosis: a review of mobile applications in the diagnostic process. *Diagnosis 2*(2), 89–96. https://doi.org/10.1515/dx-2014-0068

Kay, M., Santos, J., & Takane, M. (2011). mHealth: New horizons for health through mobile technologies. *World Health Organization, 64*(7), 66–71.

Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: definition, trends, techniques, and cases. *Artificial Intelligence, 1.*

Kräenbring, J., Penza, T. M., Gutmann, J., Muehlich, S., Zolk, O., Wojnowski, L., … Sarikas, A. (2014). Accuracy and completeness of drug information in Wikipedia: a comparison with standard textbooks of pharmacology. *PloS One, 9*(9), 106930. doi:10.1371/journal.pone.0106930

Kratzke, C., & Cox, C. (2012). Smartphone Technology and Apps: Rapidly Changing Health Promotion. Global *Journal of Health Education and Promotion, 15*(1).

Lim, N. L. Y., & Shorey, S. (2019). Effectiveness of technology-based educational interventions on the empowerment related outcomes of children and young adults with cancer: A quantitative systematic review. *Journal of Advanced Nursing, 75*(10), 2072–2084. doi:10.1111/jan.13974

Liquid State. (2018). *4 Digital Health App Trends to Consider for 2018*. Retrieved September 8, 2019, from https://liquid-state.com/digital-health-app-trends-2018/

Luger, T. M., Houston, T. K., & Suls, J. (2014). Older Adult Experience of Online Diagnosis: Results From a Scenario-Based Think-Aloud Protocol. *Journal of Medical Internet Research, 16*(1), e16. https://doi.org/10.2196/jmir.2924

Lupton, D., & Jutel, A. (2015). 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Social Science & Medicine, 133*, 128–135. https://doi.org/https://doi.org/10.1016/j.socscimed.2015.04.004

Mars, M., & Scott, R. E. (2010). Global E-Health Policy: A Work In Progress. *Health Affairs, 29*(2), 237-243. doi:10.1377/hlthaff.2009.0945

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology, 66*(2), 219–245.

Middleton, K., Butt, M., Hammerla, N., Hamblin, S., Mehta, K., & Parsa, A. (2016). Sorting out symptoms: design and evaluation of the "babylon check" automated triage system. *arXiv preprint arXiv:1606.02041.* Retrieved from http://arxiv.org/abs/1606.02041

Morita, T., Rahman, A., Hasegawa, T., Ozaki, A., & Tanimoto, T. (2017). The potential possibility of symptom checker. *International Journal of Health Policy and Management, 6*(10), 615–616. doi: 10.15171/ijhpm.2017.41

Mueller, J., Jay, C., Harper, S., Davies, A., Vega, J., & Todd, C. (2017). Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review. *Journal of Medical Internet Research, 19*(6), e202. https://doi.org/10.2196/jmir.6755

Murphy, M. (2019). *Dr Google will see you now: Search giant wants to cash in on your medical queries*. Retrieved September 9, 2019, from https://www.telegraph.co.uk/technology/2019/03/10/google-sifting-one-billion-health-questions-day/

MyClinic Healthsite Apps. (2019). *MyClinic Symptom Checker.* (Version 1.0) [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=base.com.healthsite.myclinic&hl=en

Ng, J. Y., Fatovich, D. M., Turner, V. F., Wurmel, J. A., Skevington, S. A., & Phillips, M. R. (2012). Appropriateness of healthdirect referrals to the emergency department compared with self-referrals and GP referrals. *Medical Journal of Australia, 197*(9), 498–502. doi: 10.5694/mja12.10689

NHS England. (2017). *Next steps on the NHS five year forward view.* Retrieved April 10, 2018, from https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf

Ninh, A. Q. (2014). DocBot: A novel clinical decision support algorithm. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6290–6293). https://doi.org/10.1109/EMBC.2014.6945067

North, F., Varkey, P., Laing, B., Cha, S. S., & Tulledge-Scheitel, S. (2011). Are e-health Web users looking for different symptom information than callers to triage centers? *Telemedicine and E-Health, 17*(1), 19–24. doi:10.1089/tmj.2010.0120

Payne, K. F. B., Wharrad, H., & Watts, K. (2012). Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. *BMC Medical Informatics and Decision Making, 12*, 121. https://doi.org/http://dx.doi.org/10.1186/1472-6947-12-121

Rainie, L. (2012). The rise of the e-patient: understanding social networks and online health information seeking. Grand Rounds Lecture. *Burbank (CA): Providence St. Joseph Medical Center*. Retrieved from http://www. https://www.pewresearch.org/internet/category/presentations/page/17/

Rector, L. H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review, 36*(1), 7–22. https://doi.org/10.1108/00907320810851998 https://ecu.on.worldcat.org/oclc/5347391922

Rege, A. (2019). *Dr. David Feinberg: Google's plan to fix fake news in healthcare.* Retrieved September 10, 2019, from https://www.beckershospitalreview.com/hospital-management-administration/dr-david-feinberg-google-s-plan-to-fix-fake-news-in-healthcare.html

Reynoso, R. (2019). *4 main types of Artificial Intelligence*. Retrieved September 10, 2019, from https://learn.g2.com/types-of-artificial-intelligence

Robertson, N., Polonsky, M., & McQuilken, L. (2014). Are my symptoms serious Dr Google? A resource-based typology of value co-destruction in online self-diagnosis. *Australasian Marketing Journal (AMJ), 22*(3), 246–256.

Sanders, M. J., Lewis, L. M., & Quick, G. (2012). *Mosby's paramedic textbook* (4th ed.) St. Louis, Mo.: Elsevier/Mosby Jems.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ (clinical research ed.), 351,* 3480. doi:10.1136/bmj.h3480

Skowronski, J., Kaanapu, J., Oakkar, O., & Mostafa, J. (2014, January 16). Methods, systems, and devices for online triage. *U.S. Patent Application No. 13/940,079.*

Statista. (2018). *Number of smartphone users in Australia from 2015 to 2022 (in millions).* Retrieved April 14, 2018, from https://www.statista.com/statistics/467753/forecast-of-smartphone-users-in-australia/

Taylor, H. (1999). Explosive growth of a new breed of cyberchondriacs. Harris Poll, 11.

US Food and Drug Administration. (2019). Developing a software precertification program: A working model. US Department of Health and Human Services Food and Drug Administration. Retrieved May 14, 2019, from

https://www.fda.gov/media/119722/download

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 575–582).

WebMD. (2017). WebMD Symptom Checker. Retrieved November 14, 2018, from https://symptoms.webmd.com/default.htm

White, R. W., & Horvitz, E. (2009). Experiences with web search on medical concerns and self diagnosis. In AMIA annual symposium proceedings (Vol. 2009, pp. 696). American Medical Informatics Association.

Wikipedia. (2019). Health information on Wikipedia. Retrieved November 27, 2019, Retrieved from https://en.wikipedia.org/wiki/Health_information_on_Wikipedia

Wong, C., Harrison, C., Britt, H., & Henderson, J. (2014). Patient use of the internet for health information. Australian Family Physician, 43(12), 875–877. Retrieved from http://www.racgp.org.au/afp/2014/december/patient-use-of-the-internet-for-health-information/

CHAPTER 2

REVIEW OF THE LITERATURE

Thesis Map
Chapter One – Introduction
➡ Chapter Two – Literature Review
Chapter Three – Methodology
Chapter Four – Results
Chapter Five – Discussion
Chapter Six – Conclusion

2.1 Methodology for the literature review

The purpose of this literature review was to determine what research had been done pertaining to symptom checkers that were freely available to the general public and were designed to 'symptom check' a wide variety of medical conditions (not for one specific area such as 'mental health' or 'dental health'). This was not a systematic review of the literature, rather a search for research articles relating to this research topic. An extensive search was undertaken in May 2018 to locate articles meeting this criterion. While other articles have been included for discussion and comparison purposes, only 16 articles regarding symptom checker research were identified for this literature review. Figure 2.1 below shows the databases reviewed and the search words used in this process.

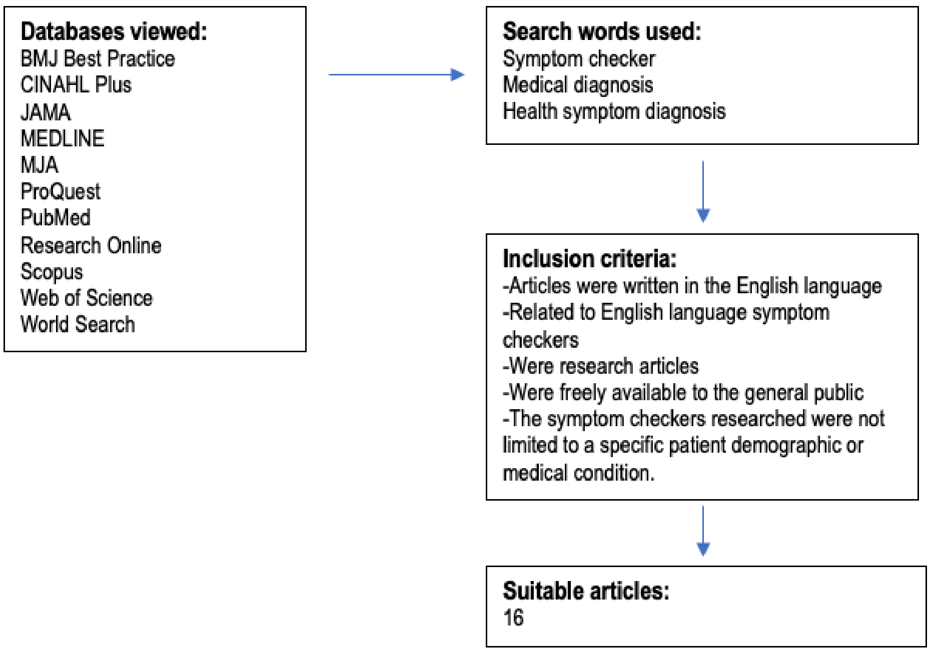


Figure 2.1: Methodology for searching articles regarding symptom checkers

2.2 Using Google to search symptoms

Tang and Ng (2006) sought to determine how accurate Google would be in determining a patient's correct diagnosis if physicians were to enter symptoms of specific medical conditions into the system. The researchers used 26 diagnostic cases published in the New England Journal of Medicine but withheld the actual diagnosis from the physician's evaluating Google. Therefore, the physicians had the case information and symptoms but were blind to the diagnosis. The physicians were asked to enter between three and five search terms into Google, and whenever possible used more unique search phrases such as 'cardiac arrest sleep'. Google located the correct diagnosis in 58% of instances. It was conjectured that Google would probably be more effective in locating and diagnosing conditions with unique symptoms, and less effective for complex conditions with non-specific symptoms or common illnesses with unusual presentations. Tang and Ng (2006) suspected that 'human experts' such as doctors would receive more accurate search results than the general public, as they would have greater experience at selecting the relevant documents [i.e. the knowledge to select the relevant links on a webpage to appropriate or suitable sites]. They determined Google may be a useful tool for physicians to use to compile a differential diagnosis list for more complex cases.

Also reviewing Google, Chen and Turner (2010) found there are serious limitations for generalised web-based searches regarding self-diagnosis for symptoms relating to amyotrophic lateral sclerosis or motor neuron disease. Some symptoms (for instance, 'weak grip' or 'tripping') incurred over 400 search results prior to the mention of amyotrophic lateral sclerosis and/or motor neuron disease as a possible diagnosis. They determined the diagnostic sensitivity and specificity were poor for these medical conditions. Information provided regarding prognostication was variable and possible curative measures were potentially misleading. Overall, the researchers suggest Google does not appear to be a reliable diagnostic source for the layperson and suggest doctors guide consumers to reliable sources of online information. It could be expected then, for specialised programs specifically designed to diagnose and/or triage symptoms to perform with greater accuracy.

2.3 Symptom checkers designed for healthcare professionals

A systematic review of symptom checkers used by physicians found differential diagnosis generators pooled accuracy rate was ~70%, however the performance of these tools fell as cases became more complex. Many generators gave extensive differential diagnoses lists. Consequently, while the 'correct' answer would frequently be included in the list, the lists comprehensiveness (as opposed to the provision of a single probable diagnosis) served to provide less value to physicians. Research has shown users may miss a correct diagnosis in a differential diagnosis list, with this effect being more profound in longer lists (Riches et al., 2016). Diagnostic symptom checkers used by physicians usually involve a fee-for-purchase (Payne, Wharrad, & Watts, 2012) and are more expensive than the layperson's counterpart, with more specialised and detailed medical information. These products are targeted to medical professionals but are still available for the general public to purchase (Jutel & Lupton, 2015). Such programs support medical practitioners with drug information, diagnosis and disease management, and can be referred to several times a day; espoused as being especially useful for students and junior doctors (Payne et al., 2012). The accuracy and reliability of these products is not necessarily guaranteed, and there is call for peer-review and regulation of these tools to ensure patient safety in clinical practice (Buijink, Visser, & Marshall, 2013). Many studies discuss the need for regulation of online medical information, including symptom checkers for both health professionals and lay people, as clinical effectiveness and safety for these products have not been determined (Buijink et al., 2013; Lupton & Jutel, 2015; Millenson, Baldwin, Zipperer, & Singh, 2018; Semigran, Linder, Gidengil, & Mehrotra, 2015; Verzantvoort, Teunis, Verheij, & van der Velden, 2018).

2.4 Symptom checkers designed for the layperson

Research pertaining to symptom checkers designed for the general public is sparse. Chambers et al.'s (2019) recent systematic review identified only eight research articles relating to the general public, and not targeting a specific medical condition or demographic group. Of these eight articles, two were reviewing symptom checkers designed in the Netherlands (Nijland, Cranen, Boer, van Gemert-Pijnen, & Seydel, 2010) or Norway (Marco-Ruiz et al., 2017) and therefore would not appear in an online search as symptom checkers available for the Australian population. The remaining six articles (Middleton et al., 2016; Poote, French, Dale, & Powell, 2014; Razzaki et al.,

2018; Semigran, Levine, Nundy, & Mehrotra, 2016; Semigran et al., 2015; Sole, Stuart, & Deichen, 2006) are discussed below as well as being summarised with their findings. A further five articles reviewed by Chambers et al. (2019) were for symptom checker programs embedded within other electronic consultation systems such as participating medical centres or the NHS pathways system (Carter, Fletcher, Sansom, Warren, & Campbell, 2018; Cowie, Calveley, Bowers, & Bowers, 2018; Madan, 2014; NHS England, 2017; Nijland, van Gemert-Pijnen, Boer, Steehouder, & Seydel, 2009). Again, these programs would not be available to the Australian general public as they are region-specific.

Chambers et al.'s (2019) review determined that symptom checkers used for urgent healthcare problems are overly cautious with disposition advice and are generally inferior to a physician's assessment diagnostically. Evidence for patient safety was weak, and there was limited evidence of patient conformity with triage advice. It was found younger and more highly educated people were more likely to use such online health-care tools which may have implications for inequity in healthcare (Chambers et al., 2019). Whilst the internet is an easy and accessible resource for health information, there seems to be a growing divide between community members who access these online services. In a United States investigation, being older, of a lower socioeconomic background and of Hispanic origin were determinants for using a healthcare professional or traditional media source to seek out healthcare information—therefore less likely to explore the internet for health-related information (Jacobs, Amuta, & Jeon, 2017). A 2019 scoping review of the literature pertaining to AI 'self-diagnosing platforms' determined that diagnostic accuracy differed substantially depending on the condition searched and platform used by consumers. Women and consumers with higher education qualifications were more inclined to select the correct diagnosis from the differential diagnosis list, and those with potentially embarrassing conditions or without healthcare were more likely to utilise such technology (Aboueid, Liu, Desta, Chaurasia, & Ebrahim, 2019).

WebMD's diagnostic symptom checker, and the National Health Service's (NHS's) triage symptom checker in the United Kingdom (UK) were assessed pertaining to inflammatory arthritis; with only 19% of patients receiving a first diagnosis of rheumatoid or psoriatic arthritis via WebMD. The triage advice provided by the NHS's symptom checker was frequently inappropriate with 44% of cases suggesting unwarranted referral to the ED or for ambulance assistance (Powley, McIlroy, Simons, & Raza, 2016). WebMD was again evaluated on its ability to

diagnose ophthalmic conditions (Shen, Nguyen, Gregor, Isaza, & Beattie, 2019). Assessing 42 ophthalmic vignettes, 26% of vignettes had the correct diagnosis listed first, and 38% within the top three options. Nevertheless, 43% of vignettes did not have the correct diagnosis listed at all, and triage advice was determined as appropriate in merely 39% of scenarios.

A symptom checker versus physician comparison was performed retrospectively on a complex patient group using ED data (Berry et al., 2019). HIV and Hepatitis C patient data was run through five popular symptom checkers; Mayo Clinic, WebMD, Symptomate, Symcat and Isabel Healthcare, and the diagnostic and triage advice given by the symptom checkers was compared to the ED physician's assessment (considered the gold standard). Less than 20% of cases had the correct diagnoses listed first by the symptom checkers, with less than 45% listing the correct diagnosis at all. Symptom checker triage advice for these select cases should have always been 'emergency', however only 37% of the HIV cases, 60% of the Hepatitis C cases and 46% of cases with combined Hepatitis C and HIV were advised to seek emergency care (Berry et al., 2019). While improving from the more generalised web-search for symptoms on Google, the results from this research indicate symptom checkers may not provide exemplary results for complex or speciality conditions.

Consumer encounters with the Buoy Health symptom checker have recently been reviewed, investigating patients' intent regarding seeking care following use of their online service (Winn, Somai, Fergestrom, & Crotty, 2019). Buoy Health asks the user concerning their intended level of care both before commencing the online triage and after the assessment. It was found that in 65% of cases the pre and post encounter care level remained the same. In 32% of instances, the intended level of care was reduced and in 4% of instances the intended urgency of care being sought increased. The most common organ system queries pertained to reproductive organs (18%); general (17%); and gastrointestinal systems (15%). The most common symptom type was pain (44%) followed by abnormal functioning (22%) and discharge (8%). The limitations with this research include that only consumer interactions providing both pre and post encounter intent were included, and while consumer behaviour may be influenced by these tools this has not actually been determined, nor was the appropriateness of triage advice established. The researchers suggest that by linking online tools with patient healthcare records, more insights into patient use of these digital services may be obtained (Winn et al., 2019).

Medical centres and related clinics are beginning to use technology to support their everyday business, as internet based clinical support systems have been shown to reduce expenses and improve scheduling (Kinney, 2003). A university in the United States has incorporated a web-based symptom checker into their university campus' medical centre which provides health information, triage advice and an email booking system (Sole et al., 2006). Good congruence between the program's diagnostic classification and the physician's diagnosis with patients who presented to the medical centre for follow-up care was observed ($\kappa = .89$). In 61% of instances the patient was counselled to seek care within 24 hours by the symptom checker, and another 22% were offered self-care advice (Sole et al., 2006). More recently, an app for self-triaging urgent care was reviewed in the Netherlands. The apps triage suggestions were compared to the gold standard of nursing triage advice provided at the time of booking an appointment for at an out-of-hours medical clinic (Verzantvoort et al., 2018). The app corresponded with the nursing triage advice in 81% of instances, with 58% of patients being advised to contact the out-of-hours clinic, another 34% of cases self-care was recommended and 8% of consumers were given 'wait and see' advice. Consumers generally rated the app's clarity at 87% and user satisfaction at 89% (Verzantvoort et al., 2018). This more contemporary research found two-thirds (65%) of consumers intended to follow the triage recommendations from the symptom checker app and have high concordance with traditional nursing advice. The triage accuracy is higher in this research than other studies have reported (Berry et al., 2019; Powley et al., 2016; Semigran et al., 2015; Sole et al., 2006) possibly because the triage advice was limited to a dichotomous choice in the analysis; either 'call a doctor' or 'don't call a doctor'. It is easier to for a program to be 'correct' under these circumstances, less so when the triage disposition range extends beyond two choices or the medical conditions become more complex.

2.5 The seminal article researching online symptom checkers

A Harvard Medical School investigation in the USA researched 23 free online symptom checkers, using 45 clinical vignettes derived from material used to train physicians and other healthcare professionals. They found the correct diagnosis was provided first in only 34% of evaluations (Semigran et al., 2015). Considering the top 20 differential diagnoses presented, only 58% of cases resulted in the correct diagnosis being included in the list. Appropriate triage advice was recommended in 57% of instances. Correctness in care advice increased at higher levels of

urgency: emergent situations (corresponding with the term ‘emergency’ in Australia) were correctly determined in 80% of occasions, 55% in non-emergent (non-urgent) situations and just 33% of self-care scenarios. Their findings suggested symptom checkers tend to be risk-averse, with some sites always triaging emergency care regardless of the situation. Their findings were comparable to research into automated self-assessment tools used in a medical clinic, where the automated assessment was more risk-averse than the physician’s assessment in more than 50% of instances. In this research, the triage categories were ‘999; seek immediate GP care; seek care within 6 hours; seek care within 24 hours; seek a routine appointment; and self-care’. One hundred percent agreement between the automated assessment and physician occurred in 39% of cases. The automated assessment deemed 56% of patients as requiring more urgent care than the GP determined was necessary (in a post-consultation assessment), and just five percent as requiring less urgent care than the GP’s own assessment. The physician’s determined 47% of consultations could have been self-care or been reviewed by a GP routinely (Poote et al., 2014).

There were some limitations with Harvard Medical School’s research. The vignettes used were typically uncomplicated with few comorbidities included. By providing limited information, the results may not provide a truly reflective representation of real-world cases. Further, Semigran et al. (2015) utilised medical terminology when inputting vignette data into the symptom checkers trialled. The authors acknowledged it is unlikely a layperson would use medical jargon when entering information into symptom checkers and suggested this may have impacted on the performance of those tested. Cheng and Dunn (2015) explain that health literacy is a major determinant in how the consumer interprets and applies health-care information supplied electronically. Furthermore, Semigran et al. (2015) excluded symptom checkers based on the same algorithmic logic (i.e. different symptom checkers for which the decision pathways/matrix was derived from the same underlying system), assuming the same decision would be derived from each. This assumption ignores the possibility that local adaptations can be made to the parent product, potentially resulting in differing outcomes.

Semigran et al (2016) later utilised these same vignettes to assess how physicians would diagnose these cases, then compared the physician’s results to symptom checker outcomes. Physicians listed the correct diagnosis first in 72.1% of vignettes and the correct diagnosis within the top three options in 84.3% of vignettes, exceeding the

symptom checker's more modest results (34% and 51.2% respectively). These results reinforce the belief that a physician's diagnostic ability remains the gold standard.

2.6 Are symptom checkers now comparable to physicians?

While most research has found symptom checkers to be inferior to physicians, a recent investigation found Babylon Health's product to be equivalent. Babylon Health completed a study in 2018 comparing their symptom checker entitled 'Babylon Triage and Diagnostic System', which uses artificial intelligence, to human physicians. One hundred clinical vignettes were tested against both the Babylon Health symptom checker and seven GPs. The diagnostic findings were presented in terms of recall (sensitivity) and precision (positive predictive value). The GPs recall was 83.9% and precision was 43.6%; Babylon Health's symptom checker obtained 80.8% recall and 44.4% precision. The results suggest the Babylon Health symptom checker performs comparably with human doctors regarding diagnosis, with only a 0.01% difference in harmonic mean of recall and precision between the physicians and the symptom checker in favour of Babylon Health (Razzaki et al., 2018).

Pertaining to triage advice, the symptom checker may be safer than a GP by 3.9% and almost identical in regard to appropriateness—0.5% lower than the GP average (Middleton et al., 2016; Razzaki et al., 2018). However, methodological flaws have been identified with this research (Fraser, Coiera, & Wong, 2018). For example, during the mock consultations used to evaluate GP performance, other medical personnel acted as patients. When assessing the symptom checker performance, GPs were utilised to input the information from the vignette into the program—it is unknown how these methodologies may have affected the results, but it is entirely possible results would differ if replicated using lay people. Results from the GP performance were also skewed by an outlier with one GP performing below par. Furthermore, no analysis was performed in respect to statistical significance (Fraser et al., 2018), providing little to no indication on the extent to which the results are reflective of true mean differences. Another comparison was performed testing Babylon Health's symptom checker against a subset of 30 clinical vignettes used by Semigran et al. (2015). Babylon Health's symptom checker identified the correct differential diagnosis first in 70% of the vignettes (Razzaki et al., 2018), however, vignettes involving children, dermatological conditions and tetanus were not used potentially creating bias in these results. Therefore, Fraser (2018) advises

any findings from this research should be considered with caution. Additionally, as Babylon Health conducted their own research into their symptom checker, there is potential for sponsorship bias to influence these outcomes.

2.7 The Australian symptom checker

An initial online investigation for Australian symptom checkers appeared to display multiple websites. Subsequent inspection established that all these websites were linked to the Healthdirect Australia symptom checker, essentially operating as a national monopoly. Their portfolio includes the 'after hours GP Helpline'; 'Healthdirect Video Call'; 'My Aged Care; Pregnancy, Birth and Baby'; 'Healthdirect Australia'; 'Carer Gateway' and the 'National Health Services Directory' (B & T Weekly, 2013), each of which are well integrated into the Australian healthcare system. For example, the 'after hours GP Helpline' can upload encounter summaries to a patient's My Health Record, and Healthdirect Australia frequently transfers consumers to agencies such as Triple Zero Emergency Response. Their symptom checker asserts to direct consumers to the correct point of care and help people to understand their "symptoms and possible signs of illness, causes and complications" (Healthdirect Australia, 2016) as well as supporting GPs by increasing the patient's health literacy prior to attending appointments, thereby supporting improvements in patient-physician relations (Healthdirect Australia, 2016, 2017).

Healthdirect Australia's symptom checker was released in 2014, and the smartphone app followed in 2015 with some initial software issues which were later resolved. Symptom checkers can complement or supersede the use of health triage call centres, and when Healthdirect Australia's symptom checker was released, 33% of contacts shifted from the call centre to online (Armstrong, 2018; Healthdirect Australia, 2017; NHS England, 2017). This symptom checker works on a derivative of the NHS symptom checker and follows the same underlying algorithms (Healthdirect Australia, 2016; Semigran et al., 2015). Healthdirect Australia adheres to the Digital Service Standard published by the Digital Transformation Agency (Healthdirect Australia, 2017). They do, however, clearly state their symptom checker is not a diagnostic tool (Healthdirect Australia, 2017), thereby avoiding coming under the Therapeutic Goods Act 1989.

Research conducted by Ng et al. (2012) determined that Healthdirect Australia's nursing call centre referred 2–4% of callers to an ED via ambulance and a further 8–12% of callers to an ED using private transport. The appropriateness of Healthdirect Australia's call centre referrals to an ED was similar to self-referred patients, whilst GP referrals were the most appropriate as deemed by the researchers. Furthermore, Healthdirect Australia's referrals were the least likely to be admitted to hospital (Ng et al., 2012), perhaps indicating inappropriate and risk-averse triage advice. The algorithms used for online symptom checkers and health call centres are very similar (Armstrong, 2018; Semigran et al., 2015), potentially creating inappropriate referrals to emergency services from both services. In addition, approximately 47% of Healthdirect Australia's patients presenting at the ED had been referred to other healthcare services (Ng et al., 2012) suggesting consumer behaviour is not being changed by contacting Healthdirect Australia; that is even when Healthdirect Australia provided triage advice to use an alternative service, consumers still chose to go to the ED. While the appropriateness of ED referrals from Healthdirect Australia's online symptom checker have not been assessed, this issue could possibly be extrapolated to digital referrals as well.

Australia, like many developed nations, suffers over-crowding in EDs, partly from low-acuity patients presenting to the hospital where a primary healthcare service would be more appropriate (Toloo et al., 2011). This places substantial strain on the limited resources available to EDs and causes unnecessary delays in treatment for many patients (Crawford et al., 2014). It is therefore important that any service or tool used by the consumer for triage purposes is effective in recommending the correct care pathway and explains pathway decisions to consumers in a manner that encourages adherence with the advice given.

2.8 Conclusion

Using Google to search for a diagnosis appears to have several shortcomings for the layperson. Some symptoms provided an extensive list of potential diagnoses and the prognostic advice given was variable. A physician may obtain better results through the use of more unique search words for symptoms and being more discriminate with document selection, however even physicians received only 58% diagnostic accuracy using Google.

Diagnostic symptom checkers reviewed for specific medical conditions (such as HIV/Aids) or tested for a range of medical conditions tended towards mediocracy, with accuracy for listing the first diagnosis correctly ranging from 19% to 34% and appropriate triage advice ranged from 37% to 81%. Babylon Health's symptom checker was something of an outlier, suggesting their symptom checker results were comparable to physicians for diagnostic and triage advice, though quite different methods were used to determine diagnostic and triage accuracy and no statistical significance testing was applied to the outcomes. Concerns have been raised regarding the outcomes of this research due to methodological flaws. Additionally, this study is potentially subject to sponsorship bias as the investigation was conducted by the company who would most benefit from the results being favourable.

Healthdirect Australia is the only Australian symptom checker and has not been evaluated in any previously published research. The Healthdirect nursing triage call centre makes referrals to ED with similar appropriateness as self-referred patients, however, patients did not always adhere to the triage advice given.

Limited research has been conducted into symptom checkers, especially pertaining to their diagnostic accuracy and triage advice provided, with Harvard Medical School's American-based investigation being the seminal article on this subject. Given the scarcity of data relating to symptom checkers the Australian population are exposed to and may utilise, it is timely for this shortfall to be addressed. This research will therefore independently investigate the performance of freely available symptom checker apps and websites to which the Australian public are likely to refer to for healthcare advice.

2.9 References

Aboueid, S., Liu, R. H., Desta, B. N., Chaurasia, A., & Ebrahim, S. (2019). The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review. *JMIR Medical Informatics, 7*(2), e13445–e13445. https://doi.org/10.2196/13445

Armstrong, S. (2018). The apps attempting to transfer NHS 111 online. *BMJ (clinical research ed.), 360,* 156. doi: 10.1136/bmj.k156. Retrieved from http://www.bmj.com/content/360/bmj.k156.abstract

B & T Weekly. (2013). *Workshop and Bohemia take out Healthdirect Australia*. Retrieved April 14, 2018, from http://www.bandt.com.au/advertising/Workshop-and-Bohemia-take-out-Healthdirect-Austral

Berry, A. C., Cash, B. D., Wang, B., Mulekar, M. S., Van Haneghan, A. B., Yuquimpo, K., … Green, W. K. (2019). Online symptom checker diagnostic and triage accuracy for HIV and hepatitis C. *Epidemiology & Infection, 147,* 104. doi:10.1017/S0950268819000268

Buijink, A. W. G., Visser, B. J., & Marshall, L. (2013). Medical apps for smartphones: lack of evidence undermines quality and safety. *BMJ Evidence-Based Medicine, 18*(3), 90–92.

Carter, M., Fletcher, E., Sansom, A., Warren, F. C., & Campbell, J. L. (2018). Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. *BMJ Open, 8*(2), e018688.

Chambers, D., Cantrell, A. J., Johnson, M., Preston, L., Baxter, S. K., Booth, A., & Turner, J. (2019). Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open, 9*(8), e027743.

Chen, Z., & Turner, M. R. (2010). The internet for self-diagnosis and prognostication in ALS. *Amyotrophic Lateral Sclerosis, 11*(6), 565–567. doi:10.3109/17482968.2010.513054

Cheng, C., & Dunn, M. (2015). Health literacy and the Internet: a study on the readability of Australian online health information. *Australian and New Zealand Journal of Public Health, 39*(4), 309–314. https://doi.org/10.1111/1753-6405.12341

Cowie, J., Calveley, E., Bowers, G., & Bowers, J. (2018). Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. *International Journal of Environmental Research and Public Health, 15*(5), 896. doi:10.3390/ijerph15050896

Crawford, K., Morphet, J., Jones, T., Innes, K., Griffiths, D., & Williams, A. (2014). Initiatives to reduce overcrowding and access block in Australian emergency departments: a literature review. *Collegian, 21*(4), 359–366. doi:10.1016/j.colegn.2013.09.005

Fraser, H., Coiera, E., & Wong, D. (2018). Safety of patient-facing digital symptom checkers. *The Lancet, 392*(10161), 2263–2264. doi:10.1016/S0140-6736(18)32819-8

Healthdirect Australia. (2016). *About Healthdirect Symptom Checker.* Retrieved February 2, 2019, from https://www.healthdirect.gov.au/about-healthdirect-symptom-checker

Healthdirect Australia. (2017). *Annual Report 2016 – 2017 Healthdirect Australia.* Retrieved April 14, 2018, from https://media.healthdirect.org.au/publications/Annual_Reports_Business_Highlights_2016-2017.pdf

Jacobs, W., Amuta, A. O., & Jeon, K. C. (2017). Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. *Cogent Social Sciences, 3*(1), 1302785.

Jutel, A., & Lupton, D. (2015). Digitizing diagnosis: a review of mobile applications in the diagnostic process. *Diagnosis, 2*(2), 89–96. https://doi.org/10.1515/dx-2014-0068

Kinney, W. C. (2003). Web-based clinical decision support system for triage of vestibular patients. *Otolaryngology—Head and Neck Surgery, 128*(1), 48–53.

Lupton, D., & Jutel, A. (2015). 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Social Science & Medicine, 133*, 128–135. https://doi.org/10.1016/j.socscimed.2015.04.004

Madan, A. (2014). *webGP: the virtual general practice.* Pilot Report (Hurley Group, London).

Marco-Ruiz, L., Bønes, E., de la Asunción, E., Gabarron, E., Aviles-Solis, J. C., Lee, E., … Bellika, J. G. (2017). Combining multivariate statistics and the think-aloud protocol to assess Human-Computer Interaction barriers in symptom checkers. *Journal of Biomedical Informatics, 74,* 104–122. doi:10.1016/j.jbi.2017.09.002

Middleton, K., Butt, M., Hammerla, N., Hamblin, S., Mehta, K., & Parsa, A. (2016). Sorting out symptoms: design and evaluation of the "babylon check" automated triage system. *arXiv preprint arXiv:1606.02041.* Retrieved April 04, 2018, from https://arxiv.org/abs/1606.02041

Millenson, M. L., Baldwin, J. L., Zipperer, L., & Singh, H. (2018). Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. *Diagnosis, 5*(3), 95–105.

Ng, J. Y., Fatovich, D. M., Turner, V. F., Wurmel, J. A., Skevington, S. A., & Phillips, M. R. (2012). Appropriateness of healthdirect referrals to the emergency department compared with self-referrals and GP referrals. *Medical Journal of Australia, 197*(9), 498–502. doi:10.5694/mja12.10689

NHS England. (2017). *NHS111 Online evaluation.* Retrieved October 9, 2019, from https://askmygp.uk/wp-content/uploads/111-Online-Evaluation-DRAFT_.pdf

Nijland, N., Cranen, K., Boer, H., van Gemert-Pijnen, J. E. W. C., & Seydel, E. R. (2010). Patient use and

compliance with medical advice delivered by a web-based triage system in primary care. *Journal of Telemedicine and Telecare, 16*(1), 8–11. doi:10.1258/jtt.2009.001004

Nijland, N., van Gemert-Pijnen, J. E. W. C., Boer, H., Steehouder, M. F., & Seydel, E. R. (2009). Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. *International Journal of Medical Informatics, 78*(10), 688–703.

Payne, K. F. B., Wharrad, H., & Watts, K. (2012). Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. *BMC Medical Informatics and Decision Making, 12*, 121. https://doi.org/http://dx.doi.org/10.1186/1472-6947-12-121

Poote, A. ., French, D. ., Dale, J., & Powell, L. (2014). A study of automated self-assessment in a primary care student health centre setting. *Journal of Telemedicine and Telecare, 20*(3), 123–127. https://doi.org/10.1177/1357633X14529246

Powley, L., McIlroy, G., Simons, G., & Raza, K. (2016). Are online symptoms checkers useful for patients with inflammatory arthritis? *BMC Musculoskeletal Disorders, 17*(1), 362.

Razzaki, S., Baker, A., Perov, Y., Middleton, K., Baxter, J., Mullarkey, D., … Majeed, A. (2018). A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *ArXiv Preprint ArXiv:1806.10698.*

Riches, N., Panagioti, M., Alam, R., Cheraghi-Sohi, S., Campbell, S., Esmail, A., & Bower, P. (2016). The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PloS One, 11*(3), e0148991. doi:10.1371/journal.pone.0148991

Semigran, H. L., Levine, D. M., Nundy, S., & Mehrotra, A. (2016). Comparison of physician and computer diagnostic accuracy. *JAMA Internal Medicine, 176*(12), 1860–1861.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ (clinical research ed.), 351*, 3480. doi:10.1136/bmj.h3480.

Shen, C., Nguyen, M., Gregor, A., Isaza, G., & Beattie, A. (2019). Accuracy of a Popular Online Symptom Checker for Ophthalmic Diagnoses. *JAMA Ophthalmology, 137*(6), 690–692. doi:10.1001/jamaophthalmol.2019.0571

Sole, M. Lou, Stuart, P. L., & Deichen, M. (2006). Web-based triage in a college health setting. *Journal of American College Health, 54*(5), 289–294. doi: 10.3200/JACH.54.5.289-294

Tang, H., & Ng, J. H. K. (2006). Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. *Bmj, 333*(7579), 1143–1145.

Toloo, S., FitzGerald, G., Aitken, P., Ting, J., Tippett, V., & Chu, K. (2011). Emergency health services: demand

and service delivery models. Monograph 1: literature review and activity trends.

Verzantvoort, N. C. M., Teunis, T., Verheij, T. J. M., & van der Velden, A. W. (2018). Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? *PloS One, 13*(6). doi:10.1371/journal.pone.0199284

Winn, A. N., Somai, M., Fergestrom, N., & Crotty, B. H. (2019). Association of Use of Online Symptom Checkers With Patients' Plans for Seeking Care. *JAMA Network Open, 2*(12), 1918561. doi:10.1001/jamanetworkopen.2019.18561

CHAPTER 3

METHODOLOGY

Thesis Map
- Chapter One – Introduction
- Chapter Two – Literature Review
- ➡ Chapter Three – Methodology
  - Design and Materials
  - Procedure
- Chapter Four – Results
- Chapter Five – Discussion
- Chapter Six – Conclusion

Design and Materials

3.1 Conceptual Framework

Evaluative research was undertaken to assess the performance of online symptom checkers and apps. According to Patton (1986; p.224) evaluative research "is the systematic collection of information about the activities, characteristics, and outcomes of programs for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs are doing and affecting." A less structured concept is that evaluation involves assessing the worth or merit of something (Mugenda & Mugenda, 2003). The essential characteristics of evaluative research is to apply social research to existing knowledge; testing the effectiveness of this knowledge by applying scientific methods to appraise and guide practical action (refer to Figure 3.1 below). Evaluative research does not attempt to prove anything, rather seeks to improve by facilitating change (O'Leary, 2004, p.133; Patton, 1986).

Please refer to the following references to review the diagram for Figure 3.1:

O'Leary, Z. (2004). *The essential guide to doing research*. *Sage* (3rd editio). London: SAGE Publications Ltd. Retrieved from https://study.sagepub.com/oleary3e

Patton, M. (1986). *Utilization-focusd evaluation* (2nd edition). Beverly Hill: SAGE Publications.

Figure 3.1: **From knowledge to change – the goals of research** (O'Leary, 2004).

Clarke and Dawson (1999) suggest evaluative research does not seek to make predictions, it simply describes what is already there. Evaluative research can be further divided into three broad categories; descriptive, normative

and cause-and-effect. Descriptive evaluation investigates how users are engaged with and participate in a particular program; normative evaluation seeks to determine if the program operates as originally intended; while cause-and-effect evaluation observes what changes can be seen as a result of a specific program intervention (Chelimsky, 1985).

Shriven (1967) further separates evaluative research into 'formative' and 'summative'; formative evaluation being concerned with providing feedback to improve something (for example a process or program), while summative evaluation aims to determine the impact or effectiveness of a program. Summative evaluative research is targeted towards policymakers, funders or the public and focuses on implementation issues and outcome measures. The methodology tends to be quantitative, and the frequency of data collection is limited (ie: conducted once). The role of the evaluator is independent, and a formal report is typically completed at the end of the evaluation (Herman, Morris, & Fitz-Gibbon, 1987).

The methods utilised by this research are normative and summative in nature, and reports on the output or performance from online symptom checkers. As variables are not manipulated nor interventions applied, other research methodologies such as experimental, quasi-experimental or correlational studies were deemed unsuitable for this investigation.

This evaluative process was applied to symptom checkers readily available to Australian consumers, whereby each symptom checker was evaluated on the ability of that website or app to diagnose and/or triage consumers to the appropriate outcome concordant with expert assessment. Healthdirect Australia (2018), as Australia's only symptom checker and supported by the Council of Australian Governments and funded by the Australian Government, was used as the 'benchmark' by which other triage symptom checkers were assessed.

3.2 Overview of research design

Figure 3.2 below provides an overview of the research design. The flowchart implies a simplistic linear approach to the design, in reality the process was more organic and dynamic in nature with multiple aspects being revised in tandem.

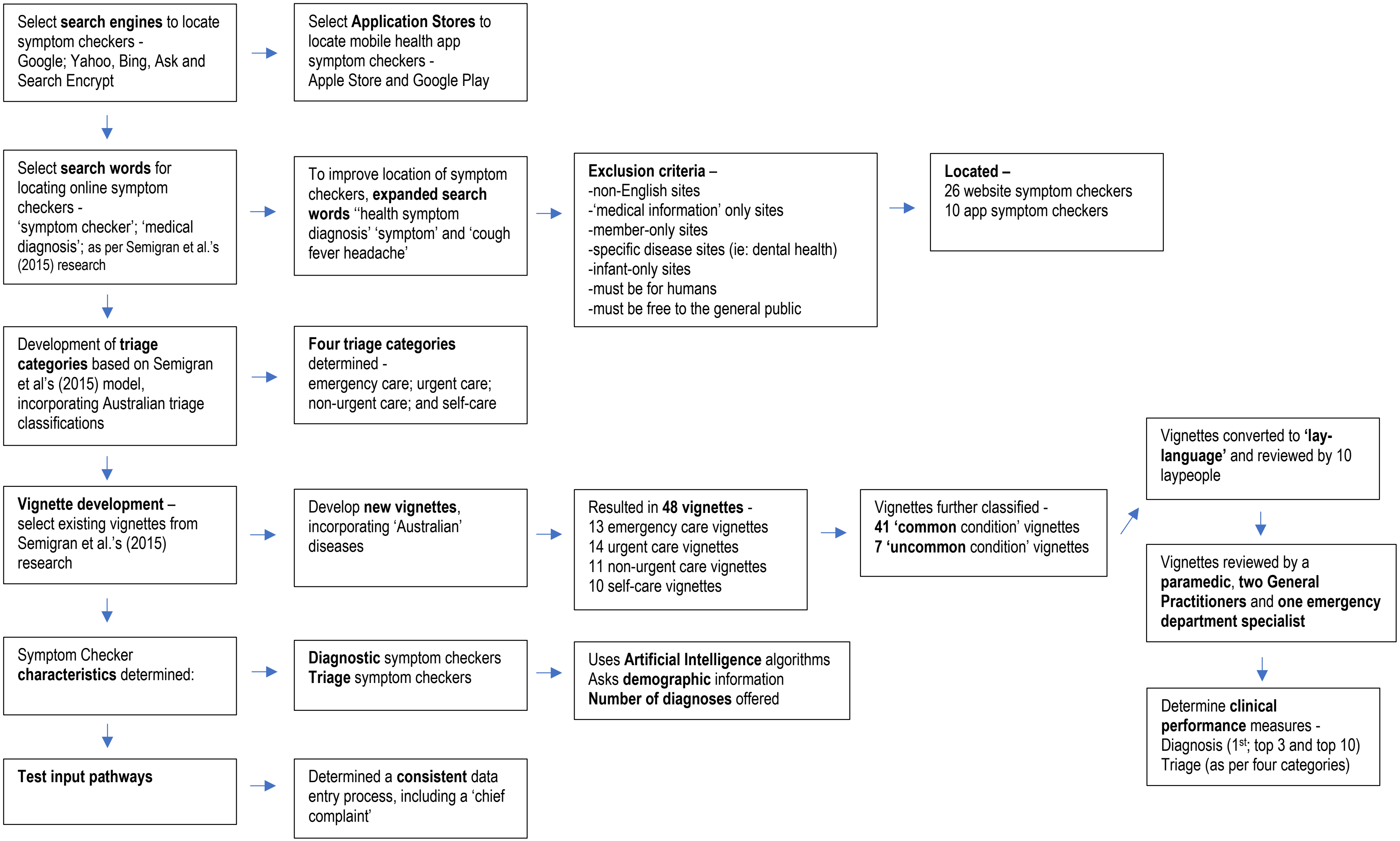


Figure 3.2: Overview of research design

3.3 Selection of symptom checkers

To determine the free English language symptom checkers most prominently available on the worldwide web to the Australian population, a systematic search was undertaken of symptom checkers available in five popular search engines (Google, Yahoo, Bing, Ask, and Search Encrypt) (eBiz, 2017). Semigran Linder, Gidengil, and Mehrotra (2015) reviewed Google and Google Scholar to locate symptom checkers for their research. Google Scholar is technically a search engine, however, is largely devoted to academic literature and so was not utilised in this evaluation. The first three pages of listings from each search engine were filtered, to identify symptom checkers that were publicly available, in English, for humans (not animals or cars) and were not restricted to a specific medical specialty such as psychological health or dental health. Research suggests many consumers do not go past the 'fold' on the screen (where they would be required to scroll down to view further links), or at most only scroll through the first page or two to locate information they desire (Eysenbach & Köhler, 2002; Lauckner & Hsieh, 2013). Therefore, for thoroughness, the first three pages were utilised from each individual search.

Five search phrases were used to identify symptom checkers; 'symptom checker', 'medical diagnosis', 'health symptom diagnosis', 'symptom' and 'cough fever headache'. This was based on research conducted by Semigran et al. (2015), where the first two phrases were employed in their online search. However, the search phrase 'health symptom diagnosis' was subsequently included in this research as the phrase 'medical diagnosis' produced poor results in Australia for symptom checkers. The phrases 'symptom' and 'cough fever headache' were also incorporated, as the general public may search symptoms directly, then follow a link to a symptom checker. Generally, these two search options came up with dictionary definitions and websites providing general medical and health information, with far less symptom checkers being filtered into the results.

The use of multiple search engines and search phrases is important as each search engine (such as Google or Bing) uses slightly different algorithms to index information held on the internet (Franklin, 2002). Search engines use 'spiders' to build lists of search words and phrases to index in their own personal database. Spiders are software robots specifically designed to construct lists of words found on Web pages. They do this by 'web crawling', sending the spiders to browse heavily used servers and popular web pages. The spider will then follow every web-link on that page and every subsequent page, 'crawling' through the internet adding information to the index (Sullivan, 2002). Certain websites block crawlers, (especially gaming sites) whilst others use 'meta-tags'

which flag key words and phrases, essentially telling the spider the concepts the page owner wants indexed (Franklin, 2002). On each webpage the spider visits it takes note of key words and where they are placed on the webpage. For instance, Google gives added weighting to key words in titles, sub-titles and URL addresses than more common words such as 'them' or 'and' (Franklin, 2002). Each search engine will therefore gather information for their own indexing system in a different way, producing different results when consumers search the internet. In essence, a consumer is not technically searching the worldwide web, rather they are searching Google or Bing's personal index of the worldwide web (Franklin, 2002).

There are two main application distribution services for smartphones; Google Play for android products and the App Store for Apple devices. Ranking algorithms contained within the app stores programming commonly result in the most popular apps appearing at the top of the list when conducting a search (Boudreaux et al., 2014). Research conducted by Choi and Stvilia (2014) found students were motivated to download apps which were free, had positive comments regarding the app, had no advertisements and were designed well. Participants were also more inclined to use apps which were ranked first in their search. Therefore, for the purposes of this research, the first fifteen listed apps were considered, using the same search phrases and filters mentioned above.

Symptom checkers limited to providing medical information or acting as medical dictionaries were excluded. Additionally, member-only sites were omitted, as registering for the symptom checker frequently required more information than supplied in each vignette (e.g. height, weight, or health insurance details). Sites specific to infants (e.g. children under two years of age) were excluded as they are specialised to very young children and fall outside the scope of this research. Advertisements were ignored as research indicates they are frequently overlooked by consumers (McDonald & Cranor, 2009). Most symptom checkers were free, although several applications incurred a fee for download and were similarly excluded. Some sites even advised they were free, but the directory of signs and symptoms the user could access were significantly reduced on their 'free' app, encouraging users to upgrade their app by purchasing the more comprehensive package. Such sites were also excluded from this research. A total of 750 websites and 150 apps were screened, with 179 websites and 77 apps deemed ineligible for this research. The filtering process yielded 26 websites and ten apps providing a symptom checker readily available to the Australian general public, with Internet Protocol (IP) addresses operating from the US, Israel, Canada, Poland

and Australia. Symptom checker user interfaces can be marketed in one country however, the ‘back-end’ logic (such as IP addresses) can operate elsewhere in the world. So, while there is an IP address identified as Israeli, there is not a specific symptom checker marketed to Israel included in this evaluation. A list of the free symptom checkers used in this evaluation is depicted in Appendix A.

Analysis was undertaken during the research phase to identify websites that link to the same symptom checker in much the same way all Australian sites link to Healthdirect Australia (Appendix A). In some instances, multiple websites filter to a common database and single underlying algorithm, where it would be reasonable to expect the diagnosis to be identical for each website. However, given triage advice may be adapted to local services and healthcare systems, (again for thoroughness) each website was individually tested. Other websites may purchase a product from another company then customise it to meet their own needs, as Healthdirect Australia (2016) has from the NHS in England. Healthdirect Australia’s symptom checker has an IP address located in Australia, not the UK. Accordingly, these websites were also included in the study (Semigran et al., 2015). The premise for this is that some algorithms may be adapted locally, potentially changing diagnostic or triage advice.

3.4 Symptom checker characteristics

The symptom checkers were characterised by whether they assessed and provided diagnostic and/or triage advice. The number of listed differential diagnoses was identified, whether demographic questions were asked (as a minimum, asking the consumer’s age and gender) and whether they clearly stated they utilised AI algorithms or not. AI programs are currently the promotional material marketed to consumers; therefore, this has been the characteristic assessed for this research. If a site claimed the use of AI technology for their symptom checker, this was taken at ‘face value’. It was outside the scope of this current research to determine the format or nature of the AI used by a symptom checker.

An attempt to identify whether the system used Schmitt or Thompson nursing triage guidelines was made, however almost no websites provided this level of information. Schmitt or Thompson guidelines are frequently utilised in

telephone triage call centres and comparisons have been drawn in the past to these types of call centres and symptom checkers (Armstrong, 2018; Semigran et al., 2015).

3.5 Vignettes

Research conducted by Semigran et al. (2015) used 45 patient vignettes drawn from various sources such as medical resource websites and material used to educate health professionals. We based a number of vignettes off those utilised by Semigran et al.'s (2015) investigation. However, as this research is being conducted in Australia, certain vignettes were adapted or created to incorporate Australian-specific illnesses (Appendix B). The existing and newly created vignettes resulted in 48 scenarios, which were then scrutinised by one paramedic with 22 years on-road experience (a lecturer from Edith Cowan University's School of Medical and Health Science). This process considered the original vignette Semigran et al. (2015) utilised, reviewed the original core signs and symptoms and amended medical terminology and triage categories as required. Agreement between the researcher and paramedic was obtained, confirming the describing signs and symptoms* were valid for the diagnoses, and that each vignette was ascribed the appropriate level of triage acuity. [*Signs are objective evidence of an illness visible to a clinician, for example abdominal tenderness, whereas symptoms are subjective, a phenomenon experienced only by the individual, such as a headache. Whilst vignettes are based on both signs and symptoms, only 'symptoms' are input into the program by consumers, as the computer cannot 'see' signs. Accordingly, only the term 'symptom' will be utilised throughout the thesis.] Additional examination and consensus were then gained by a further three currently practising medical practitioners: two GPs and one ED physician with 87 years of accumulated clinical experience; confirming face and content validity of the vignettes from a medical perspective. The range of disposition advice provided by the symptom checkers were sorted into four categories of triage acuity (emergency, urgent, non-urgent and self-care). A further review was conducted by ten members of the general public. This was to confirm the vignettes and been converted appropriately to layperson's language, thereby ensuring the face and content validity of the vignettes from this perspective.

As a consequence of the validation process, some of Semigran et al.'s (2015) vignettes were excluded, whilst some scenarios were recategorised. For example, Semigran et al. (2015) utilised a scenario where a patient with

discoloured sputum was classified into the self-care triage category. In Australia, any person presenting with discoloured sputum would be expected to attend a medical practitioner, so this vignette was recategorised to the urgent care triage category. Some vignette triage classifications generated debate prior to resolution of consensus, as clinical decisions usually consider contextual factors such as timing and local access to resources (e.g. diagnostic imaging services). Notably the triage classification for the deep vein thrombosis vignette was categorised as urgent for this research, differing from the emergency category determined by Semigran et al. (2015) reflecting recent advances with new oral anticoagulant therapy (Kruger, Eikelboom, Douketis, & Hankey, 2019).

One limitation Semigran et al. (2015) identified with their research was that medical terminology was utilised to input information into the symptom checkers. This research addressed this shortfall by using 'lay' terms for symptoms. The symptom checkers were, therefore, tested in a similar way a non-medically trained member of the general public would interact with the product. The 'lay' terms used were based on the experiences of the paramedic and medical practitioners who reviewed the vignettes. For example, amending the symptom 'febrile' to the more commonly used term 'fever'. Each vignette was modified to a core set of symptoms, as symptom checkers ask a series of questions or request a set of symptoms to establish a possible diagnosis and provide triage advice.

3.5.1 Customising vignettes for Australian conditions

The vignettes have been customised from Semigran et al.'s (2015) investigation to target the Australian audience more directly. A vignette from Semigran et al.'s (2015) study for Rocky Mountain Spotted Fever has been substituted for a Queensland Tick Typhus vignette, as this is a close Australian equivalent. Whilst uncommon to have Queensland Tick Typhus as an emergency illness, it is possible for this presentation to occur (McBride, Hanson, Miller, & Wenck, 2007; Wilson, Tierney, Lai, & Graves, 2013). In addition, a vignette for Ross River Virus has been designed, as this disease is more endemic to Australia and is generally a self-limiting condition (Dore & Auld, 2004; Wertheim, Horby, & Woodall, 2012; Westhorpe, 2014). Despite treatment typically being self-care, the symptoms synonymous with this condition would suggest a patient needs to be reviewed by a medical practitioner to ensure the appropriate diagnostic tests were performed, as this virus closely mimics other diseases such as

mononucleosis (Murtagh & Rosenblatt, 2015). The last vignette designed for Australian conditions is for Hendra Virus. Hendra Virus is uniquely Australian and results from infection spread by bats of the genus Pteropus through an intermediate equine host, and can be fatal to infected humans (Queensland Government, 2017; Young, Selvey, & Symons, 2011). This vignette was therefore deemed an emergency triage example. This resulted in a total of 48 vignettes, covering the spectrum of triage categories; emergency (*n* = 13), urgent (*n* = 14), non-urgent (*n* = 11) and self-care (*n* = 10). Refer to Appendix B for vignette information and associated symptoms.

3.5.2 Common and uncommon vignettes

Semigran et al. (2015) also categorised vignettes into whether they were common or uncommon presentations based on the frequency of the diagnosis among ambulatory visits in the USA 2008–2010. These categories are also appropriate to be utilised for this research. The common and uncommon categories were substantiated based on the Australian healthcare system statistics which identify the prevalent conditions found in Australia. This information was largely sourced from the BEACH research data for the decade 2005–06 to 2015–16 (Britt et al., 2016) which analysed general practice activity from approximately one thousand GPs for one hundred consecutive encounters with consenting patients in Australia. Amongst other information, the 'Most frequent reasons for a patient's visit to a GP*'* and the 'Most frequently managed chronic problems' were evaluated, which allowed confirmation of most vignettes' categorisation into 'common' or 'uncommon' conditions. The remaining common or uncommon categorisations were confirmed by the Australian Institute of Health and Welfare's ED care statistics (Australian Institute of Health and Welfare, 2017) or the Victoria State Government's website, 'Better Health Channel' (VIctoria State Government, 2018). Ultimately, 41 vignettes were deemed common medical conditions with the remaining seven being classified as uncommon. This was ratified by the same two practising GPs mentioned above (Appendix B). Uncommon vignettes represent 15% in this research, similar to the 'real-world' distribution presenting to Australian GPs which is reported at 10% (Cooke, Valenti, Glasziou, & Britt, 2013).

3.5.3 Input pathways assessment for vignettes

A preliminary assessment was performed to determine how the vignettes would work on different symptom checker programs, using an independent 'test' vignette not being utilised in this research (refer to Figure 3.3 below). The test vignette was 'influenza' with symptoms of a cough, fever and headache.

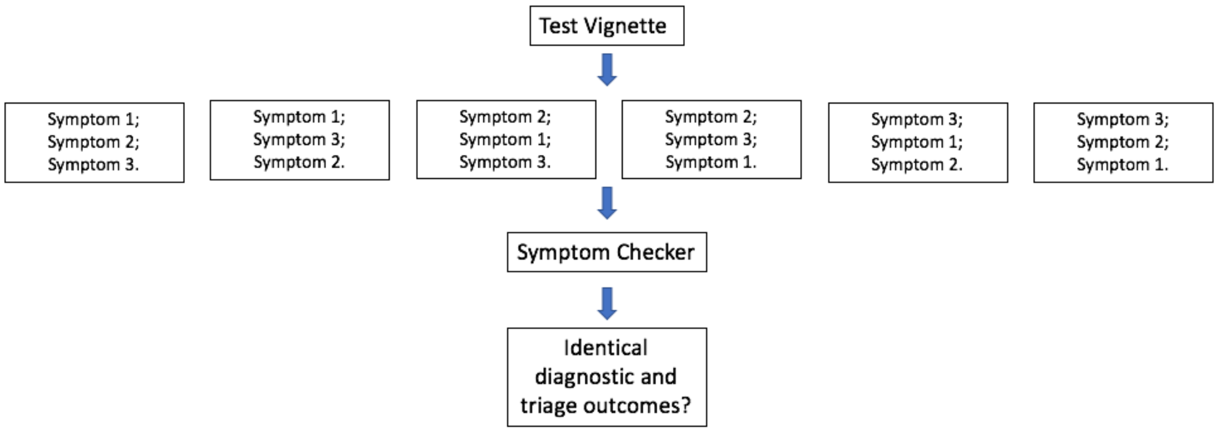


Figure 3.3: **Input pathways assessment for vignettes**

The vignette was entered into four popular symptom checkers (Healthdirect Australia, Mayo Clinic, WebMD and Isabel Healthcare), changing the order of the symptoms with each attempt. The test vignette had three symptoms, allowing a total of six possible orders in which the symptoms could be input into each program. Accordingly, each symptom checker had six input pathway assessments performed on the test vignette to determine how the symptom checkers would respond to such variations in input.

With Isabel Healthcare (2019), no matter which order the symptoms were entered into the program the list of possible diagnoses was identical, as was the disposition advice. However, the other three symptom checkers worked quite differently. Mayo Clinic (2018) for example, only provides a list of symptoms for the user to select from, and this list does not include fever as a possible chief complaint in their program, the 'chief complaint' being the main symptom of a medical condition (Murtagh & Rosenblatt, 2015, p. 15). Users could only select headache or cough as the primary symptom (or chief complaint). Fever was only asked about in subsequent questioning. Healthdirect Australia (Healthdirect Australia, 2016b) only allows a chief complaint to be entered and asks a series

of follow-up questions which may or may not ask about other symptoms. WebMD (2017) works differently again; appearing to add extra 'weight' to the first listed symptom, as their lists of potential diagnoses altered depending on the order information was input. WebMD (2017) does not provide triage advice. A brief example of results for WebMD is displayed below in Table 3.1 (WebMD, 2017).

Table 3.1

WebMD: diagnostic results for test vignette

Order of symptoms	Potential diagnoses provided
Cough, fever, headache	Common cold Acute sinusitis Asthma (teen & adult) Viral pneumonia Whooping cough
Cough, headache, fever	Common cold Acute sinusitis Asthma (teen & adult) Viral pneumonia Whooping cough
Fever, cough, headache	Common cold Acute sinusitis Viral pneumonia Asthma (teen & adult) Strep throat
Fever, headache, cough	Common cold Acute sinusitis Viral pneumonia Asthma (teen & adult) Strep throat
Headache, fever, cough	Common cold Acute sinusitis Viral pneumonia Asthma (teen & adult) Tension headache
Headache, cough, fever	Common cold Acute sinusitis Viral pneumonia Asthma (teen & adult) Tension headache

Consequently, if the system was designed to add extra weighting to the chief complaint, different differential diagnoses would be provided with each shuffling of the symptoms. Programs that were not designed to add weight to the first listed symptom (usually considered as the chief complaint), then the answers would be identical

irrespective of the order of symptoms. This issue highlighted the necessity for uniformity with inputting the chief complaint and symptoms in identical order for each symptom checker. Regardless of the algorithms used by various products the vignettes would work as intended. As many of the vignettes had been used by Semigran et al. (Semigran, Levine, Nundy, & Mehrotra, 2016; Semigran et al., 2015) to evaluate both online symptom checkers and physicians for diagnostic accuracy, we deemed the vignettes as being reliable.

3.6 Triage categories

This research methodology was adapted from Semigran et al.'s (2015) research design in which their triage categories were emergent, non-emergent and self-care (refer to Table 3.2 below). Semigran et al. (2015) developed their triage options from various symptom checkers as they conducted their research.

Table 3.2

Triage categories and symptom checker advice options for each category.

Derived from Semigran et al.'s (2015) research.

Triage Category	Symptom checker care advice options
Emergent	Call an ambulance Go to the emergency department See a GP immediately
Non-emergent	Call a GP or primary healthcare provider See a GP or primary healthcare provider Go to an urgent care facility Go to a Specialist Go to a retail clinic Have an e-visit
Self-care	Stay at home Go to a pharmacist

Some of the triage care options suggested by overseas symptom checkers are unsuitable for the Australian healthcare system. For example, Semigran et al. (2015) had 'Go to a Specialist' as a non-emergent care option. Under Medicare in Australia it is not possible to self-refer to a Specialist without a referral from a GP. However, as it is a triage option suggested by some symptom checkers, this is still categorised as a non-urgent option for triage even though it will not be possible to follow this advice in Australia.

Table 3.3 below depicts the triage options Healthdirect Australia use for their nursing triage call centre (Healthdirect Australia, 2017) and incorporates the triage information found in Semigran et al.'s (2015) research and subsequent triage advice provided during this research. This system of triage categorisation more accurately reflects the Australian healthcare system. 'Emergency care', for the purposes of this research refers to an illness or injury that requires immediate ED expertise. 'Urgent care' refers to a condition requiring prompt attention by a medical professional within 24 hours but is not immediately life-threatening. 'Non-urgent care' refers to needing to see a healthcare provider sometime in the near future. 'Self-care' is deemed as assistance a person provides to themselves often utilising over-the-counter products.

Table 3.3

Triage categories and disposition advice provided by symptom checkers.

Triage Category	**Disposition advice provided by symptom checkers.**
Emergency (Requiring immediate medical expertise)	Call an ambulance Activate triple zero (000) Go to the emergency department Seek medical help immediately- this may be a life-threatening condition.
Urgent Requiring medical attention within 24 hours)	See a GP immediately See a GP within 4 hours See a GP within 24 hours Go to an urgent care facility Call a nurse immediately
Non-urgent (See a healthcare provider in the near future)	Call a GP or primary healthcare provider See a GP or primary healthcare provider Go to a Specialist* Go to a retail clinic Have an e-visit Minor injuries unit Walk-in clinic Telemedicine Speak to a nurse about your child
Self-care (Assistance a person provides to themselves)	Stay at home Go to a pharmacist
*requires GP referral in Australia and therefore not used by Australian websites	

Procedure

3.7 Clinical performance

Forty-eight vignettes were systematically entered into 26 websites and 10 smartphone apps, generating 1,858 symptom checker vignette tests (SCVTs). Refer to Figure 3.4 below for the process, and Appendix C for the full selection of symptom checkers and summary of diagnostic and triage advice. The resulting diagnosis and/or triage advice was recorded into an SPSS v25 database. There was only one person entering vignettes to ensure consistency in data entry and collection.

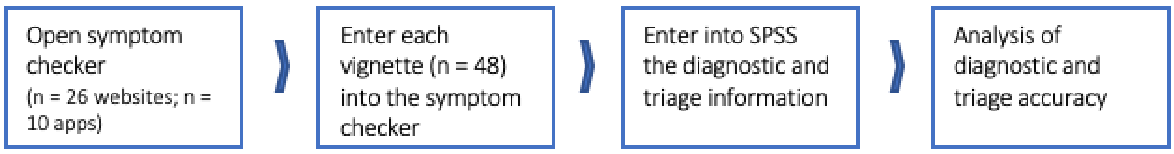


Figure 3.4: Evaluation process for this research

3.8 Data entry

Exactly how the data was entered into the symptom checker was dependent on how the individual system operated, but frequently the demographic information was entered first followed by the simplified symptoms for each vignette. Semigran et al. (2015) found that for two scenarios additional information was required by at least one symptom checker in order to gain a diagnosis. When this occurred, the researchers added additional information into the symptom checker to complete the process. In their abbreviated version of the 'signs and symptoms' for each condition in their Appendix, they incorporated this extra information and italicised the exact wording they used to test the program. It appears this extra information was only supplied by Semigran et al. (2015) if it was requested. For example, if symptoms A, B and C were input into the program and a symptom checker enquired about symptom D, Semigran included symptom D's information in the assessment. Therefore, for this research, Semigran et al.'s additional information ('symptom D') has been included in the vignettes already but was not provided to the symptom checker in the first instance.

A logical process was followed for entering symptoms. Given the outcome of the input pathways assessment above, symptoms were carefully input in the same order for each symptom checker. The first symptom entered was the chief complaint, as deemed by the two GPs ratifying the vignettes. Each subsequent entry followed the order the symptom appeared in the vignette. An example is provided below in Figure 3.5 (Vignette: Viral upper respiratory).

Vignette 38 Viral upper respiratory	Mr. R. is a 56 year-old man who presents to you with 6 days of non-productive cough, nasal congestion, and green nasal discharge. He has had intermittent fevers as high as 100.8. His physical examination is normal except for rhinorrhea. He is otherwise healthy, except for chronic osteoarthritis of the right knee. He has no drug allergies.	56 year old male, 6 day cough, nasal congestion, green nasal discharge. Fever (38.2) and runny nose. Chief complaint = cough

Figure 3.5: Vignette: Viral upper respiratory

For the vignette 'viral upper respiratory', the demographic information input would be the gender as 'male' and age of '56'. The chief complaint was entered first as a 'cough'. Subsequent symptoms were entered in the order of 'nasal congestion'; 'green nasal discharge'; 'fever' (the temperature of 38.2^{o}C would have been input if requested); and finally, 'runny nose'.

Some symptom checkers did not allow a symptom to be entered. In this instance, the flow chart in Figure 3.6 below was adhered to.

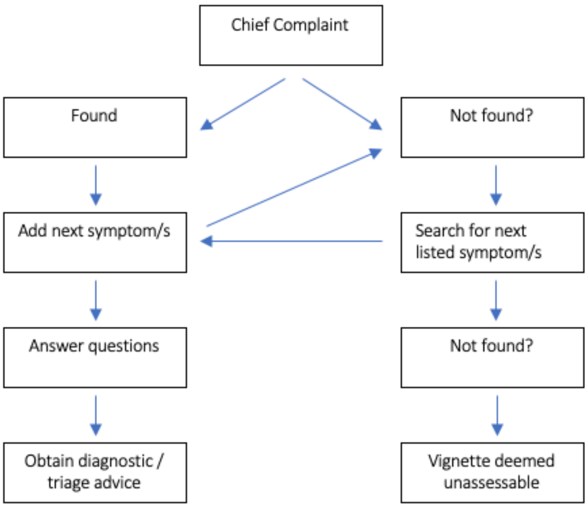


Figure 3.6: **Flow chart for selecting symptoms**

Some symptom checkers only provided a limited list of symptoms which could be select from. This included sites where an image of a body could be clicked and then a drop-down list of symptoms could be selected from. What's My Diagnosis (n.d.) is one such website and provides a list of only 35 common symptoms. A 'diagnosis' list appears even before anything has been input into the program; this diagnosis list updates every time a symptom is selected from the list provided.

Some symptom checkers allowed typing in of symptom/s and simultaneously provided a drop-down list of their accepted symptoms. Several symptom checkers also provided an image of a body which would work by selecting the body part and selecting a symptom relating to that region of the body. Symptomate (2019) is an example of a website which utilises both options.

For consistency, unless the symptom checker only allowed the selection of symptoms via clicking on a body part, the information was typed in. The most appropriate option from the drop-down list provided was then selected. Rarely, sites such as Isabel Healthcare (2019) allowed provision of one's own symptom/s verbatim and/or select from a drop-down box. Both options are equally acceptable on this site. Most sites would only give a drop-down list and the symptom had to be selected from this list.

Multiple symptom checkers asked questions after the symptoms has been input. The way questions were answered followed the flow chart below (Figure 3.7).

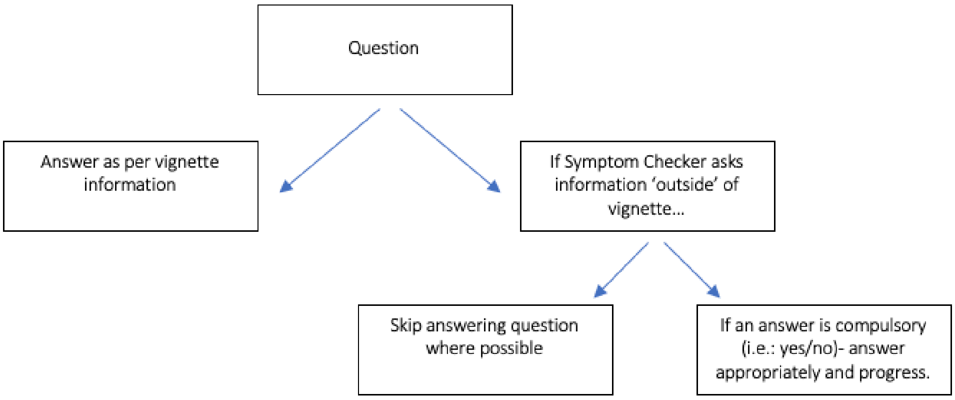


Figure 3.7: **Flow chart for answering follow-up questions consistently**

For instance, if a symptom checker asked, 'Have you been diagnosed with cancer?' This information would not be provided in the vignette. If at all possible, the question would be skipped. If it was a yes/no answer where it was compulsory to respond in order to progress to the next question and gain a diagnosis, this particular question would be answered 'no' as the vignette does not indicate the patient has cancer. If the question was 'Is this the worst headache you have ever experienced?' and it was pertaining to the meningitis vignette, then the answer in that instance would be 'yes' as meningitis headaches are usually severe.

Where a symptom checker was on multiple platforms (for example Symptomate has a webpage, and apps available from Google Play and Apple Store), all platforms were assessed in tandem to ensure each vignette was input identically and each follow-up question was answered identically. In the instance a website's symptom checker was identified to use a product from another company (such as MedicineNet and RxList, who use a WebMD product), one vignette was run in tandem on both websites, to establish whether the sites were identical or not. If they were not identical, each site was assessed independently.

3.9 Diagnostic accuracy

Clinical performance of the symptom checkers was determined by diagnostic and triage accuracy. The assessment for diagnostic accuracy was based on whether the correct differential diagnosis appeared first; in the top three or in the top ten diagnoses options. In Semigran et al.'s (2015) research some symptom checkers yielded up to 99 potential diagnoses. This seemed an unreasonably long list of possible conditions to present to a consumer, making any diagnostic list nonsensical. Therefore, the list of 'correct diagnosis' conditions was capped to the first ten options. An 'incorrect diagnosis' would result from the symptom checker failing to identify the correct condition within the first ten possibilities. This is an arbitrary number established on the basis that a range of illnesses can present with identical symptoms. For example, the back-pain vignette could have seven acceptable 'correct' answers, based on the core symptoms. Similarly, many upper respiratory tract conditions have commonality with symptoms, so limiting the 'correct' diagnosis to only three options was deemed too narrow (Appendix B). A certain amount of precision was still required for a diagnosis to be accepted as correct. Vignette #1 was a scenario for acute liver failure. Given the history described in the vignette, acetaminophen (known as paracetamol in Australia) poisoning was also considered an acceptable answer. The liver failure was a result of digesting too much paracetamol, also known as acetaminophen, therefore this is what has poisoned the patient and resulted in acute liver failure. However, the more generalised 'medication reaction or side effect' was not accepted as this was a more ambiguous response.

In addition, if a symptom checker failed to allow the key symptoms to be entered into the program, this resulted in an 'unassessable' vignette (Appendix C). The vignettes were also classified by whether the correct diagnosis was likely to appear in the top three options listed, based on the symptoms provided (Appendix B and Table 4.6). Some conditions such as Ross River Virus may have very similar symptoms to other disorders (mononucleosis) and it may not be reasonable to expect a symptom checker to list Ross River Virus as the most likely differential diagnosis, given it is less common. Where a symptom checker was specific to adult or paediatric patients, these symptom checkers were assessed but only the age-appropriate vignettes were included in the assessment. This resulted in 1,170 SCVT for diagnostic accuracy, covering 27 symptom checkers (Appendix C).

3.10 Triage accuracy

Assessing for triage accuracy was determined by whether the symptom checker offered triage advice concordant with our assessment; emergency, urgent, non-urgent or self-care (refer to Table 3.3 above). As with Semigran et al.'s (2015) study, if multiple triage options were presented the most urgent suggestion was utilised. The most urgent suggestion was frequently listed first, and it would be reasonable to expect a consumer to take the most risk-averse advice (Hancock, 2017). If the website or app did not diagnose, but offered triage advice, this was still included in the analysis. It has been argued that appropriate triage advice is more immediately useful to a consumer than a potential diagnosis (Middleton et al., 2016). For example, pulmonary embolism and heart attack can present with very similar symptoms and regardless of the diagnosis it is more important that emergency care is sought.

If a diagnosis needed to be selected in order to obtain triage advice, this resulted in the triage aspect of the symptom checker being deemed 'unassessable' for this study. For example, for the 'Appendicitis' vignette, Buoy Health (2018) provided triage advice ranging from 'self-care' if the diagnosis of rotavirus was selected, to 'emergency department' if appendicitis was selected. As a result, no meaningful triage advice could be recorded against symptoms provided. Since the triage advice related to the consumer's choice of diagnosis, no meaningful triage advice could be recorded against symptoms provided in the vignette.

For some vignettes, it was decided that either emergency or urgent care could be sought with equal accuracy, so either response by a symptom checker was accepted as correct (refer to Appendix B). Where a symptom checker was specific to adult or paediatric patients, these symptom checkers were assessed but only the age-appropriate vignettes were included in assessment. This resulted in 688 SCVT for triage assessments, covering 19 symptom checkers (Appendix C). Healthdirect Australia's performance, as the Australian benchmark, was compared to international websites and apps for triage only, as Healthdirect Australia does not offer diagnostic possibilities.

3.11 Summary of Resulting Study Sample

The resulting search for symptom checkers resulted in 36 symptom checkers being selected for evaluation in this research, with IP addresses operating from the USA, Israel, British Columbia, Poland and Australia. Online websites contributed 26 symptom checkers and mobile phone apps provided a further 10 symptom checkers. Only

10 symptom checkers provided both diagnostic and triage advice; 17 provided only a diagnosis and 9 were specific to triage advice. Therefore, 27 symptom checkers provided diagnostic information and 19 provided triage advice. Of the 48 vignettes 13 were classified as emergency, 14 as urgent, 11 as non-urgent and 10 for the self-care category. Common conditions comprised 41 vignettes. Diagnostic performance was based on 1170 assessable vignettes over the 27-symptom checkers providing diagnostic information. Triage performance was based on 688 assessable vignettes over the 19-symptom checkers providing triage advice (Appendix C).

3.12 Data Analysis

Summary statistics were calculated for diagnostic accuracy and triage advice with a 95% confidence interval (CI) based on binomial distribution using SPSS v25. One-way ANOVA and independent samples t-tests were used to compare diagnostic and triage accuracy by level, vignette acuity and symptom checker characteristic. Normality was tested using Shapiro-Wilks. Where data was non-normally distributed, log transformation was undertaken and found to be normally distributed thereafter. Levene's test was used for equal variance, and when unequal, Welch's ANOVA was used. A sensitivity analysis was conducted, removing four symptom checkers that never suggest self-care from triage analysis.

Analysis of diagnostic accuracy was determined for the correct differential diagnosis being listed as the 1st option; within the top 3 possible diagnoses; or within the top 10 differential diagnoses. Individual performance was expressed as a percentage for diagnostic accuracy (with the correct diagnosis listed first, within top three and ten differentials). A diagnosis was determined as incorrect after the 10th option.

Analysis of triage accuracy was calculated overall, as well as by each individual category (i.e. emergency; urgent; non-urgent; self-care) and expressed as a percentage. Healthdirect Australia's triage results were compared to the international symptom checkers for both websites and apps.

Symptom checkers which have both a website and an app had both products compared to see if advice given in both systems was identical. Comparison was also be made between common and uncommon medical conditions to see if diagnostic and triage accuracy varies between the categories.

3.13 References

Armstrong, S. (2018). The apps attempting to transfer NHS 111 online. *BMJ (clinical research ed.)*, *360,* 156. doi: 10.1136/bmj.k156. Retrieved from http://www.bmj.com/content/360/bmj.k156.abstract

Australian Institute of Health and Welfare. (2017). *Emergency department care 2016-17: Australian hospital statistics.* Retrieved from https://www.aihw.gov.au/reports/hospitals/ahs-2016-17-emergency-department-care/contents/table-of-contents

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014). Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine: Practice, Policy, Research*, *4*(4), 363–371. https://doi.org/10.1007/s13142-014-0293-9

Britt, H., Miller, G. C., Bayram, C., Henderson, J., Valenti, L., Harrison, C., … Wong, C. (2016). *A Decade of Australian General Practice Activity 2006-07 to 2015-16 (No.41). Sydney University Press.* Retrieved from https://ses.library.usyd.edu.au/bitstream/handle/2123/15482/9781743325162_ONLINE.pdf;jsessionid=6DA6EB5836FBFB5D9DA29ACA4325A4D8?sequence=5

Buoy Health. (2018). *Buoy.* Retrieved December 15, 2018, from https://www.buoyhealth.com/

Chelimsky, E. (1985) Comparing and contrasting auditing and evaluation: some notes on their relationship. Evaluation Review*, 9*(4): 483–503.

Choi, W., & Stvilia, B. (2014). How do college students choose mobile health/wellness applications? *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1–4. https://doi.org/doi:10.1002/meet.2014.14505101115

Clarke, A., & Dawson, R. (1999). *Evaluation research : an introduction to principles, methods, and practice.* London ; SAGE. Retrieved from https://methods-sagepub-com.ezproxy.ecu.edu.au/book/evaluation-research

Cooke, G., Valenti, L., Glasziou, P., & Britt, H. (2013). Common general practice presentations and publication frequency. *Australian Family Physician*, *42*(1/2), 65–68. Retrieved from http://www.racgp.org.au/afp/2013/januaryfebruary/common-general-practice-presentations/

Dore, A., & Auld, J. (2004). Barmah Forest viral exanthems. *Australasian Journal of Dermatology*, *45*(2), 125–129. https://doi.org/10.1111/j.1440-0960.2004.00062.x

eBiz. (2017). *Top 15 most popular search engines.* Retrieved April 12, 2018, from http://www.ebizmba.com/articles/search-engines

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ*, *324*(7337),

573–577. doi:10.1136/bmj.324.7337.573

Franklin, C. (2002). *How internet search engines work.* Retrieved May 02, 2019, from https://computer.howstuffworks.com/internet/basics/search-engine.htm

Hancock, D. (2017). Digital transformation in practice. *Practice Management*, *27*(4), 16–21. https://doi.org/10.12968/prma.2017.27.4.16

Healthdirect Australia. (2016a). *About healthdirect symptom checker.* Retrieved April 14, 2018, from https://www.healthdirect.gov.au/about-healthdirect-symptom-checker

Healthdirect Australia. (2016b). *Healthdirect Symptom Checker.* Retrieved October 9, 2019, from https://about.healthdirect.gov.au/healthdirect-symptom-checker

Healthdirect Australia. (2017). *Annual Report 2016 – 2017 Healthdirect Australia.* Retrieved April 14, 2018, from https://media.healthdirect.org.au/publications/Annual_Reports_Business_Highlights_2016-2017.pdf

Healthdirect Australia. (2018). *Healthdirect Australia Annual Report Business Highlights 2017-2018*. Retrieved from http://media.healthdirect.org.au/publications/healthdirect-australia-annual-report_17-18_business-highlights.pdf

Herman, J. L., Morris L. L., Fitz-Gibbon, C. T., and University of California, Los Angeles. Center for the Study of Evaluation. 1987. *Evaluator's Handbook*. Newbury Park, Calif.: Sage Publications.

Isabel Healthcare. (2019). *Isabel symptom checker.* Retrieved January 31, 2019, from https://symptomchecker.isabelhealthcare.com/private/suggest_diagnosis.jsp

Kruger, P. C., Eikelboom, J. W., Douketis, J. D., & Hankey, G. J. (2019). Deep vein thrombosis: update on diagnosis and management. *Medical Journal of Australia*, *210*(11), 516–524. https://doi.org/10.5694/mja2.50201

Lauckner, C., & Hsieh, G. (2013). The presentation of health-related search results and its impact on negative emotional outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 333–342). Paris, France: ACM.

Mayo Clinic. (2018). *Symptom Checker.* Retrieved November 15, 2018, from https://www.mayoclinic.org/symptom-checker/select-symptom/itt-20009075

McBride, W. J. H., Hanson, J. P., Miller, R., & Wenck, D. (2007). Severe Spotted Fever Group Rickettsiosis, Australia. *Emerging Infectious Diseases*, *13*(11), 1742–1744. https://doi.org/10.3201/eid1311.070099

McDonald, A. M., & Cranor, L. F. (2009). *An empirical study of how people perceive online behavioral advertising (CMU-CyLab-09-015).* Retrieved from http://repository.cmu.edu/cylab/2/

Middleton, K., Butt, M., Hammerla, N., Hamblin, S., Mehta, K., & A, P. (2016). Sorting out symptoms: design and evaluation of the "babylon check" automated triage system. *arXiv preprint arXiv:1606.02041.* Retrieved April 04, 2018, from https://arxiv.org/abs/1606.02041

Mugenda, O., & Mugenda, A. (2003). Research methods: Quantitative and Qualitative methods. *Revised in Nairobi*.

Murtagh, J., & Rosenblatt, J. (2015). *John Murtagh's general practice* (Sixth edit). North Ryde, NSW : McGraw-Hill Education.

O'Leary, Z. (2004). *The essential guide to doing research. Sage* (3rd editio). London: SAGE Publications Ltd. Retrieved from https://study.sagepub.com/oleary3e

Patton, M. (1986). *Utilization-focusd evaluation* (2nd edition). Beverly Hill: SAGE Publications.

Queensland Government. (2017). *Hendra virus infection.* Retrieved April 14, 2018, from http://conditions.health.qld.gov.au/HealthCondition/condition/14/217/363/Hendra-Virus-Infection

Semigran, H. L., Levine, D. M., Nundy, S., & Mehrotra, A. (2016). Comparison of physician and computer diagnostic accuracy. *JAMA Internal Medicine*, *176*(12), 1860–1861.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ (clinical research ed.)*, *351,* 3480. doi:10.1136/bmj.h3480

Scriven, M. (1967) *The methodology of evaluation*, in R.W. Tyler, R.M. Gagne and M. Scriven(eds), Perspectives of Curriculum Evaluation, Chicago: Rand McNally. pp. 39–83.

Sullivan, D. (2002). *How search engines work.* Retrieved May 02, 2019, from https://didattica-2000.archived.uniroma2.it//prog_web/deposito/search_engine.pdf

Symptomate. (2019). *Symptomate*. Retrieved January 8, 2019, from https://symptomate.com/diagnosis/

VIctoria State Government. (2018). *Better health channel.* Retrieved May 25, 2018, from https://www.betterhealth.vic.gov.au/

WebMD. (2017). *WebMD Symptom Checker.* Retrieved November 14, 2018, from https://symptoms.webmd.com/default.htm

Wertheim, H. F. L., Horby, P., & Woodall, J. P. (2012). *Atlas of human infectious diseases*. John Wiley & Sons.

Westhorpe, C. (2014). Ross River virus. *Geodate*, *27*(3), 4.

What's My Diagnosis. (n.d.). *What's My Diagnosis.* Retrieved January 9, 2019, from www.whatsmydiagnosis.com

Wilson, A., Tierney, L., Lai, K., & Graves, S. (2013). Queensland tick typhus: three cases with unusual clinical features. *Internal Medicine Journal*, *43*(7), 823–825. https://doi.org/10.1111/imj.12184

Young, J., Selvey, C., & Symons, R. (2011). Hendra virus. *The Medical Journal of Australia*, *195*(5), 250–251. doi:10.5694/mja11.10967

CHAPTER 4

RESULTS

Thesis Map
Chapter One – Introduction
Chapter Two – Literature Review
Chapter Three – Methodology
➡ Chapter Four – Results
Chapter Five – Discussion
Chapter Six – Conclusion

Results for diagnostics

4.1 Listing the correct diagnosis first

Overall, the correct diagnosis was listed first in 36% of patient SCVT (95% confidence interval [CI], 31—42; Table 4.1 below). Diagnostic performance varied according to patient acuity presented in the vignettes with emergency at 27% (95% CI 21—34), urgent at 45% accuracy (95% CI 39—51), non-urgent at 39% (95% CI 33—47), and self-care with 32% (95% CI 25— to 42) respectively (p = 0.004, Table 4.1). Table 4.2 depicts where significant differences lie. Table 4.7 depicts each symptom checkers accuracy for diagnosis per vignette, stratified by patient severity.

Common vignettes were listed first more frequently than uncommon vignettes (42%, 95% CI 37—50 and 4%, 95% CI 1—7, p = <0.001, Table 4.1). Non-significant differences were found between sites asking demographic questions than those that do not (38%, 95% CI 32—47 and 32%, 95% CI 26—38, p = 0.154). Sites that provide between six and ten differential diagnoses had the strongest performance with 42% accuracy (95% CI 32—58, p = 0.120). There was a significant difference between symptom checkers using AI algorithms. Sites and apps clearly stating they use AI had an accuracy of 46% (95% CI 40—57) versus 32% non-AI sites (95% CI 26—38, p = 0.007) listing the first diagnosis of correctly (Table 4.3 below). Performance was variable across individual symptom checkers, with ePain Assist Google (2018) listing the correct diagnosis first in 12% of SCVT, whilst Symptomate (2019a; 2019b; 2019c) listed the correct diagnosis first in 61% of SCVT consistently across all three of their platforms (Table 4.4 below).

Table 4.1: Accuracy of diagnosis decision and triage advice for all symptom checkers; stratified by severity of patient condition in vignette and by frequency of the condition's diagnosis

		Diagnosis									Triage		
		Listed first			Listed in top 3			Listed in top 10					
Type of vignette or diagnosis	No. of vignettes (%)	Rate*	% (95% CI†)	P value	Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value
All vignettes	48 (100%)	421/1170	36% (31 to 42)		606/1170	52% (47 to 59)		681/1170	58% (53 to 65)		338/688	49% (44 to 54)	
Type of patient condition vignette‡													
Emergency	13 (27%)	92/343	27% (21 to 34)	**0.004**	139/343	41% (33 to 50)	**<0.001**	161/343	47% (40 to 56)	**<0.001**	121/191	63% (52 to 71)	**<0.001**
Urgent	14 (29%)	163/365	45% (39 to 51)		231/365	63% (58 to 70)		245/365	67% (62 to 73)		115/206	56% (52 to 75)	
Non-urgent	11 (23%)	99/252	39% (33 to 47)		145/252	58% (51 to 67)		168/252	67% (61 to 75)		46/151	30% (11 to 39)	
Self-care	10 (21%)	67/210	32% (25 to 42)		91/210	43% (36 to 53)		107/210	51% (42 to 60)		56/140	40% (26 to 49)	
Type of diagnosis**#													
Common	41 (85%)	414/986	42% (37 to 50)	**<0.001**	593/986	60% (54 to 69)	**<0.001**	666/986	68% (61 to 76)	**<0.001**	297/587	51% (46 to 55)	**0.012**
Uncommon	7 (15%)	7/184	4% (1 to 7)		13/184	7% (3 to 12)		15/184	8% (4 to 13)		41/101	41% (22 to 45)	

† CI = confidence interval
*Number of correct vignette evaluations divided by the applicable evaluations. Some vignettes could not be applied to a given symptom checker; for example, a paediatric vignette could not be evaluated on an adult-only symptom checker. Also, some vignettes were not assessable as key signs or symptoms were unable to be entered into the program.
**Based on the BEACH (Bettering the Evaluation and Care of Health) research data for the decade 2005-06 to 2015-16 (Britt et al., 2016), the Australian Institute of Health and Welfare's Emergency Department care statistics 2017 (Australian Institute of Health and Welfare, 2017) or the Victoria State Government's website, 'Better Health Channel' (VIctoria State Government, 2018). Reviewed by two practicing General Practitioners. Uncommon vignettes represent 15% in this research, similar to the 'real-world' distribution which is reported at 10% for General Practitioners (Cooke, Valenti, Glasziou, & Britt, 2013).
‡ANOVA (Analysis of variance) analysis. Assumptions behind the ANOVA: Shapiro-Wilk test for normality, log transformation when non-normality was found, and Levene's test for equal variance with Welch's ANOVA where equal variance was violated.
#Independent sample t-test analysis

Table 4.2. Diagnostic Symptom Checkers: Depicts the mean differences between triage categories (Emergency, Urgent, Non-urgent and Self-care).				
Dependent variable	**Category**	**The triage acuity being compared to:**	**Mean Difference (95% Confidence Interval)**	***p*-value**
Correct diagnosis listed first	Emergency (vs)	Urgent	-17.556 (-30.71 to -4.04)	**0.004***
		Non-urgent	-12.889 (-26.05 to 0.27)	0.057
		Self-care	-6.111 (-19.27 to 7.05)	0.620
	Urgent (vs)	Non-urgent	4.667 (-8.49 to 17.83)	0.791
		Self-care	11.444 (-1.71 to 24.60)	0.112
	Non-urgent (vs)	Self-care	6.778 (-6.38 to 19.94)	0.537
Correct diagnosis listed in top 3 options	Emergency (vs)	Urgent	-22.370 (-36.86 to -7.88)	**0.001***
		Non-urgent	-17.556 (-32.04 to -3.07)	**0.011***
		Self-care	-3.037 (-17.52 to 11.45)	0.947
	Urgent (vs)	Non-urgent	4.815 (-9.67 to 19.30)	0.821
		Self-care	19.333 (4.85 to 33.82)	**0.004***
	Non-urgent (vs)	Self-care	14.519 (0.03 to 29.00)	**0.049***
Correct diagnosis listed in top 10 options	Emergency (vs)	Urgent	-19.852 (-34.16 to -5.55)	**0.003***
		Non-urgent	-20.259 (-34.57 to -5.95)	**0.002***
		Self-care	-3.556 (-17.86 to 10.75)	0.916
	Urgent (vs)	Non-urgent	-0.407 (-14.71 to 13.90)	1.000
		Self-care	16.296 (1.99 to 30.60)	**0.019***
	Non-urgent (vs)	Self-care	16.704 (2.40 to 31.01)	**0.015***

Analysis performed by one-way ANOVA (ANOVA = Analysis of variance)
*The mean difference is significant at the 0.05 level.
'Mean difference' refers to the difference in mean percentage correct values (Top 1, Top 3, Top 10 correct) for symptom checker's for vignettes from different triage categories

4.2 Listing the correct diagnosis in the top 3 or 10 differential options.

Across all symptom checkers the correct diagnosis was listed in the top three differentials in 52% of SCVT (95% CI 47—59) and 58% (95% CI 53—65) of SCVT when considering the top ten listed differentials (Table 4.1 above). Considering the patient acuity represented in the vignette, accuracy for emergency, urgent, non-urgent and self-care SCVT resulted in the correct diagnosis listed within the top three options in 41% (95% CI 33—50), 63% (95% CI 58—70), 58% (95% CI 51—67) and 43% (95% CI 36—53, $p = <0.001$) of SCVT respectively and within the top ten options in 47% (95% CI 40—56), 67% (95% CI 62—73), 67% (95% CI 61—75) and 51% (95% CI 42—60, $p = <0.001$) of SCVT respectively (Table 4.1 above). Table 4.2 depicts where significant differences lie.

Common vignettes outperformed uncommon vignettes, having 60% (95% CI 54—69, $p = <0.001$) accuracy for listing the correct diagnosis within the first three options versus 7% (95% CI 3—12), and 68% (95% CI 61—76, $p=<0.001$) compared to 8% (95% CI 4—13) within the top ten differential diagnoses (Table 4.1 above).

Significant differences were found in sites asking demographic information providing the correct diagnosis within the first ten differentials in 62% (95% CI 55—72, $p = 0.024$) of SCVT, as opposed to sites asking no demographic information (50% 95% CI 44—57, Table 4.3). No significant differences were determined between sites providing 1—5, 6—10 or more than 11 potential differential diagnoses in either the top three or top ten listed differential diagnoses. Symptom checkers using AI had a significant difference in performance. Listing the correct diagnosis within the top three differentials, sites utilising AI achieved 64% accuracy with SCVTs (95% CI 59—74, $p= 0.001$) compared to 47% (95% CI 40—54) for sites not using AI, and within the top ten differentials 71% (95% CI 65—80, $p= 0.002$, Table 4.3) versus 54% (95% CI 46—61).

Concerning individual symptom checker performance, the percentage of correct diagnoses listed within the top three options varied from 23% for ePain Assist Google (2018) and What's My Diagnosis (n.d.) to 77% for all three Symptomate programs (2019a; 2019b; 2019c) for SCVT accuracy. The percentage of correct diagnoses listed within the top ten differentials ranged in accuracy from 30% for What's My Diagnosis (n.d.) and Symptify (2018) to 81% for all Symptomate symptom checkers (2019a; 2019b; 2019c) for SCVTs. Refer to Table 4.4 below.

Whether the diagnosis was expected to appear within the top three options was also assessed. Of the 38 vignettes where the diagnosis was expected to appear within the top three listed differential diagnoses, 60.44% of SCVTs achieved this measure (547 of the 905 assessed vignettes; 121 vignettes were unassessable). Ten vignettes were not expected to have the correct diagnosis appear within the top three listed differentials, however 22.26% of SCVT still appeared within the top three options (59 of the 206 assessed vignettes; 5 vignettes were unassessable). Refer to Table 4.6 below.

Table 4.3: Accuracy of diagnosis given and triage advice for all symptom checkers; based on certain characteristics of the tool

		Diagnosis										Triage			
			Listed first			Listed in top 3			Listed in top 10						
Symptom checker characteristics	All Symptom checkers (websites & apps^)	No. of diagnostic symptom checkers (%)	Rate*	% (95% CI†)	P value	Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value	No. of triage symptom checkers (%)	Rate*	% (95% CI)	P value
All symptom checkers	36	27 (100%)	421/1170	36% (31 to 42)		606/1170	52% (47 to 59)		681/1170	58% (53 to 65)		19 (100%)	338/688	49% (44 to 54)	
Advises use of Artificial Intelligence Algorithms#															
Yes	8	8 (30%)	144/312	46% (40 to 57)	**0.007**	201/312	64% (59 to 74)	**0.001**	220/312	71% (65 to 80)	**0.002**	5 (26%)	88/172	52% (50 to 54)	0.247
No	28	19 (70%)	277/858	32% (26 to 38)		405/858	47% (40 to 54)		461/858	54% (46 to 61)		14 (74%)	250/516	48% (41 to 54)	
Asks demographic questions? #															
Yes	23	18 (67%)	293/770	38% (32 to 47)	0.154	421/770	55% (47 to 64)	0.105	481/770	62% (55 to 72)	**0.024**	11 (58%)	240/451	53% (49 to 57)	**0.047**
No	13	9 (33%)	128/400	32% (26 to 38)		185/400	46% (40 to 53)		200/400	50% (44 to 57)		8 (42%)	98/237	41% (34 to 51)	
Max. no. of diagnoses listed.‡															
0	9	-	-	-	-	-	-	-	-	-	-	9 (47%)	167/301	55% (48 to 58)	0.288
1-5	5	5 (18%)	83/218	38% (32 to 43)	0.120	111/218	51% (43 to 58)	0.316	111/218	52% (44 to 58)	0.305	4 (21%)	66/169	39% (11 to 65)	
6-10	8	8 (30%)	132/312	42% (32 to 58)		178/312	57% (44 to 76)		197/312	63% (49 to 82)		6 (32%)	105/218	49% (42 to 56)	
≥11	14	14 (52%)	206/640	32% (23 to 40)		317/640	50% (39 to 59)		373/640	58% (48 to 68)		-	-	-	

^ apps = applications
† CI = confidence interval
*Number of correct vignette evaluations divided by the applicable evaluations. Some vignettes could not be applied to a given symptom checker: for example, a paediatric vignette could not be evaluated on an adult-only symptom checker. Also, some vignettes were not assessable as key signs or symptoms were unable to be entered into the program.
+At least both age and gender needed to be requested for a symptom checker to be classified as asking demographic questions.
‡ANOVA Welch analysis. (ANOVA = Analysis of variance) Assumptions behind the ANOVA: Shapiro-Wilk test for normality, log transformation when non-normality was found, and Levene's test for equal variance with Welch's ANOVA where equal variance was violated.
#Independent samples t-test analysis

Results for triage

4.3 Providing the correct triage advice

The concordant triage advice was provided in 49% (95% CI 44—54, Table 4.1) of SCVT. Emergency care provided the greatest accuracy of triage information at 63% (95% CI 52—71, p = <0.001), urgent care at 56% (95% CI 52—75), non-urgent care at 30% (95% CI 11—39) and self-care with 40% (95% CI 26—49) of SCVT. Table 4.5 depicts where the significant differences lie. Table 4.8 depicts each symptom checker's accuracy for triage advice per vignette, stratified by patient severity. Appropriate triage advice was higher for common SCVTs at 51% (95% CI 46—55, p = 0.012) than uncommon vignettes 41% (95% CI 22— 45). Excluding four symptom checkers that never suggest self-care from the overall triage performance (Isabel Healthcare (2019); Everyday Health (2019); Symcat (2019); Doctor Diagnose Symptom (2018)), appropriate triage advice was given in 51% (95% CI 48—55) of vignettes.

Referring to Table 4.3 above, symptom checkers requesting demographic information provided triage advice concordant with our assessment in 53% (95% CI 49—57, p = 0.047) of SCVTs versus 41% (95% CI 34—51) for symptom checkers that did not. Symptom checkers that only provided triage advice (no diagnostic functionality), provided concordant triage information in 55% (95% CI 48—58, p = 0.288) of SCVTs. Symptom checkers where the differential diagnostic lists were between 1—5 long, provided concordant triage advice in 39% (95% CI 11—65) of instances, whereas differential lists 6—10 possible options provided concordant triage advice in 49% (95% CI 42—56) of SCVTs. Triage symptom checkers utilising AI did not yield significant differences, with sites using AI technology providing concordant triage advice in 52% (95% CI 50—54, p = 0.247) of SCVTs versus 48% (95% CI 41—54) for sites that do not use AI.

4.4 Individual symptom checkers triage advice

Triage performance varied widely across symptom checkers, with Healthdirect Australia (2016, 2018) (website and app) achieving 61% accuracy for all SCVTs whereas the Doctor Diagnose Symptom app for Google (2018) achieved merely 17% concordant with our assessment (Table 4.4). For emergency care, Isabel Healthcare's website (2019) provided the most concordant triage advice in 100% of emergency SCVTs versus 8% for the Doctor

Diagnose Symptom Google app (2018). Urgent care SCVTs had four children's symptom checkers (Healthy Children (2018); Hopkins All Children's (2019); CHW (2019); St Luke's Online (n.d.)) all providing 100% concordant triage advice as opposed to Symcat (2019) failing to provide concordant advice at all (0%). For non-urgent care, Symcat (2019) appropriately provided triage advice in 100% of instances whilst 8 symptom checkers (Healthy Children (2018); Hopkins All Children's (2019); CHW (2019); St Luke's Online (n.d.); Symptomate website, Apple and Google Play platforms (2019a; 2019b; 2019c); Doctor Diagnose Symptom (2018)) attained 0%. Healthwise (My Health Alberta, 2018) and University of Michigan Health (UofMHealth) (2019) performed the best under the self-care triage category with 70% concordance, whilst four symptom checkers never suggested self-care (Isabel Healthcare (2019); Everyday Health (2019); Symcat (2019); Doctor Diagnose Symptom (2018)) resulting in 0% concordant advice.

Table 4.4 Accuracy of diagnosis decision and triage advice for each symptom checker

	Diagnosis						Triage													
	Listed first		Listed in top 3		Listed in top 10		All cases		Emergency care required		Urgent care required		Non-urgent care reasonable		Self-care reasonable					
Symptom checker (n=36)	Rate*	%	Rate*	%	Rate*	%	Rate*	%	Rate*	%	Rate*	%	Rate*	%	Rate*	%				
AARP† Health Tools	15/48	31%	18/48	38%	19/48	40%	---	---	---	---	---	---	---	---	---	---				
APWUHP††††††	12/48	25%	17/48	35%	19/48	40%	---	---	---	---	---	---	---	---	---	---				
Buoy Health	17/48	35%	24/48	50%	24/48	50%	∇	∇	∇	∇	∇	∇	∇	∇	∇	∇				
CHW†††††	-	-	-	-	-	-	8/17	47%	1/2	50%	3/3	100%	0/5	0%	4/7	57%				
Doctor Diagnose (Google)	12/36	33%	16/36	44%	16/36	44%	6/35	17^	1/13	8%	5/13	38%	0/4	0%	0/5	0%				
Drugs.com	16/43	37%	24/43	56%	24/43	56%	22/43	51%	7/13	54%	7/14	50%	6/8	75%	2/8	25%				
Drugs.com (Google)	19/43	44%	25/43	58%	25/43	58%	23/43	53%	7/13	54%	7/14	50%	7/8	88%	2/8	25%				
ePain Assist (Apple)	6/43	14%	12/43	28%	17/43	40%	---	---	---	---	---	---	---	---	---	---				
ePain Assist (Google)	5/43	12%	10/43	23%	14/43	33%	---	---	---	---	---	---	---	---	---	---				
Everyday Health	16/31	52%	22/31	71%	22/31	71%	17/31	55%	9/11	82%	7/11	64%	1/6	17%	0/3	0%				
Family Doctor	19/48	40%	22/48	46%	22/48	46%	15/48	31%	5/13	38%	4/14	29%	5/11	45%	1/10	10%				
Healthdirect Australia	-	-	-	-	-	-	28/46	61%	9/12	75%	9/13	69%	4/11	36%	6/10	60%				
Healthdirect Aust. (Apple)	-	-	-	-	-	-	28/46	61%	9/12	75%	9/13	69%	4/11	36%	6/10	60%				
Healthline	13/48	27%	22/48	46%	29/48	60%	---	---	---	---	---	---	---	---	---	---				
Healthwise††	-	-	-	-	-	-	27/48	56%	8/13	62%	10/14	71%	2/11	18%	7/10	70%				
Health Link BC	-	-	-	-	-	-	26/48	54%	8/13	62%	10/14	71%	2/11	18%	6/10	60%				
Health Status	17/44	37%	26/44	59%	32/44	73%	---	---	---	---	---	---	---	---	---	---				
Healthy Children	-	-	-	-	-	-	8/17	47%	1/2	50%	3/3	100%	0/5	0%	4/7	57%				
Hopkinsallchildrens††††	-	-	-	-	-	-	8/17	47%	1/2	50%	3/3	100%	0/5	0%	4/7	57%				
Isabel Healthcare	16/48	33%	30/48	63%	37/48	77%	23/48	48%	13/13	100 %	7/14	50%	3/11	27 %	0/10	0%				
Mayo Clinic	16/42	38%	25/42	60%	28/42	67%	---	---	---	---	---	---	---	---	---	---				
MedicineNet	13/48	27%	26/48	54%	32/48	67%	---	---	---	---	---	---	---	---	---	---				
Patient.info	21/48	44%	27/48	56%	30/48	63%	---	---	---	---	---	---	---	---	---	---				
Right Diagnosis	6/44	14%	16/44	36%	18/44	41%	---	---	---	---	---	---	---	---	---	---				
RxList	15/48	31%	28/48	58%	34/48	71%	---	---	---	---	---	---	---	---	---	---				
St Luke's online	-	-	-	-	-	-	8/17	47%	1/2	50%	3/3	100%	0/5	0%	4/7	57%				
Symcat	17/46	37%	22/46	48%	30/46	65%	17/46	40%	7/13	54%	0/14	0%	10/10	100%	0/9	0%				
Symptify (Google)	11/46	24%	14/46	30%	14/46	30%	∇	∇	∇	∇	∇	∇	∇	∇	∇	∇				
Symptomate	19/31	61%	24/31	77%	25/31	81%	16/31	52%	9/11	82%	6/11	55%	0/6	0%	1/3	33%				
Symptomate (Apple)	19/31	61%	24/31	77%	25/31	81%	16/31	52%	9/11	82%	6/11	55%	0/6	0%	1/3	33%				
Symptomate (Google)	19/31	61%	24/31	77%	25/31	81%	16/31	52%	9/11	82%	6/11	55%	0/6	0%	1/3	33%				
UofMHealth†††	-	-	-	-	-	-	26/45	58%	7/11	64%	10/13	77%	2/11	18%	7/10	70%				
WebMD	25/47	53%	33/47	70%	36/47	77%	---	---	---	---	---	---	---	---	---	---				
WebMD (Apple)	24/47	51%	32/47	68%	35/47	74%	---	---	---	---	---	---	---	---	---	---				
WebMD (Google)	25/47	53%	33/47	70%	36/47	77%	---	---	---	---	---	--	---	---	---	---				
What's My Diagnosis	8/43	19%	10/43	23%	13/43	30%	---	---	---	---	---	---	---	---	---	---				

* Number of correct vignette evaluations divided by the applicable evaluations. Some vignettes could not be applied to a given symptom checker: for example, a paediatric vignette could not be evaluated on an adult-only symptom checker. Also, some vignettes were not assessable as key signs and symptoms were unable to be entered into the program.
- Symptom checker does not provide diagnosis.
--- Symptom checker does not provide triage advice.
∇ Triage advice was not assessed. The triage advice given varied according to each potential differential diagnosis. (i.e. Buoy Health for the Appendicitis Vignette, differentials varied from rotavirus = selfcare to appendicitis = ED.)
†AARP Healthtools = American Association of Retired Persons Healthtools
†† Healthwise = Healthwise (My Health Alberta)
††† UofMHealth = University of Michigan Health
†††† Hopkinsallchildrens = John Hopkins Children's Hospital
††††† CHW = Children's Hospital of Wisconsin
†††††† APWUHP = American Postal Workers Union Health Plan

Other Results

4.5 Discrepancies between platforms

Drugs.com (2019a, 2019b), ePain Assist (2018a, 2018b) and WebMD (2017, 2018a, 2018b) had some variation in their results between platforms—website, Apple Store apps and Google Play apps (Table 4.4 above). All three companies had discrepancies in diagnostic listings, and Drugs.com (2019a, 2019b) also had differences in triage advice. There is no platform consistently providing better results.

Table 4.5: Triage Symptom Checkers:
Depicts the mean differences between triage categories
(Emergency, Urgent, Non-urgent and Self-care).

Category	The triage acuity being compared to:	Mean Difference (95% Confidence Interval)	*p*-value
Emergency (vs)	Urgent	-1.526 (-23.97 to 20.92)	0.998
	Non-urgent	36.632 (14.19 to 59.08)	**<0.001***
	Self-care	24.579 (2.13 to 47.03)	**0.026***
Urgent (vs)	Non-urgent	38.158 (15.71 to 60.60)	**<0.001***
	Self-care	26.105 (3.66 to 48.55)	**0.016***
Non-urgent	Self-care	-12.053 (-34.50 to 10.39)	0.496

Analysis performed by one-way ANOVA (ANOVA = Analysis of variance)
*The mean difference is significant at the 0.05 level.
'Mean difference' refers to the difference in mean percentage correct values (Top 1, Top 3, Top 10 correct) for symptom checker's for vignettes from different triage categories.

Table 4.6 - Did the correct diagnosis appear in the top three options?

		AARP Health Tools	Apwuhp (APWU Health Plan)	Bouy Health	Doctor Diagnose (Google Play)	Drugs.com	Drugs.com (Google Play)	ePain Assist (Apple App)	ePain Assist (Google Play)	Everyday Health	Family Doctor	Healthline	Healthstatus	Isabel	Mayo Clinic	MedicineNet	Patient.info	Right Diagnosis	Rxlist	Symcat	Symptify (Google Play)	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	WebMD	WebMD (Apple App)	WebMD (Google Play)	Whats My Diagnosis
Patient Vignettes	**Diagnosis not expected in top 3 options**																											
	Acute liver failure	No	No	Yes	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No
	Haemolytic uremic syndrome	No	No	No	No	No	No	No	No	/	No	No	No	No	No	No	No	No	No	No	No	/	/	/	No	No	No	No
	Malaria	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
	Hendra Virus	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
	Back strain- foot drop	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Cellulitis	No	No	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No	No	No	No
	Queensland tick typhus	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No	No	No	No
	Ross River Virus	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No
	Shingles	No	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	Stinging nettle	No	No	No	/	No	No	No	No	No	Yes	No	No	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No

Correct diagnosis was in top 3 options	Correct diagnosis did not appear in top 3	/ Vignette unassessable

Patient Vignettes	Diagnosis expected in top 3 options	AARP Health Tools	Apwuhp (APWU Health Plan)	Bouy Health	Doctor Diagnose (Google Play)	Drugs.com	Drugs.com (Google Play)	ePain Assist (Apple App)	ePain Assist (Google Play)	Everyday Health	Family Doctor	Healthline	Healthstatus	Isabel	Mayo Clinic	MedicineNet	Patient.info	Right Diagnosis	Rxlist	Symcat	Symptify (Google Play)	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	WebMD	WebMD (Apple App)	WebMD (Google Play)	Whats My Diagnosis
	Appendicitis									/												/	/	/				
	Asthma																											
	COPD exacerbation																											
	Myocardial infarction																											
	Ureteric stones																											
	Meningitis																											
	Pulmonary embolism																											
	Stroke																											
	Tetanus																											
	Deep vein thrombosis																											
	Pneumonia																											
	Acute bronchitis																											
	Acute otitis media									/												/	/	/				
	Acute sinusitis																											/
	Bowel cancer																											
	Infectious mononucleosis									/												/	/	/				
	Migraine																											
	Tonsilitis									/												/	/	/				
	Peptic ulcer disease																											
	Allergic rhinitis																											
	Apthous ulcers				/	/	/			/					/							/	/	/				/
	Back pain -unremarkable				/																							
	Bee sting without anaphylaxis				/	/	/			/			/		/			/		/	/	/	/	/				
	Bursitis				/																							
	Eczema				/					/												/	/	/				
	Molluscum contagiosum				/					/												/	/	/				
	Solar keratosis				/	/	/	/	/				/		/													/
	Sprained ankle									/												/	/	/				
	Viral upper respiritory																											
	Blue bottle jellyfish sting				/	/	/	/	/	/			/					/		/	/	/	/	/	/	/	/	
	Conjunctivitis									/								/				/	/	/				/
	Cradle cap				/			/	/	/												/	/	/				
	Dysmenorrhoea									/												/	/	/				
	Head lice				/			/	/	/					/							/	/	/				
	Herpes simplex virus type 1				/	/	/			/					/			/				/	/	/				/
	Planter warts												/															
	Threadworm							/	/	/					/							/	/	/				
	Tinea pedis																											

Correct diagnosis was in top 3 options

Correct diagnosis did not appear in top 3 options

/ Vignette unassessable

Table 4.7 - Was the diagnosis listed 1st, within top 3 or within top 10 options for each Symptom Checker? Stratified by severity of the patient vignette.

		AARP Health Tools	Apwuhp (APWU Health Plan)	Bouy Health	Doctor Diagnose (Google Play)	Drugs.com	Drugs.com (Google Play)	ePain Assist (Apple App)	ePain Assist (Google Play)	Everyday Health	Family Doctor	Healthline	Healthstatus	Isabel	Mayo Clinic	MedicineNet	Patient.info	Right Diagnosis	Rxlist	Symcat	Symptify (Google Play)	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	WebMD	WebMD (Apple App)	WebMD (Google Play)	Whats My Diagnosis
Patient Vignettes	**Requires Emergency Care**																											
	Acute liver failure	Incorrect	Incorrect	Top 10	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	1st	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	1st	1st	1st	Top 10	Top 10	Top 10	Incorrect
	Appendicitis	Top 10	Top 10	1st	Top 10	1st	1st	1st	1st	/	1st	Top 10	1st	1st	Top 10	Incorrect	1st	Top 10	Incorrect	Top 10	Incorrect	/	/	/	Top 10	Top 10	Top 10	Incorrect
	Asthma	1st	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	1st	1st	1st	1st	Top 10	1st	1st	1st	1st	1st	Incorrect	1st	Top 10	Incorrect	1st	1st	1st	1st	1st	1st	Top 10
	COPD exacerbation	Incorrect	Incorrect	Incorrect	Top 10	Incorrect	Incorrect	Top 10	Incorrect	Top 10	1st	Top 10	Incorrect	Top 10	1st	Top 10	Top 10	1st	Top 10	1st	Incorrect	1st	1st	1st	1st	1st	1st	Incorrect
	Haemolytic uremic syndrome	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	/	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	/	/	/	Incorrect	Incorrect	Incorrect	Incorrect
	Myocardial infarction	Incorrect	Incorrect	1st	1st	1st	1st	Incorrect	Incorrect	Top 10	1st	1st	Top 10	1st	Top 10	Top 10	Incorrect	Incorrect	Top 10	Incorrect	Incorrect	Top 10	Top 10	Top 10	Top 10	Top 10	Top 10	Incorrect
	Ureteric stones	Incorrect	1st	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	1st	Incorrect	Top 10	1st	Top 10	Incorrect	Top 10	Incorrect	Top 10	Top 10	Top 10	Incorrect	1st	1st	1st	1st	1st	1st	Top 10
	Malaria	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect
	Meningitis	1st	1st	Incorrect	1st	Top 10	Top 10	Top 10	Top 10	Top 10	1st	Incorrect	1st	1st	Top 10	1st	Incorrect	1st	1st	1st	Incorrect	Top 10	Top 10	Top 10	Top 10	Top 10	Top 10	Incorrect
	Pulmonary embolism	1st	1st	Incorrect	1st	Incorrect	Incorrect	Incorrect	Incorrect	1st	Incorrect	Incorrect	1st	Top 10	Top 10	Top 10	Incorrect	Incorrect	Top 10	Incorrect	1st	1st	1st	1st	1st	1st	1st	Incorrect
	Stroke	Incorrect	Incorrect	Incorrect	1st	1st	1st	Incorrect	Incorrect	1st	Incorrect	Top 10	Top 10	1st	Incorrect	Top 10	1st	Incorrect	Top 10	Top 10	Incorrect	1st	1st	1st	Top 10	Top 10	Top 10	Incorrect
	Tetanus	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Top 10	1st	Top 10	Incorrect	Incorrect	Incorrect	1st	Incorrect	Incorrect	Incorrect	1st	1st	1st	Incorrect	Incorrect	Incorrect	Incorrect
	Hendra Virus	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect
	Requires Urgent Care																											
	Deep vein thrombosis	1st	Incorrect	Incorrect	1st	Incorrect	Incorrect	Incorrect	Incorrect	Top 10	1st	1st	1st	Top 10	Top 10	1st	1st	Top 10	Top 10	1st	Incorrect	Top 10	Top 10	Top 10	1st	1st	1st	Incorrect
	Pneumonia	Top 10	Top 10	1st	Incorrect	Incorrect	Incorrect	Top 10	Incorrect	1st	Top 10	Top 10	Top 10	Top 10	Top 10	1st	Top 10	Incorrect	1st	Top 10	Incorrect	1st	1st	1st	1st	1st	1st	1st
	Acute bronchitis	1st	1st	1st	1st	1st	1st	1st	Incorrect	1st	1st	Top 10	Top 10	Top 10	1st	1st	Top 10	Top 10	1st	1st	1st	1st	1st	1st	Top 10	Top 10	Top 10	1st
	Acute otitis media	1st	Incorrect	1st	1st	1st	1st	Top 10	Top 10	/	1st	1st	1st	1st	1st	Top 10	1st	1st	Top 10	1st	1st	/	/	/	1st	1st	1st	1st
	Acute sinusitis	Incorrect	Top 10	Incorrect	1st	1st	1st	Incorrect	Incorrect	1st	1st	Incorrect	Top 10	1st	1st	1st	Top 10	Top 10	1st	Top 10	Incorrect	1st	1st	1st	1st	1st	1st	~
	Back strain- foot drop	Top 10	Top 10	Top 10	Incorrect	Top 10	Top 10	Top 10	Top 10	1st	1st	1st	Top 10	Top 10	1st	Top 10	1st	Incorrect	Top 10	1st	Top 10	1st	1st	1st	1st	1st	1st	Top 10
	Bowel cancer	1st	Top 10	1st	1st	1st	1st	1st	1st	1st	Top 10	Top 10	1st	Top 10	1st	Top 10	1st	Top 10	Top 10	Top 10	Incorrect	1st	1st	1st	Top 10	Top 10	Top 10	Top 10
	Cellulitis	Incorrect	Incorrect	1st	Incorrect	Top 10	Top 10	1st	1st	Incorrect	Incorrect	Incorrect	Top 10	Top 10	Top 10	Incorrect	Top 10	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect
	Infectious mononucleosis	Incorrect	Incorrect	1st	1st	Top 10	1st	Top 10	Top 10	/	Incorrect	1st	1st	Top 10	1st	1st	1st	Top 10	1st	1st	Incorrect	/	/	/	1st	1st	1st	1st
	Migraine	1st	1st	Top 10	Incorrect	1st	1st	Incorrect	Incorrect	1st	Incorrect	1st	1st	1st	1st	1st	1st	Top 10	1st	1st	Incorrect	1st	1st	1st	1st	1st	1st	Incorrect
	Queensland tick typhus	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	1st	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect
	Ross River Virus	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Incorrect	Top 10	Top 10	Top 10	Incorrect	Incorrect	Incorrect	Incorrect
	Shingles	Incorrect	Incorrect	Incorrect	Incorrect	Top 10	1st	Incorrect	Incorrect	1st	Incorrect	Incorrect	1st	Top 10	Incorrect	Incorrect	1st	Incorrect	Incorrect	Top 10	1st	Top 10	Top 10	Top 10	1st	1st	1st	Incorrect
	Tonsilitis	1st	1st	1st	1st	1st	1st	Incorrect	Incorrect	/	1st	1st	1st	1st	1st	1st	1st	Incorrect	1st	1st	Incorrect	/	/	/	1st	1st	1st	1st

Key	
1st	Correct diagnosis listed first
Top 10	Correct diagnosis cited in top 10
Incorrect	Incorrect diagnosis (diagnosis not within first 10 options)
/	Process unable to be started (ie: age restrictions)
~	Process unable to be started (ie: could not enter key s/s)

		AARP Health Tools	Apwuhp (APWU Health Plan)	Bouy Health	Doctor Diagnose (Google Play)	Drugs.com	Drugs.com (Google Play)	ePain Assist (Apple App)	ePain Assist (Google Play)	Everyday Health	Family Doctor	Healthline	Healthstatus	Isabel	Mayo Clinic	MedicineNet	Patient.info	Right Diagnosis	Rxlist	Symcat	Symptify (Google Play)	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	WebMD	WebMD (Apple App)	WebMD (Google Play)	Whats My Diagnosis
Patient Vignettes	**Requires Non-urgent Care**																											
	Peptic ulcer disease																											
	Allergic rhinitis																											
	Apthous ulcers				~	~	~			/					~							/	/	/				~
	Back pain -unremarkable				~																							
	Bee sting without anaphylaxis				~	~	~			/			~		~			~		~	~	/	/	/				
	Bursitis				~																							
	Eczema				~					/												/	/	/				
	Molluscum contagiosum				~					/												/	/	/				
	Solar keratosis				~	~	~	~	~				~		~													~
	Sprained ankle									/												/	/	/				
	Viral upper respiritory																											
	Requires Self Care																											
	Blue bottle jellyfish sting				~	~	~	~	~	/			~					~		~	~	/	/	/	~	~	~	
	Conjunctivitis									/								~				/	/	/				~
	Cradle cap				~			~	~	/												/	/	/				
	Dysmenorrhoea									/												/	/	/				
	Head lice				~			~	~	/					~							/	/	/				
	Herpes simplex virus type 1				~	~	~			/					~			~				/	/	/				~
	Planter warts												~															
	Stinging nettle				~																							
	Threadworm							~	~	/					~							/	/	/				
	Tinea pedis																											

Legend:
- Correct diagnosis listed first
- Correct diagnosis cited in top 10
- Incorrect diagnosis (diagnosis not within first 10 options)
- / Process unable to be started (ie: age restrictions)
- ~ Process unable to be started (ie: could not enter key s/s)

Table 4.8 - Triage advice for each Symptom Checker. Stratified by severity of the patient vignette.

Patient Vignettes		CHW (Children's Hospital of Winsconsin)	Doctor Diagnose	Drugs.com	Drugs.com (Google Play)	Everyday Health	Family Doctor	Healthdirect	Healthdirect (Apple App)	Healthlink BC	Healthwise (My Health Alberta)	Healthy Children	Hopkins all Childrens	Isabel	St Lukes Online	Symcat	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	UofMHealth (University of Michigan)
	Requires Emergency Care																			
*emergency or urgent	Acute liver failure	/	Unable to assess	Unable to assess	Unable to assess	Emergency	Self-care	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Non-urgent	Emergency	Emergency	Emergency	Emergency
	Appendicitis	Emergency	Urgent	Unable to assess	Unable to assess	/	Emergency	Emergency	Emergency	Urgent	Urgent	Emergency	Emergency	Emergency	Emergency	Emergency	/	/	/	Urgent
	Asthma	/	Non-urgent	Emergency	Emergency	Emergency	Urgent	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Emergency	Emergency	Emergency	Emergency	Emergency
*emergency or urgent	COPD exacerbation	/	Urgent	Emergency	Emergency	Urgent	Urgent	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Emergency	Urgent	Urgent	Urgent	Emergency
	Haemolytic uremic syndrome	Urgent	Unable to assess	Non-urgent	Non-urgent	/	Non-urgent	Emergency	Emergency	Urgent	Urgent	Urgent	Urgent	Emergency	Urgent	Emergency	/	/	/	~
	Myocardial infarction	/	Urgent	Emergency	Emergency	Emergency	Emergency	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Emergency	Emergency	Emergency	Emergency	Emergency
	Ureteric stones	/	Urgent	Emergency	Emergency	Urgent	Emergency	Urgent	Urgent	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Malaria	/	Non-urgent	Non-urgent	Non-urgent	Emergency	Non-urgent	Non-urgent	Non-urgent	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Emergency	Emergency	Emergency	Urgent
	Meningitis	/	Urgent	Urgent	Urgent	Emergency	Emergency	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Emergency	Emergency	Emergency	Emergency	Emergency
	Pulmonary embolism	/	Urgent	Emergency	Emergency	Emergency	Urgent	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Emergency	Emergency	Emergency	Emergency	~
	Stroke	/	Urgent	Emergency	Emergency	Emergency	Non-urgent	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Non-urgent	Emergency	Emergency	Emergency	Emergency
	Tetanus	/	Non-urgent	Emergency	Emergency	Urgent	Urgent	~	~	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Emergency	Emergency	Emergency	Urgent
	Hendra Virus	/	Urgent	Urgent	Urgent	Emergency	Non-urgent	Urgent	Urgent	Emergency	Emergency	/	/	Emergency	/	Urgent	Urgent	Urgent	Urgent	Emergency
	Requires Urgent Care																			
	Deep vein thrombosis	/	Urgent	Urgent	Urgent	Emergency	Emergency	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Non-urgent	Emergency	Emergency	Emergency	~
	Pneumonia	/	Unable to assess	Self-care	Self-care	Urgent	Non-urgent	Urgent	Urgent	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Acute bronchitis	/	Unable to assess	Non-urgent	Non-urgent	Urgent	Self-care	Urgent	Urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Acute otitis media	Urgent	Unable to assess	Urgent	Urgent	/	Self-care	Non-urgent	Non-urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Emergency	/	/	/	Urgent
	Acute sinusitis	/	Unable to assess	Non-urgent	Non-urgent	Non-urgent	Non-urgent	~	~	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Self-care	Self-care	Self-care	Urgent
	Back strain- foot drop	/	~	Non-urgent	Non-urgent	Urgent	Emergency	Emergency	Emergency	Emergency	Emergency	/	/	Emergency	/	Non-urgent	Urgent	Urgent	Urgent	Emergency
	Bowel cancer	/	Urgent	Non-urgent	Non-urgent	Urgent	Non-urgent	Urgent	Urgent	Non-urgent	Non-urgent	/	/	Non-urgent	/	Emergency	Urgent	Urgent	Urgent	Non-urgent
	Cellulitis	/	Urgent	Urgent	Urgent	Emergency	Emergency	Urgent	Urgent	Urgent	Urgent	/	/	Emergency	/	Emergency	Emergency	Emergency	Emergency	Urgent
	Infectious mononucleosis	Urgent	Unable to assess	Urgent	Urgent	/	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Emergency	/	/	/	Urgent
	Migraine	/	Urgent	Non-urgent	Non-urgent	Urgent	Emergency	Urgent	Urgent	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Queensland tick typhus	/	Unable to assess	Urgent	Urgent	Urgent	Non-urgent	Urgent	Urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Ross River Virus	/	Unable to assess	Urgent	Urgent	Emergency	Urgent	Urgent	Urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Non-urgent	Non-urgent	Non-urgent	Urgent
	Shingles	/	Urgent	Non-urgent	Non-urgent	Urgent	Urgent	Non-urgent	Non-urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Self-care	Self-care	Self-care	Urgent
	Tonsilitis	Urgent	Unable to assess	Urgent	Urgent	/	Urgent	Urgent	Urgent	Self-care	Self-care	Urgent	Urgent	Emergency	Urgent	Non-urgent	/	/	/	Self-care

Emergency care | Urgent Care | Non-urgent Care | Self-care | / Process unable to be started (ie: age restrictions) | Unable to assess (ie: triage advice not given) | ~ Unable to assess; unable to enter key signs/symptoms

Patient Vignettes		CHW (Children's Hospital of Winsconsin)	Doctor Diagnose	Drugs.com	Drugs.com (Google Play)	Everyday Health	Family Doctor	Healthdirect	Healthdirect (Apple App)	Healthlink BC	Healthwise (My Health Alberta)	Healthy Children	Hopkins all Childrens	Isabel	St Lukes Online	Symcat	Symptomate	Symptomate (Apple App)	Symptomate (Google Play)	UofMHealth (University of Michigan)
	Requires Non-urgent Care																			
Non-urgent Care	Peptic ulcer disease	/	Unable to assess	Non-urgent	Non-urgent	Urgent	Urgent	Self-care	Self-care	Non-urgent	Non-urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Non-urgent
Non-urgent Care	Allergic rhinitis	/	Unable to assess	Non-urgent	Non-urgent	Urgent	Non-urgent	Self-care	Self-care	Self-care	Self-care	/	/	Non-urgent	/	Non-urgent	Urgent	Urgent	Urgent	Self-care
Non-urgent Care	Apthous ulcers	Self-care	~	~	~	/	Non-urgent	Non-urgent	Non-urgent	Urgent	Urgent	Self-care	Self-care	Non-urgent	Self-care	Non-urgent	/	/	/	Urgent
Non-urgent Care	Back pain -unremarkable	/	~	Non-urgent	Non-urgent	Urgent	Non-urgent	Self-care	Self-care	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
Non-urgent Care	Bee sting without anaphylaxis	Self-care	~	~	~	/	Urgent	Emergency	Emergency	Urgent	Urgent	Self-care	Self-care	Emergency	Self-care	~	/	/	/	Urgent
Non-urgent Care	Bursitis	/	~	Non-urgent	Non-urgent	Urgent	Urgent	Non-urgent	Non-urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
Non-urgent Care	Eczema	Urgent	~	Non-urgent	Non-urgent	/	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Emergency	Urgent	Non-urgent	/	/	/	Urgent
Non-urgent Care	Molluscum contagiosum	Urgent	~	Non-urgent	Non-urgent	/	Non-urgent	Non-urgent	Non-urgent	Self-care	Self-care	Urgent	Urgent	Emergency	Urgent	Non-urgent	/	/	/	Self-care
Non-urgent Care	Solar keratosis	/	~	~	~	Non-urgent	Urgent	Non-urgent	Non-urgent	Non-urgent	Non-urgent	/	/	Non-urgent	/	Non-urgent	Urgent	Urgent	Urgent	Non-urgent
Non-urgent Care	Sprained ankle	Urgent	Urgent	Non-urgent	Non-urgent	/	Non-urgent	Self-care	Self-care	Urgent	Urgent	Urgent	Urgent	Emergency	Urgent	Non-urgent	/	/	/	Urgent
Non-urgent Care	Viral upper respiritory	/	Unable to assess	Self-care	Self-care	Urgent	Urgent	Self-care	Self-care	Urgent	Urgent	/	/	Emergency	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
	Requires Self Care																			
Self-care	Blue bottle jellyfish sting	Self-care	~	~	~	/	Urgent	Non-urgent	Non-urgent	Self-care	Self-care	Self-care	Self-care	Urgent	Self-care	~	/	/	/	Self-care
Self-care	Conjunctivitis	Urgent	Non-urgent	Urgent	Urgent	/	Unable to assess	Non-urgent	Non-urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Urgent	Non-urgent	/	/	/	Urgent
Self-care	Cradle cap	Self-care	~	Self-care	Self-care	/	Urgent	Self-care	Self-care	Self-care	Self-care	Self-care	Self-care	Urgent	Self-care	Non-urgent	/	/	/	Self-care
Self-care	Dysmenorrhoea	Self-care	Unable to assess	Self-care	Self-care	/	Urgent	Self-care	Self-care	Self-care	Self-care	Self-care	Self-care	Urgent	Self-care	Non-urgent	/	/	/	Self-care
Self-care	Head lice	Self-care	~	Non-urgent	Non-urgent	/	Urgent	Self-care	Self-care	Self-care	Self-care	Self-care	Self-care	Urgent	Self-care	Non-urgent	/	/	/	Self-care
Self-care	Herpes simplex virus type 1	Urgent	~	~	~	/	Self-care	Non-urgent	Non-urgent	Self-care	Self-care	Urgent	Urgent	Urgent	Urgent	Non-urgent	/	/	/	Self-care
Self-care	Planter warts	/	Unable to assess	Emergency	Non-urgent	Non-urgent	Non-urgent	Self-care	Self-care	Non-urgent	Non-urgent	/	/	Emergency	/	Non-urgent	Self-care	Self-care	Self-care	Non-urgent
Self-care	Stinging nettle	/	~	Non-urgent	Non-urgent	Urgent	Non-urgent	Non-urgent	Non-urgent	Urgent	Urgent	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Urgent
Self-care	Threadworm	Non-urgent	Non-urgent	Non-urgent	Non-urgent	/	Urgent	Self-care	Self-care	Self-care	Self-care	Non-urgent	Non-urgent	Urgent	Non-urgent	Non-urgent	/	/	/	Self-care
Self-care	Tinea pedis	/	Non-urgent	Non-urgent	Non-urgent	Urgent	Non-urgent	Self-care	Self-care	Self-care	Self-care	/	/	Urgent	/	Non-urgent	Urgent	Urgent	Urgent	Self-care

Legend: Emergency care (red) · Urgent Care (orange) · Non-urgent Care (green) · Self-care (blue) · / Process unable to be started (ie: age restrictions) · Unable to assess (ie: triage advice not given) (black) · ~ Unable to assess; unable to enter key signs/symptoms

4.6 References

Australian Institute of Health and Welfare. (2017). *Emergency department care 2016-17: Australian hospital statistics.* Retrieved from https://www.aihw.gov.au/reports/hospitals/ahs-2016-17-emergency-department-care/contents/table-of-contents

Britt, H., Miller, G. C., Bayram, C., Henderson, J., Valenti, L., Harrison, C., … Wong, C. (2016). *A Decade of Australian General Practice Activity 2006-07 to 2015-16 (No.41).* Sydney University Press. Retrieved from https://ses.library.usyd.edu.au/bitstream/handle/2123/15482/9781743325162_ONLINE.pdf;jsessionid=6DA6EB5836FBFB5D9DA29ACA4325A4D8?sequence=5

Children's Wisconsin. (2019). *Symptom Checker.* Retrieved January 7, 2019, from https://chw.org/health-information/symptom-checker

Cooke, G., Valenti, L., Glasziou, P., & Britt, H. (2013). Common general practice presentations and publication frequency. *Australian Family Physician, 42*(1/2), 65–68. Retrieved from http://www.racgp.org.au/afp/2013/januaryfebruary/common-general-practice-presentations/

Doctor Diagnose Symptom. (2018). *Doctor Diagnose Symptoms Check.* (Version 1.0.5) [Mobile software application]. Retrieved January 17, 2019, from https://play.google.com/store/apps/details?id=com.appcolliders.doctordiagnose

Drugs.com. (2019a). *Symptom Checker.* Retrieved January 5, 2019, from https://www.drugs.com/symptom-checker/

Drugs.com. (2019b). *Symptom Checker.* (Version 2.9.3) [Mobile application software]. Retrieved January 5, 2019, from https://play.google.com/store/apps/details?id=com.drugscom.app

ePain Assist. (2018a). *Symptom Checker.* (Version 1.5.5) [Mobile software application]. Retrieved December 7, 2018, from https://play.google.com/store/apps/details?id=com.epainassist.symptomchecker

ePain Assist. (2018b). *Symptom Checker.* (Version 1.5.5) [Mobile application software]. Retrieved December 7, 2018, from https://apple.com/au/ios/app-store

Everyday Health. (2019). *Symptom Checker.* Retrieved January 8, 2019, from https://www.everydayhealth.com/symptom-checker/

Healthdirect Australia. (2016). *Healthdirect Symptom Checker.* Retrieved October 9, 2019, from https://about.healthdirect.gov.au/healthdirect-symptom-checker

Healthdirect Australia. (2018). *Check Your Symptoms.* (Version 2.2.1) [Mobile software application]. Retrieved December 18, 2018, from https://apps.apple.com/au/app/healthdirect/id1021494621

Healthy Children. (2018). *KidsDoc Symptom Checker.* Retrieved November 25, 2018, from https://www.healthychildren.org/english/tips-tools/symptom-checker/Pages/default.aspx

Isabel Healthcare. (2019). *Isabel symptom checker.* Retrieved January 31, 2019, from https://symptomchecker.isabelhealthcare.com/private/suggest_diagnosis.jsp

Johns Hopkins All Chilren's Hospital. (2019). *Symptom Checker.* Retrieved January 6, 2019, from https://www.hopkinsallchildrens.org/Patients-Families/Patient-Family-Resources/Symptom-Checker#!/index/child/body

My Health Alberta. (2018). *Check Your Symptoms.* Retrieved December 12, 2018, from https://myhealth.alberta.ca/Health/pages/symptom-checker.aspx

St. Luke's Online. (n.d.). *Symptom Checker.* Retrieved January 12, 2019, from https://www.stlukesonline.org/symptom-checker

Symcat. (2019). *What is bothering you today?* Retrieved January 14, 2019, from https://www.symcat.com

Symptify. (2018). *How are you feeling today?* (Version 2) [Mobile application software]. Retrieved January 3, 2019, from https://play.google.com/store

Symptomate. (2019a). *Symptomate.* Retrieved January 8, 2019, from https://symptomate.com/diagnosis/

Symptomate. (2019b). *Symptomate.* (Version 2.1) [Mobile application software]. Retrieved January 8, 2019, from https://apps.apple.com/au/app/symptomate/id837725433

Symptomate. (2019c). *Symptomate.* (Version 2) [Mobile application software]. Retrieved January 8, 2019, from https://play.google.com/store/apps/details?id=com.symptomate.mobile

University of Michigan. (2019). *Symptom Checker*. Retrieved January 4, 2019, from http://www.uofmhealth.org/health-library/sx

VIctoria State Government. (2018). *Better health channel.* Retrieved May 25, 2018, from https://www.betterhealth.vic.gov.au/

WebMD. (2017). *WebMD Symptom Checker*. Retrieved November 14, 2018, from https://symptoms.webmd.com/default.htm

WebMD. (2018a). *WebMD Symptom Checker.* (Version 7.4) [Mobile application software]. Retrieved November 14, 2018, from https://apps.apple.com/au/app/webmd-symptoms-doctors-rx/id295076329

WebMD. (2018b). *WebMD Symptom Checker.* (Version 7.0) [Mobile application software]. Retrieved November 14, 2018, from https://play.google.com/store/apps/details?id=com.webmd.android

What's My Diagnosis. (n.d.). *What's My Diagnosis*. Retrieved January 9, 2019, from www.whatsmydiagnosis.com

CHAPTER 5

DISCUSSION

Thesis Map
Chapter One – Introduction
Chapter Two – Literature Review
Chapter Three – Methodology
Chapter Four – Results
➡ Chapter Five – Discussion
▪ Overall performance – Diagnosis
▪ Overall performance – Triage
▪ Symptom checker user interface
▪ Nursing triage call centres and relationship with symptom checkers
▪ Emerging trends
▪ Regulation of symptom checkers and health-related sites
▪ Other considerations
▪ Methodological differences between this research and previous research
▪ Strengths and limitations of this research
Chapter Six – Conclusion

This research aimed to ascertain the diagnostic and triage performance of free online symptom checkers readily available to the Australian public. A total of 26 websites and 10 smartphone apps were included in this evaluative research, with the Australian Government funded Healthdirect Australia being the only Australian symptom checker. Healthdirect Australia's triage symptom checker was used as the 'gold standard' and compared to other international triage symptom checkers. Diagnostic and triage accuracy was assessed using clinical vignettes; 30 vignettes were utilised in previous research (Semigran, Linder, Gidengil, & Mehrotra, 2015), with some new scenarios created to test Australian specific conditions. The results are discussed below with implications and considerations regarding these findings explored.

5.1 Overall performance – Diagnosis

5.1.1 Accuracy of symptom checkers

The diagnostic performance for the 36-symptom checkers was assessed using 48 clinical vignettes. Diagnostic accuracy was defined by whether the correct differential diagnosis was listed first, within the top three options or within the top 10 differential diagnoses options. If the correct diagnosis did not appear or was listed outside of the top ten options, it was assessed as incorrect. Symptom checkers continue to suffer shortcomings in diagnostic accuracy as was found in previous research (Semigran et al., 2015). Averaging results across all symptom checkers, the correct diagnosis was provided first only 36% of the time (refer to Table 4.1). This figure increased

to 52% when reviewing the top three differential diagnoses, and then increased further to 58% when including the top ten differential diagnoses.

These results are similar to previous research in the USA which found 34% accuracy with the first listed diagnosis, 51% for the inclusion of the correct diagnosis within the top three differential diagnoses and 58% for reviewing the top 20 differential diagnoses (Semigran et al., 2015).

A major difference between this research and previous research (Semigran et al., 2015) was the separation of triage acuity into four classifications—emergency care, urgent care, non-urgent care and self-care—to more accurately reflect the Australian healthcare system, contrasting with Semigran et al.'s (2015) emergent, non-emergent and self-care classifications. Table 5.1 below shows the results from both studies for the first listed diagnosis.

Table 5.1: Comparison of diagnosis results between studies

This research				Semigran et al.'s (2015) research			
	First listed diagnosis	Top three options	Top ten options		First listed diagnosis	Top three options	Top ten options
Overall combined performance of all 4 categories	36%	52%	58%	Overall combined performance of all 3 categories	34%	51%	58%
Emergency care	27%	41%	47%	Emergent care	24%	40%	50%
Urgent care	45%	63%	67%		-	-	-
Combined emergency and urgent care*	36%	52%	57%	Emergent care	24%	40%	50%
Non-urgent care	39%	58%	67%	Non-emergent care	38%	57%	60%
Self-care	32%	43%	51%	Self-care	40%	57%	65%
Best individual diagnostic symptom checker:	Symptomate 61%	Symptomate 77%	Symptomate 81%		DocResponse 50%	Symcat 71%	Isabel Healthcare 84%

*For comparison purposes to Semigran et al.'s (2015) research, the 'urgent care' triage classification created for this investigation could be combined with the 'emergency care' triage classification. This is not a perfect comparison as the methodologies are not identical between the two research projects.

While an imprecise comparison, if the results for the emergency and urgent care triage classifications were combined for the 'first diagnosis' category, the percentage of correct diagnoses is 36% for this research. Doing the same for the top three and top ten listed differential diagnoses and the results are 52% and 57% respectively. Despite the five-year difference between study's and the reduction from including the 'top 20' differentials down to only including the 'top ten' and the difference in symptom checkers evaluated, these results are remarkably comparable. Thirty of the vignettes utilised by Semigran et al. (2015) were incorporated into this study plus the inclusion of 18 new scenarios. Some of Semigran et al.'s (2015) vignettes were adapted to enhance the relativity to the Australian context. For some vignettes, we accepted multiple 'correct' diagnoses given the similarity of symptoms with many other conditions, which differs from Semigran et al.'s (2015) methodology which had one acceptable diagnosis. Despite these adjustments and methodological differences between studies, the similarity in findings across studies indicates a robustness to these data.

5.1.2 Variation between symptom checkers

There was wide discrepancy between the performance of individual symptom checkers, with the most proficient diagnostic symptom checker—Symptomate (2019a, 2019b, 2019c)—achieving 61% accuracy for the first listed diagnosis and the least proficient—ePain Assist Google (2018)—identifying the correct first listed diagnosis in only 12% of instances. DocResponse was the best performing individual diagnostic symptom checker in Semigran et al.'s (2015) research (Table 5.1 above). DocResponse was not included in the research as it did not appear in any of the online searches. Some symptom checkers provided extensive lists of differential diagnoses, for example Right Diagnosis provided 676 possible differential diagnoses for the 'viral upper respiratory' vignette. Previous research (Semigran et al., 2015) had also found some differential diagnosis lists extensive, and for this reason we capped the number of differentials to be considered in this research at ten.

Diagnostically, programs advertising the use of AI performed with greater accuracy for the first listed diagnosis, top three and top ten listed differential diagnoses (Table 4.3). Whether a program was categorised as AI or not was based on information supplied on each app or website and was therefore an imprecise determination. It was beyond the scope of this research to establish the legitimacy of these claims or attempt to determine the actual AI

logic behind the symptom checker program. Symptomate's (2019a, 2019b, 2019c) AI program was the strongest performing diagnostic symptom checker overall, with the highest percentage for listing the correct diagnosis first as well as listing the correct diagnosis within the top three and top ten differentials (61%, 77% and 81% respectively). Buoy Health (2018) has widely promoted the use of AI in developing their symptom checker (Baum & News, 2017; Hauer, 2018). However, its individual performance was 35% accuracy in listing the correct diagnosis first, 50% for listing the correct diagnosis in both the top three and top ten differentials. As it is likely that Buoy Health's (2018) program utilises limited memory AI (i.e. designed to learn from historical information continuously building up its database), it is possible this performance may improve over time. Symptomate's (2019) AI symptom checker was launched at least two years before Buoy Health's (2018) and has enjoyed additional years to 'learn' through the building up of their database. It is also entirely plausible that a 'real consumer' would gain differing results from Buoy Health's (2018) symptom checker, as unlike the research protocol, they would answer every question according to symptoms they are experiencing. In this research, only the symptoms described in the vignette were input into the program and wherever possible questions were 'skipped' if they did not relate to a symptom contained in the scenario. By skipping questions, performance may have been affected for certain symptom checkers such as Buoy Health (2018). This was done to achieve as much consistency as possible between assessing each individual symptom checker; essentially, each was given the same information with which to make their diagnosis.

Surprisingly, requesting demographic information did not provide a significant contribution to diagnostic accuracy, for the first listed diagnosis and the top three differentials (Table 4.3, 'Max. No. of diagnoses listed'). Symcat (Symcat, 2019) was one of the few programs to ask about ethnic origin. Integral to current healthcare paradigms is the concept of high-quality patient-centred care. Each age category, gender type, and ethnic group have unique patterns of injury and illness, as well as the potential for distinctive language and health beliefs (U.S. Department of Health & Human Services, 2018). Therefore, to aid in diagnostic accuracy, it would seem reasonable to expect modern symptom checkers to request these forms of data routinely from consumers, given certain populations may be predisposed to particular conditions. In this research, a minimum of age and gender needed to be solicited for a symptom checker to be classified as 'requesting demographic information', however 42% of symptom checkers failed to meet this criterion. Overall, 58% of symptom checkers asked for age and gender, and 64%

requested age, gender and ethnicity. All symptom checkers suggesting they utilised AI did request demographic information. It is possible demographic data aided AI symptom checkers to provide more accurate diagnoses than symptom checkers which did not use AI and/or did not request demographic information; however, this cannot be verified. Symptom checkers are largely in the first generation of such tools. In the future, with the continuing evolution of symptom checkers, attention should be paid to the inclusion of demographic data to aid provision of more accurate differential diagnoses.

5.1.3 Results from different vignette classifications

The diagnostic accuracy was significantly higher for urgent care vignettes compared to emergency vignettes (1st and 4th ranked for urgency out of 4 respectively), for the first listed diagnosis, top three and top ten differential diagnoses (Table 4.1). This differs from Semigran et al.'s (2015) research where self-care vignettes were the most accurate diagnostically, followed by non-emergent then emergent (1st, 2nd and 3rd ranked for urgency out of 3 respectively). It was expected this research would follow previous findings. It is possible the inclusion of 'Australian' vignettes such as the blue bottle jellyfish sting and stinging nettle in the self-care category may have influenced these results as both vignettes were poorly diagnosed. Other self-care vignettes such as cradle cap and tinea pedis were also diagnosed inaccurately and not included in previous studies. Emergency vignettes are more likely to have life-saving implications—if the accuracy figure is lower for emergency care than for urgent care, then this is important from a safety point of view. For the first listed diagnosis, emergency care vignettes had 27% accuracy, the lowest percentage of all the triage classifications, while urgent care had the highest accuracy rate of 45%. Table 4.7 depicts which vignettes had the correct diagnosis listed first, within the top ten, or were deemed incorrect (the correct diagnosis did not appear in the top ten differentials), for each symptom checker.

Diagnostic accuracy was also significantly higher for common vignettes than uncommon (42% versus 4% respectively for first listed diagnosis, Table 4.1), echoing prior research by Semigran et al. (2015). Symptoms which can be attributed to a common condition will usually be diagnosed as that common condition—as by definition it is more probable to be the correct diagnosis rather than the rare one (Aaronson, 2011). A wide variety of uncommon conditions account for 10% of presentations in Australian general practice (Cooke, Valenti, Glasziou, & Britt, 2013), but any single rare condition is by definition encountered far less than 1% of the time (Richter et al., 2015). So,

when a physician is considering differential diagnoses, the more common options will typically be considered first. However, particularly for a serious but less common disease, the more serious diagnosis should also be considered. For example, when a patient presents with chest pain, clinicians should consider an acute myocardial infarction (heart attack) or a musculoskeletal injury initially. The heart attack is reasonably common but also very serious, while the musculoskeletal injury very common. Other differentials to be considered would include a pulmonary embolus which is rare, and an abdominal aortic aneurysm which is exceptionally uncommon (Gregory, Ward, & Sanders, 2010). Therefore, it is unlikely an abdominal aortic aneurysm would be diagnosed first due to the rarity of the presentation unless there was other pertinent information such as a family history of this condition, or the patient had previously suffered from an aneurysm.

Agreement on what defines a disease as 'rare' is difficult. The European Union suggests a prevalence rate of 1:2000 would merit a condition to be considered as rare (Simone Baldovino, Domenica Taruscio, & Dario Roccatello, 2016). For those suffering from a rare condition, the diagnosis often arrives late as the diagnostic process takes longer due to the rarity of the of the disease. The internet now provides access to information on even the rarest illness, with Google Scholar providing public access to many peer-reviewed articles formerly limited to university databases. Niche organisations can have an online presence with webpages devoted to specialist subjects, including rare diseases. Rare Voices Australia (n.d.) is one such organisation. Their website provides information, advocacy, education for policy and decision makers, and fellowship for patients and families with rare diseases. While a single disease in itself might be rare, those living with any rare disease are less rare, with conservative estimates suggesting 6 to 8% of the Australian population suffer from a rare condition (Rare Voices Australia, n.d.). Patients and concerned parties have the incentive, time and determination to put considerable hours into finding a diagnosis online to explain their symptoms. Information found online can then open up communication channels between healthcare providers and patients.

Symptom checkers are another tool used by consumers and physicians to locate these unusual conditions. For example, Isabel Healthcare (Isabel Healthcare, 2018) was specifically designed and developed to address this need. As a young child, Isabel contracted chicken pox which later developed into necrotising fasciitis and toxic shock. Fortunately, she survived the horrific experience. Her parents subsequently established the Isabel Medical

Charity with the purpose of developing a paediatric tool for clinicians. This tool (the symptom checker) helps formulate a diagnosis quickly by listing all potential conditions for a group of symptoms, from the rare through to common diseases. The hope was that other children would not suffer from a delayed diagnosis like Isabel did, with the power of a computer aiding diagnosis. This paediatric tool was later extended to include adult conditions and has been made available to the general public (Isabel Healthcare, 2018).

For this research, whether it was reasonable to expect a diagnosis to appear within the top three differentials was also assessed. The two practising GPs and ED specialist who reviewed the scenarios also made a determination on whether we expected the diagnosis to be present or not, within the top three differentials for each vignette. In some cases, we did not expect a diagnosis to appear within the top three differentials for a variety of reasons. Some conditions share many similar symptoms with other illnesses, making it more unlikely the correct diagnosis would be listed as the first diagnosis, or even within the top three differential diagnoses. As an example, the symptoms for a common cold, pharyngitis, laryngitis, tonsillitis, and glandular fever are all similar, and the 'correct' diagnosis may be placed further down the list due to the commonality of symptoms. Other more unique signs and symptoms, time, or further testing may be required to determine which condition is the actual diagnosis. Therefore, this research was flexible in accepting multiple possible 'correct' answers for some vignettes (refer to Appendix B). Another reason a diagnosis may not appear within the top three listed options is that the condition is relatively uncommon. Physicians typically do not look for the unusual diagnosis first but rather consider the most likely option given a particular grouping of symptoms (Aaronson, 2011). For example, malaria is not endemic to Australia; we do not have the particular mosquito that carries this disease. Therefore, if a patient presented to a GP with symptoms of a fever with chills, it would be unlikely malaria would be considered within the top three differentials, unless the GP was aware of recent travel to a country that has this mosquito. Instead, the more common options would be investigated first, such as a viral illness. Table 4.6 depicts which vignettes were expected to have the diagnosis appear in the top three differentials. Sixty percent of SCVTs accurately recorded the correct diagnosis within the top three options that were expected to appear in the top three differentials list. For those vignettes where the diagnosis was not expected within the top three differentials, approximately 22% still produced the correct diagnosis within the top three options.

Certain vignettes were almost universally answered incorrectly by all symptom checkers; particularly haemolytic uraemic syndrome; malaria; Hendra virus; Queensland tick typhus; Ross River virus; solar keratosis; and blue bottle jellyfish vignettes. Many of these vignettes relate to Australian specific conditions and have not been identified in overseas symptom checkers as a potential differential diagnosis, even outside of the top ten listed answers. The single symptom checker that was 'correct' for the Queensland tick typhus vignette Symptify (2018; Table 4.7) actually diagnosed Lyme Disease, another tick-related condition, which was deemed an acceptable alternative answer by our two practicing GPs. There appears to be a gap in existing symptom checkers databases concerning Australia's unique flora and fauna. Given the potential for serious or fatal consequences, it is important the Australian community have access to accurate evidence-based information relating to such conditions. While Healthdirect Australia (2018) only provides triage advice, many of these conditions are included in the informational side of their website with relevant advice and treatment options. Many symptom checkers struggled with conditions involving dermatological conditions (including bites, stings and rashes) and altered mental states, frequently not supplying appropriate search words or asking relevant follow up questions. This meant that relevant information from the vignette was unable to be entered, limiting the information provided to the program and potentially affecting the resultant output.

5.2 Overall performance – Triage

5.2.1 Appropriateness of triage disposition

Triage advice gives guidance to the consumer about the type of health service they should attend to seek care—referred to as the 'disposition'—and how urgently they should seek intervention for their condition (Middleton et al., 2016). Triage advice provides a course of action for the consumer, while a formative 'diagnosis' given by a symptom checker (in many circumstances) is unconfirmed until the formal diagnostic process has been concluded by a medical practitioner. It has therefore been suggested that triage advice may be more immediately useful to the consumer than a possible differential diagnosis (Middleton et al., 2016).

As mentioned above, this research has four classifications of triage acuity—emergency care, urgent care, non-urgent care and self-care—to more accurately reflect the Australian healthcare system, as opposed to Semigran

et al.'s (2015) emergent, non-emergent and self-care classifications. This may account for lower percentages obtained for triage accuracy in this investigation as can be seen in the table below (Table 5.2).

Table 5.2: Comparison of triage results between studies

This research		**Semigran et al.'s (2015) research**	
Overall combined performance of all 4 categories	49%	Overall combined performance of all 3 categories	57%
Emergency care	63%	Emergent care	80%
Urgent care	56%		-
Combined emergency and urgent care*	59%	Emergent care	80%
Non-urgent care	30%	Non-emergent care	55%
Self-care	40%	Self-care	33%
Best individual triage symptom checker:	Healthdirect Australia 61%		HMS Family Health Guide 78%
*For comparison purposes to Semigran et al.'s (2015) research, the 'urgent care' triage classification created for this investigation could be combined with the 'emergency care' triage classification. This is not a perfect comparison as the methodologies are not identical between the two research projects.			

Combined triage performance across all symptom checkers achieved 49% concordance with our assessment across all vignettes, lower than previous estimates in the USA which found 57% accuracy (Semigran et al., 2015). In the present research, evaluating each triage category independently, performance was best in the emergency care classification with concordant advice provided in 63% of cases, followed by urgent care with 56%, dropping to 30% for non-urgent care and rising to 40% for the self-care classification.

Semigran et al. (2015) also found the emergent care triage classification (called 'emergency care' in this research) to be the most accurate at 80%, followed by non-emergent care (or 'non-urgent care' in this research) at 55% and self-care at 33%. Individually, the most accurate triage symptom checker in this research was Healthdirect Australia (2018), providing concordant advice in 61% of SCTVs, whilst Semigran et al. (2015) determined HMS Family Health Guide (78%) provided the most concordant triage advice in their investigation. HMS Family Health Guide was not included in this research as it did not appear in any of the online searches for symptom checkers. While an imprecise comparison, if the results for the emergency and urgent care triage classifications were combined, the percentage of concordant triage advice is 59% for this research, compared to Semigran et al. (2015) which found 80% accuracy in their emergent category.

Triage advice, especially for non-urgent and self-care categories, tended to be risk-averse recommending more urgent care than necessary. Forty percent of vignettes in the non-urgent and self-care categories were given advice to attend urgent or emergency care. From an economic and medical perspective, over-triage is a burden and a serious strain on the stretched resources of the ED. There were also notable instances of under-triage which is particularly concerning for vignettes classified as requiring emergency assistance (Table 4.8). Two notable examples are provided below:

Case One: Acute Liver Failure.

This vignette had 'confusion' as the chief complaint. Accompanying symptoms were disoriented, increasingly sleepy, mild right upper abdomen tenderness, with a history of migraines and the patient has been taking paracetamol for lower back pain. As the patient has an altered mental state due to her symptoms of confusion and disorientation, she should be directed to seek emergency care. However, Family Doctor (2018) has recommended self-care which is unsuitable for these symptoms.

Case Two: Stroke.

In another instance non-urgent care was recommended for a stroke, which is a time-critical condition (Family Doctor, 2018). The chief complaint was a 'weak right arm' and other symptoms included confusion and trouble speaking. Suggesting anything but emergency care could have serious consequences for these types of conditions.

For emergency presentations, 10% of vignettes received advice to seek non-urgent or self-care which could have grave implications for consumer safety. Triage appropriateness of disposition was significantly higher for common conditions compared to uncommon, consistent with prior research (Semigran et al., 2015). Possibly, by virtue of being 'common', the condition is accompanied by symptoms which are easily recognisable and therefore easier to provide an appropriate disposition (Aaronson, 2011).

Semigran et al. (2015) found some symptom checkers always referred consumers to a healthcare service, never suggesting self-care. They conducted a sensitivity analysis removing those symptom checkers from their

evaluation. The assumption was that symptom checkers that do not have self-care as an option may have skewed the overall results, by virtue of not suggesting the full range of triage options. Analysing the group without these symptom checkers potentially provides a better reflection of overall triage performance. Removal of these symptom checkers increased appropriate triage by 4% for the Harvard Medical School's investigation. Their triage results improved from 57% to 61% for appropriate advice. Similarly, we ran analysis both with and without these types of symptom checkers. By removing the four-symptom checkers which never suggested self-care (Isabel Healthcare (2018), Everyday Health (2019), Symcat (2019) and Doctor Diagnose Symptom (2018)) appropriate triage disposition increased by only 2%; from 49% to 51% overall. In approximately half of all searches, triage advice was not concordant with our assessment for the level of triage category (or acuity) based on the symptoms presented in the vignette. Some symptom checker algorithms may not be programmed precisely enough yet to accurately determine the correct level of acuity based purely on a list of symptoms. There is a tension between providing information to enable consumers to self-manage their condition (where appropriate) and to not over-use services; and ensuring consumer safety and considering 'worst case scenarios', a balance needs to be met.

5.2.2 Red Flag conditions

Certain symptoms or grouping of symptoms raise something called a 'red flag' alert (Heneghan et al., 2009). Red flags are "clinical indicators of possible serious underlying conditions requiring further medical intervention" (New South Wales Government, 2005, p. 1). From a clinical perspective, these patients should seek urgent, if not emergency care, as soon as possible. Isabel Healthcare (2018) uses this medical metaphor effectively in their differential diagnoses list, actually raising an image of a 'red flag' next to conditions that are serious.

Yet other sites provided with the same symptoms could not find any matching condition, nor raised any sort of alert to the consumer that they should be concerned about these symptoms and act quickly to seek care. This was found with the American Association of Retired Persons (AARP) Health Tools site (2018) as described below.

Case Three: Myocardial Infarction.

Given the three symptoms of ‘breathless, chest pain and sweating’, the output produced was “We have no matching results for the symptoms that you entered” (AARP Health Tools, 2018). These are the symptoms for the myocardial infarction vignette and is an emergency condition requiring immediate medical attention.

While the AARP Health Tools (2018) symptom checker does not provide triage advice, other sites that similarly only provided diagnostic information still raised alerts with certain ‘red flag’ symptoms, such as MedicineNet’s (2018) symptom checker. MedicineNet overlaid the warning “If you are experiencing severe chest pain or discomfort please seek emergency medical attention” displaying an exclamation mark next to the warning (MedicineNet, 2018). It should be noted however that chest pain does not need to be severe in all presentations of myocardial infarction, and in some instances, there will be no chest pain at all (Canto et al., 2012).

In a recent study conducted by the NHS, 20% of consumers using various online symptom checkers in the UK were referred to the ED or 999. A further 20% of consumers received a call-back from a 111-helpline clinician (who could be a doctor, nurse or other healthcare professional) following their online transaction (Lind, 2018; NHS England, 2017). This data confirms consumers do consult online programs such as symptom checkers for emergency and urgent medical conditions. As highlighted above, some symptom checkers do not alert consumers to the need for emergency medical treatment for concerning symptoms. Symptoms such as an altered mental state (i.e. confusion, disorientation), chest pain, paralysis or weakness in a limb or face should always be ‘red flagged’ in online programs for consumer safety, due to the potentially serious or life-threatening consequences. If a consumer is not alerted to the importance of this symptom, it is possible a non-urgent care pathway may be followed, which can delay treatment and possibly result in permanent injury or possibly death. There appears to be a concerning gap in some symptom checkers programming about alerting consumers to serious symptoms. Accordingly, consumer needs may not be met, and future research needs to identify a solution to this concern. The need for ongoing research post deployment of products to the market is elaborated on in section 5.6.1 below.

5.2.3 Results from different vignette classifications

While the use of AI technology by symptom checkers was associated with an improved diagnostic accuracy, there was no association with improved triage disposition advice. Buoy Health (2018) have invested significantly in AI technology to help facilitate guiding consumers to the correct point of care in the healthcare system (Lovette, 2018). Buoy Health's (2018) triage advice was regarded as 'unassessable' for this evaluation, as the consumer was required to select a diagnosis from the three differentials offered before receiving triage advice, as illustrated below.

> Case Four: Acute Bronchitis.
>
> We determined the patient should seek care 'urgently' for acute bronchitis, therefore seek medical attention within 24-hours. Buoy Health's (2018) differential diagnoses and triage advice was as follows:
>
> - Bronchitis = self-treatment;
> - Acid Reflux (or gastroesophageal reflux disease) = see a primary care doctor within two weeks;
> - Bacterial Pneumonia = in person visit [to a doctor] today or as soon as possible.
>
> The consumer is then left deciding which diagnosis is likely to be correct and whether that triage advice applies. In this case, the consumer might decide they have bronchitis and self-treat, or alternatively lean towards the other extreme; that they have pneumonia and rush to see their GP. The former is the correct diagnosis (albeit it should specify 'acute bronchitis') while the latter is the correct disposition advice to follow.

Whether a symptom checker only provided triage advice or provided both triage and a diagnosis made no difference to triage accuracy (Table 4.3, 'Max. No. of diagnoses listed'). However, requesting demographic information did. Those triage symptom checkers requesting demographic information triaged in accordance with our assessment in 53% of instances versus 41% for those that did not request demographic data. Ten of the 19 triage symptom checkers did not request demographic information (i.e. requesting both age and gender as a minimum requirement). Of these ten, eight did request some form of demographic information, (i.e. either age or gender). While it cannot be determined if this was what influenced the improvement in triage advice, it may have been a contributing factor.

5.3 Symptom checker user interface

While the impact of symptom checker design was not specifically tested in this investigation, certain observations were made during the evaluation by the researcher. While problems faced under 'laboratory testing' may not directly translate to real world consumers, it would be reasonable to suggest consumers may face similar challenges. Therefore, a segment of this discussion has been used to discuss symptom checker user interfaces. The results discussed below are factually reporting what was input into various programs and what outputs were provided.

5.3.1 Symptom selection

Symptom checkers prescribe to set rules regarding entering information into their system, with some sites offering more flexibility than others. For example, Isabel Healthcare (2018) would allow users to type in symptoms using their own words or alternatively select from a drop-down list of options thereby completely encompassing layperson's language; a layperson did not need to know or understand medical jargon. Websites like MedicineNet (2018) allowed consumers to enter their own words such as 'feel sick' and the options that were automatically generated for selection were 'nausea or vomiting (abdomen)' or 'nausea or vomiting (upper abdomen)' exchanging layperson's language for the more medically descriptive option. Alternatively, the consumer could click on a body part and scroll through a list of options to select the most appropriate sign or symptom. However, using this option, the layperson would need to know that nausea meant 'feeling sick' in order to select the appropriate option. It is possible that symptom checkers which allowed consumers to type symptoms in freely may have had different results from those sites that only allow selection from a list, drop down box or clicking on a body part. Due to the variety of differences in how symptom checkers function is so large, it was impossible to classify them succinctly.

Certain symptom checkers provided a finite directory of symptoms the consumer can select from, potentially affecting the diagnostic capability of the program. Mayo Clinic's (2018) symptom checker provides 28 chief complaints (the most troubling symptom (Sanders, Lewis, Quick, & McKenna, 2012, p. 558)) for the adult patient to select from which does not include the common symptom 'fever', and only 17 chief complaints for children. If an adult consumer's chief complaint was 'fever', this would mean another symptom would need to be selected from

the 28 offered by Mayo Clinic (2018) as the chief complaint (refer to Figure 3.6 for the symptom selection process). It is unknown how this may affect the diagnostic options provided to a consumer in real life, however it is likely to impact upon the questioning pathway determined by the 'second' choice for a chief complaint. It would be akin to a patient failing to tell a physician that their main complaint was a fever, and unless the physician specifically asked about a fever during questioning, they would be unaware of this symptom. Similarly, unless the symptom checker specifically asked about 'fever' during questioning, the computer program remains unaware of this complaint. This provides an interesting avenue for future research. To provide an example of how this may be a limiting feature, the vignette for 'Malaria' had the symptom 'fever' for the chief complaint. As the chief complaint of 'fever' could not be used, the next symptoms listed— 'chills' and 'shivering'—were searched for but were similarly unable to be input as they were not presented within the provided list of 28 options. Therefore, the last symptom had to be used: 'diarrhea' (American spelling as used by Mayo Clinic (2018)). In subsequent questioning the consumer is asked about how long the diarrhea has occurred for; if food or antibiotics preceded this diarrhea; if eating certain foods makes this condition worse; if this condition is relieved by avoiding certain foods; and lastly what accompanying symptoms are there. At this point 'fever' can be selected amongst ten other options. No other symptoms could be selected, as 'chills' and 'shivering' were not provided in this symptom list. The line of questioning is very relevant for diarrhea being the chief complaint, however, does not narrow the differentials enough. Mayo Clinic's (2018) top nine differentials provided were: antibiotic-associated diarrhea; intestinal obstruction; ischemic colitis; traveller's diarrhea; viral gastroenteritis; Crohn's disease; food poisoning; pseudomembranous colitis; and ulcerative colitis. If for argument's sake, the vignette is tweaked to being a child patient (in which case 'fever' can be selected as the chief complaint) the following pathway occurs: the consumer is asked what symptoms accompany the fever from a list of 28 options: 'chills' and 'diarrhea' can be chosen from the list provided but no other follow-up questions are asked. The five differentials then offered for a child are: influenza (flu); pneumonia; roseola; rotavirus; and viral gastroenteritis (stomach flu). Neither the adult nor the paediatric line of questioning provides malaria as a differential. This issue was not confined to Mayo Clinic (2018), multiple symptom checkers only provided a set number of symptoms to select from (i.e. Family Doctor (2018)*;* Healthwise (My Health Alberta, 2018)*;* University of Michigan Health (2019)*;* Drugs.com (2019a & 2019b)*;* What's My Diagnosis (n.d.)*;* and Doctor Diagnose Symptom (2018)) potentially causing similar limitations.

Concern regarding consumer reliance on symptom checkers for diagnosis is perhaps exemplified by examples from AARP Health Tools (2018) symptom checker. Symptom checkers follow programming rules and how information is presented can incur 'blind spots', possibly leading to diagnostic limitations. AARP Health Tools (2018) responses to the myocardial infarction have already been discussed above. The output of "We have no matching results for the symptoms that you entered" was also given for the stroke vignette. The symptoms selected from AARP Health Tools options were 'arm paralysis, confused, feeling sick, trouble speaking, vomiting', and this condition is another time critical emergency condition.

Groupings of symptoms such as 'chest pain and breathless' or 'confusion, paralysis and trouble speaking' are readily identifiable to medical professionals as a potential heart attack, stroke or other serious condition which requires immediate action. This particular symptom checker could not provide a single differential diagnosis for either vignette. Possible differentials that could have been suggested might include unstable angina, pulmonary embolism or pericarditis for the myocardial infarction vignette (Ashwath & Gandhi, 2018), or brain aneurysm, seizure or transient ischemic attack for the stroke scenario (Ntaios, 2018). Furthermore, for the myocardial infarction vignette, inputting all three symptoms into AARP Health Tool's (2018) symptom checker provided no results (or no differential diagnoses), however, the response for the single symptom of 'chest pain' provided an extensive list of 51 differential diagnoses with the correct response of heart attack listed second. Counter-intuitively, by increasing the list of symptoms, diagnostic accuracy was reduced. While AARP Health Tools (2018) was evaluated on the answer given for the myocardial infarction vignette using all the scenario's symptoms ("We have no matching results for the symptoms that you entered"), rather than the exploratory assessment done where a single symptom was entered, this example demonstrates the variability in accuracy for some symptom checkers based on the number symptoms input into the program.

It is unlikely a 'single key symptom' or chief complaint would be representative of how most apps or websites are designed or expected to be used, without entering other symptoms into the program. While it is possible consumers may choose to input only one symptom into the program, this particular symptom checker prompts the user to 'add related symptoms', suggesting the addition of multiple symptoms is desirable. Many symptom checker sites actively suggest the addition of extra symptoms is desirable. For example, Symptomate (2019b) and Everyday Health (2019) prompts users to "Please try to add more than one symptom". Symcat (2019) goes a step further,

asking the consumer to “add more symptoms to improve your results” and the Patient.info (n.d.) site reads “Enter more symptoms for more accurate results, starting with your most severe symptom”. As with a physician’s consultation, providing a list of symptoms helps to exclude certain conditions and narrow the list of differentials down to a manageable number (Balogh, Miller, & Ball, 2015). For example, the symptom ‘cough’ relates to a vast number of medical conditions. By including other symptoms such as ‘fever’, ‘headache’, ‘nasal discharge’, this provides more information to the program (or clinician) narrowing down potential diagnoses and helps determine how urgent the condition might be. More research is required to determine if symptom checkers that offer the consumer greater flexibility with symptom selection are likely to provide more accurate advice.

5.3.2 Symptom checker language

Other interfaces and programs were even less user-friendly or had aspects which were undesirable. For example, Symcat’s (2019) symptom checker was prone to using medical terminology during follow up questioning, for instance, asking the consumer if the patient had apraxia or hypernasality. This kind of medical terminology used during questioning can potentially confuse the consumer who may in turn answer questions incorrectly, drastically altering diagnostic and/or triage results. This was a known limitation of Semigran et al.’s (2015) research where they utilised medical terminology to test online symptom checkers. Other sites including Healthdirect Australia (2018) also use medical terminology but provide corresponding explanations of these terms in lay language. By providing definitions to medical terminology, Healthdirect Australia (2018) hopes to strengthen health literacy and promote patient empowerment, which may contribute towards improving patient-physician relationships (Healthdirect Australia, 2017). This potentially aids the consumer to respond appropriately to questions from the symptom checker (Healthdirect Australia, 2017, 2018c), thereby increasing the appropriateness and relevance of the symptom checker output and improving consumer adherence with advice.

A further difficulty encountered was that symptom checkers usually accepted either UK English (which is almost identical to Australian English) or USA English which affected the ability to search for some symptoms using Australian English. Therefore, the Australian consumer may need to type in esophagus rather than oesophagus if using an American product. It is difficult to know how this would affect consumers, as there is a certain ‘global

awareness' of USA or UK spelling especially in this age of the internet (Buckby, 2016; Picardo, 2012). Patient.info (n.d.) was a one site which did accept both language conventions. The differing names for medications may be more problematic however, as symptom checkers will not accept an unknown drug name, meaning the consumer would need to be aware of other names for similar medication. For instance, WebMD (2017) would accept albuterol (not used in Australia) which is used to treat asthma in the USA, but did not recognise salbutamol or Ventolin, which are generic and brand names respectively in the same class of medication used in Australia. In a similar vein, questions regarding fevers and temperatures often required readings in Fahrenheit rather than Celsius which is used in Australia.

5.3.3 Site navigation

Some symptom checker websites and apps were problematic to navigate. Most providers had the link to their symptom checker in the tool bar at the top of the screen, or predominantly displayed on their home page. However, Healthline had the link to their symptom checker on their home page, but it was situated five 'folds' down the screen making it difficult to locate. Similarly, Drugs.com (2019b) has their symptom checker link 3 folds down the screen and located to the left with multiple other links, making it challenging to find. AARP Health Tools (2018) and ePain Assist's Apple and Google apps (2018a and 2018b) had adverts constantly pop up during questioning. This delayed the diagnostic process and could become stressful for a consumer if they were querying an urgent or emergency condition.

Right Diagnosis (2011) had an extremely 'busy' website which was difficult to navigate. Even finding where to add the chief complaint was confusing and locating the part of the screen to add subsequent symptoms was difficult to find—past the 'fold' in the screen. Once the appropriate section to 'add extra symptoms' has been located, the catalogue of symptoms provided could be extremely long. With each symptom selected by the consumer, the page reorders itself so the consumer needs to start the search from the top of the list again, nor were the symptoms catalogued in alphabetical order to allow for quick location of the symptom. In contrast, Right Diagnosis (Right Diagnosis, 2011) listed their differential diagnoses in alphabetical order giving no weighting to more probable conditions; which affected their accuracy rating. By way of an example, the appendicitis vignette has the correct

diagnosis (appendicitis) beginning with the letter 'A'. Therefore, the correct diagnosis was listed second following abdominal cancer. However, for the myocardial infarction vignette, the correct diagnosis could be either 'heart attack' or 'myocardial infarction'. These options did not appear within the top ten possible options as all of the top ten differentials began with the letter 'A' (the tenth option was 'adrenal hypertension'). ePain Assist's Apple and Google apps (2018a & 2018b) also listed their differential diagnoses alphabetically affecting their evaluation.

5.3.4 Line of questioning

Some programs narrowed in on one train of questioning, failing to investigate the patient holistically. A classic example is Symptify's Google app (2018), when evaluating the stroke vignette. After entering the chief complaint of 'weak right arm', the app asks for the consumer to click on the body part where the weakness is. The entire arm from shoulder to fingers was selected by the researcher (i.e. the shoulder, upper arm, elbow, lower arm and hand were each independently selected). All subsequent questioning focussed only on the shoulder, ignoring the other parts of the arm, and the singular differential diagnosis ultimately given was for a 'rotator cuff tear'. No questions were asked regarding any difficulties with the other parts of the arm, nor were questions asked regarding other possible symptoms like an altered mental state or speech difficulties, which might be elicited from this concerning symptom. Symptify (2018), like Buoy Health (2018), provides triage advice based on the diagnosis selected by the consumer from the list of differentials, therefore their triage advice was deemed unassessable for this research. In this particular case, the singular differential provided was 'rotator cuff tear' and the triage advice given was to 'see a specialist'.

A stroke is considered a time critical medical emergency and treatment is best instigated within three hours of the event for optimal patient outcome (Ragoschke-Schumm et al., 2014). 'Time is brain' is a common catchphrase for a stroke, with each passing minute around 1.9 million neurons and 14 billion synapses are irreversibly destroyed (Saver, 2006). Not all strokes cause severe symptoms, many can be mild but are still considered an emergency situation (Brain Foundation, 2017). Transient ischaemic attacks (TIAs) are often coined 'mini strokes', where the symptoms completely resolve within 24 hours (75% will resolve within an hour) without neurological deficit (Murtagh & Rosenblatt, 2015; Sanders, Lewis, & Quick, 2012). If a TIA is left untreated, one-in-three patients are likely to

suffer a stroke within five years (Murtagh & Rosenblatt, 2015). In Australia, strokes are the primary cause of disability and the third most common reason for death (Brain Foundation, 2017). In this scenario, if an Australian consumer received the same diagnosis and followed the triage advice given, they would need to see their GP first to obtain a referral to a specialist. This extra step could delay treatment for a time critical condition requiring emergency care, and potentially be a life-threatening mistake. Ideally, the patient should be directed to attend a specialised stroke unit via ambulance without delay. It is critical symptom checkers provide accurate advice to seek emergency care for medical emergencies to ensure optimal outcomes and patient safety.

Rarely, lines of questioning incorporated a pain scale. Healthwise (2018) was one of a few programs to include a pain scale when the user was questioned regarding their pain. Given a computer program cannot 'see' a person's pain and the use of pain scales are fundamental to a clinician's questioning, it would seem essential to have this incorporated into a symptom checker's algorithm. Pain scales provide important information pertaining to pain intensity, duration and possible changes over time and are very useful as a screening assessment tool (Breivik et al., 2008). Therefore, when a consumer suffers from abdominal pain for example, the pain scale provides an indication of severity and potential urgency involved with the complaint which helps the program (or clinician) to determine the appropriate disposition. It also should provide important clues for potential differential diagnoses. Pain scales were not specifically measured in this investigation so it is unknown if provision of pain scales improves diagnostic or triage accuracy, however, this would be an avenue for future research.

WebMD (2017) was one program that enquired about previous and current medical conditions, however, would not accept asthma as a condition which is surprising given this is a fairly common complaint. In the USA, 8% of adults and children have asthma (Asthma and Allergy Foundation of America, 2019) and in Australia, it is reported that 11% of the population suffer from this condition, higher than in the USA (Healthdirect Australia, 2018a). Knowing a person has a history of a respiratory disorder is useful when considering a differential diagnosis and/or giving triage advice. Combine this with the same program not recognising a medication name (this program did not recognise Ventolin, the trade name for salbutamol in Australia, and only recognised albuterol, the latter being a generic name that is not used in Australia) and this potentially creates another limitation within the symptom checker.

For consumers in the southern hemisphere the influenza season is usually April to October, peaking in August. In 2019, the peak was suggested to arrive early, in June/July (Healthdirect Australia, 2019). However, Drugs.com's (2019a & 2019b) symptom checker recognised the 'flu season as "only being a concern during the months October through February", disregarding the possibility of obtaining influenza at other times of the year (a distinct possibility with international travel) or that consumers outside of the USA may use the product. The NHS has precluded the possibility of international users by limiting their site to accepting only UK postcodes (NHS England, n.d.-a). This may limit the usability of international products to the Australian consumer, as epidemiological trends for America are significantly different than in Australia.

5.3.5 Symptom checkers with multiple platforms

Symptom checkers running across multiple platforms (e.g. internet website, iOS Apple and android Google Play apps) had the programs tested in tandem. This was to ensure each response would be input into the programs at the same time, guaranteeing questions were answered identically. These programs were not always consistent in questioning, with one version of a product occasionally following a different line of questioning from the other platforms, sometimes returning to the same questioning thread and other times not. The subsequent diagnostic listings resulted in some discrepancy between results, possibly due to AI algorithms or programming deficiencies. This does not imply one platform is superior to another, rather demonstrates that there are inconsistencies between platforms for reasons unknown. This draws attention to the need for adequate product testing prior to releasing it onto the market. All platforms by the same company should provide the most accurate information possible for consumer safety. Future research could investigate this phenomenon and determine the cause of such discrepancies.

5.3.6 Diagnostic process

Symptom checkers prescribe to rules within the program's algorithms. Murtagh and Rosenblatt (2015) describes diagnosis as a process, calling it the "most difficult, complex and challenging of all the healing arts" (Murtagh & Rosenblatt, 2015, p. 150), and diagnostic tests such as blood analysis or electrocardiograms often provide a physician important clues in facilitating a definitive diagnosis. Accordingly, the diagnostic process can take time,

multiple tests and examinations to arrive at a definitive answer, a process by which symptom checkers would find difficult to replicate under current conditions.

A diagnostic error is considered to be a "diagnosis that was unintentionally delayed (sufficient information was available earlier), wrong (wrong diagnosis made before the correct one), or missed (no diagnosis ever made), as judged from the eventual appreciation of more definitive information" (Abimanyi-Ochom et al., 2019, p. 1; Graber, Franklin, & Gordon, 2005, p. 1493). Research in the United States suggests physician diagnostic error rates in outpatients is approximately 5% (Singh, Meyer, & Thomas, 2014), while an investigation by Semigran, Levine, Nundy, & Mehrotra (2016) found physicians made diagnostic errors in ~15% of clinical vignettes, concordant with earlier research findings by Graber (2013). Physician error rates typically presented in previous studies are substantially lower than most of the individual symptom checker error rates in this research and in Semigran et al.'s (2015) findings.

False negatives given by symptom checkers can be dangerous and compromise patient safety (Burgess, 2017). False negatives are when a diagnostic test incorrectly indicates a disease or condition is not present (Mosby., 2016). Wolf et al. (2013) determined high false negative rates in their study of melanoma apps, with 30% of melanoma's misclassified as of 'no concern' in three out of four apps, which poses serious concerns surrounding patient safety of these tools. Symptom checkers rely entirely upon the information provided by the consumer. If the consumer does not provide all the relevant information, perhaps not realising the importance of a symptom, this could affect the output provided by the symptom checker. Some symptom checkers allow the consumer to input test results and chronic medical conditions, potentially improving the analysis of a differential diagnosis. How often consumers would use this functionality is debateable (and not determined through this research), however providing this level of detail takes time to upload and there would be concerns surrounding patient confidentiality (Jutel & Lupton, 2015). Various programs provide the ability to print out the analysis to take to a physician. For example, Isabel Healthcare (2018) and Healthdirect Australia (2018c) both provide the consumer with an option to print out their triage assessment to present to their physician, while Symcat (2019) integrates with other services and provides an option to book appointments with a primary care provider via their webpage. The degree of

inaccuracy revealed in this research demonstrates the difficulty symptom checkers have in determining an accurate differential diagnosis using current algorithms.

5.3.7 Sensitivity and specificity

When analysing any diagnostic tool, the sensitivity and specificity of that tool to provide accurate diagnosis is considered to assess the extent it can be expected to be effective in clinical practice. Sensitivity refers to the ability for a tool to correctly identify a person has a condition, and specificity being the ability to correctly identify when a person does not have a condition (Smith, 2012). If a tool/test were perfect, it would provide 100% accuracy on both specificity and sensitivity. Generally, no tool is this precise, there is frequently a trade-off between high specificity and high sensitivity. However, a test that does not work at all would achieve 50% accuracy (i.e. no better than chance) (Smith, 2012).

For example, mammograms are screening tests to determine if a woman is likely to have breast cancer. If a mass is found during imaging, she will be referred for further testing (such as a biopsy) to establish what the mass is. If the mammogram found a mass, and it was determined the patient did have breast cancer, this is a true positive. A false positive would occur if the mass was something else like a cyst. Conversely, if nothing was detected during imaging, and shortly afterwards the woman was diagnosed with breast cancer, this is a false negative. A true negative is when no mass was found, and the patient does not have breast cancer. False positives are preferable to false negatives due to the risk of serious consequences for the patient if a diagnosis is missed. There is also a cost to the healthcare system; for false negatives, early identification of disease aids reduction in morbidity and mortality, but there is a cost to additional unnecessary testing for false positives. [This is not suggesting that further testing is not desirable]. For a mammogram to be a useful clinical tool, it needs to provide true positives and true negatives with a reasonable degree of accuracy. Estimates for mammogram sensitivity are between 75%—90% and specificity of 90%—95% (Ferrini, Mannino, Ramsdell, & Hill, 1996). These statistics far exceed the random chance figure of 50% accuracy.

Admittedly, symptom checkers do not provide a dichotomous response. For example, for the question 'Do I have a cold?' the answer is either 'yes, or no'. However, symptom checkers do not provide a yes/no response. Rather, the question posed to symptom checkers is 'What could these symptoms be?'. Symptom checkers then provide a range of diagnostic possibilities that the provided symptoms could match. Nevertheless, given the overall combined accuracy of the 36-symptom checkers assessed was 36%— worse than 'chance'— in providing the correct differential diagnosis first, the clinical application of these symptom checkers as a diagnostic tool seems unlikely. Babylon Health (discussed below in-depth) suggest their symptom checker performs with a sensitivity of 80% (Razzaki et al., 2018), however there is some doubt over these results due to methodological flaws with this in-house research. The implications for consumers are considerable. Should consumers obtain the same results as the present research, around 64% of first reported diagnoses will be incorrect. In some cases, the differential diagnosis offered may invoke unnecessary anxiety (Doherty-Torstrick, Walton, & Fallon, 2016) while others may believe their symptoms are not serious enough (as with the stroke example) and fail to seek timely care.

5.3.8 Age of promise rather than delivery

The problems described with diagnostic symptom checkers support the notion that triage may be the more valuable aspect of these online tools. Foreign symptom checkers also suggest healthcare services unavailable or inappropriate for Australian consumers. Although, as an entry point to the healthcare system, a well-designed symptom checker could certainly provide direction and advice in a timely manner and may aid consumers who would not typically see a GP regarding their symptoms (White & Horvitz, 2009). In the four years since Semigran et al.'s (2015) research, this research suggests there has not been a substantial improvement in the combined diagnostic or triage capabilities of symptom checkers, with previous research establishing a combined total for the first listed diagnosis being correct at 34% and the combined total from this research is 36%. Although, there have been some positive shifts in individual symptom checkers performance. WebMD's (2017) performance for the listing the correct diagnosis first rose from 36% with Semigran et al.'s (2015) research, to 53% with this evaluation, and Symptomate's (2019b) performance went from 31% to 61%. Equally there were individual downward shifts; Doctor Diagnose Symptom (2018) fell from 41% to 33% and Healthline (2018) went from 38% to 27%. Digital technology has been touted to revolutionise healthcare by improving patient experiences, lowering costs and

improving health outcomes. However, Michie, Yardley, West, Patrick, & Greaves (2017, p. 1) advises “we are still mainly in the age of promise rather than delivery” a comment which seems applicable for many symptom checkers reviewed.

5.4 Nursing triage call centres and relationship with symptom checkers

5.4.1The effectiveness of call centres

The Australian Medical Association (AMA) released a position statement in 2014 regarding triage call centres—specifically mentioning Healthdirect Australia— expressing concern at the lack of evidence to support their cost-effectiveness, and clearly state they must comply with national standards and protocols with transparent evaluation processes (Australian Medical Association, 2014). The AMA acknowledge there is a place for such services particularly in rural and remote regions of Australia where there are barriers to health service access. However, such services have not demonstrated a positive impact on EDs with low acuity presentations at ED remaining constant despite the existence of triage call centres (Australian Medical Association, 2014; Gibson et al., 2018). While the impact and appropriateness of presentations to ED may be unchanged, there has been no evaluation of the impact (positive and negative) on the wider healthcare system. In 50% of cases Healthdirect Australia’s dispositions are to primary care with a further 13% of transactions suggesting self-care based on their 2016/2017 Annual Report (Healthdirect Australia, 2017).

A 2012 Australian study investigated the effectiveness of nursing triage call centres to appropriately refer patients to ED. It was determined that Healthdirect Australia’s dispositions were similar to consumer self-referrals in their appropriateness of referrals to the ED. It was also concluded that GPs had the highest level of appropriate referrals to EDs (Ng et al., 2012). Despite advice to the contrary, more than half of the Healthdirect-referred patients attended ED when they were actually recommended to alternative care pathways. Therefore, consumer adherence is another factor to be considered regarding call centre and symptom checker recommendations. Gibson et al., (2018) conducted a population-based data linkage study using Healthdirect Australia’s call records and ED presentation, hospital admission and death registration records in New South Wales between 2009–2012 to determine caller adherance with advice to attend the ED. It was determined that ~67% of patients were adherent

with Healthdirect Australia's disposition to attend ED and were unlikely to receive the least urgent ED triage category, implying the referral was appropriate. Over six percent of consumers who were given lower acuity dispositions (therefore advised by Healthdirect Australia not to attend ED), self-referred to the ED within 24 hours nonetheless and were more likely to self-refer if that had been their original intent prior to contacting Healthdirect Australia. Another possibility is that these consumers may have felt their condition had changed or deteriorated subsequent to receiving their disposition advice from Healthdirect Australia. Typically disposition advice includes a clause to call back or seek urgent care if a condition worsens. Therefore, if Healthdirect Australia gave a disposition to see a GP and the consumer felt their condition had deteriorated, they may feel it is appropriate to present to the ED. Research into consumer adherence with Healthdirect Australia's online symptom checker would be beneficial to determine the impact this has on the healthcare system.

5.4.2 Targeting digital audiences

Like Healthdirect Australia (2018c), companies frequently operate both a nursing triage call centre and a web-based triage option simultaneously, aiming to reduce demand on call centres and target users of digital technologies (NHS England, 2017). Cost savings may be an incentive for companies to divert consumers to online tools rather than promoting nursing triage call centres (Armstrong, 2018; Baum & News, 2017; Lanseng & Andreassen, 2007; Marklund et al., 2007; Middleton et al., 2016), and the NHS have reported a potential cost saving of ~£1.50 for each triage done via digital technology rather than utilising the nursing call centre, a theoretical saving of £250,000 per month (NHS England, 2017). The NHS has been making a purposeful 'channel shift' from telephone triage to an online self-serve option to manage rising call volumes and meet increasing demand, citing Healthdirect Australia's 33% call reduction following the introduction of their online symptom checker in 2015 as an example (NHS England, 2017).

The NHS is currently trialling and operating four different online services in four different regions of the UK (Armstrong, 2018; Limb, 2016), which has resulted in an average 6% channel shift to online tools during initial trials. Each symptom checker will always provide a disposition for every transaction, the most frequent one to-date being referral to a primary healthcare provider (40%). Online consumers are more likely to seek advice regarding

lower acuity conditions than consumers accessing the call centre, with a greater proportion receiving a selfcare disposition (18%) (NHS England, 2017). This aligns with prior research where it was identified consumers contacting call centres were seeking advice regarding acute conditions, whereas web-based searches were more inclined to be for chronic conditions. Web-based searches were more frequent for adult conditions in comparison to call centre interactions, however symptoms searched were similar between both (North, Varkey, Laing, Cha, & Tulledge-Scheitel, 2011).

Web-based symptom checker triage has been immensely popular with Australian consumers as demonstrated by Healthdirect Australia's annual 2017/18 'business highlights' report, announcing a two-fold increase in web-based triage usage from the previous financial year. In total, between the nursing triage call centre and digital technologies, Healthdirect Australia fielded 1,786,995 triages last year, with just under three quarters being digitally processed (Healthdirect Australia, 2018c). Statistics regarding Australian consumer use of international symptom checkers for triage advise and/or diagnostic information is undeterminable, but the sheer number of interactions on the Healthdirect Australia (Healthdirect Australia, 2018c) sites indicates the popularity of such tools within the Australian community.

5.4.3 Clinical decision guidelines used by symptom checkers

It has been observed that some symptom checker algorithms are based on triage guidelines utilised in nursing triage call centres, for example the Schmitt and Thompson guidelines (Armstrong, 2018; Gardiner, Cullen, Karabatsos, & St George, n.d.; Semigran et al., 2015)(Semigran et al., 2015), except technological advances have changed the algorithmic logic used for some digital triage guidelines. With the development of AI and enhanced machine learning algorithms, these types of algorithms are designed to learn as they go—no longer adhering to strict nursing triage guidelines (Armstrong, 2018).

Healthdirect Australia's symptom checker has been developed in collaboration with the NHS and an Australian clinical expert panel to ensure it is relevant for the Australian healthcare system (Healthdirect Australia, 2016). The Healthdirect Australia 'information for health professionals' page has a hyperlink directing the user to a general

NHS health website and not one specifically linked to a symptom checker, nor supplies any information pertaining to the Australian or NHS symptom checker development process (Healthdirect Australia, 2016). Gibson (Gibson et al., 2018) advises Healthdirect Australia uses CareEnhance Call Centre software as their clinical decision support system, and it is not apparent if this is also utilised by the NHS.

NHS is trialling four different symptom checkers. The NHS Pathways trialled in West Yorkshire uses the same clinical content and algorithms as the 111-telephone service, with the same clinical governance and assurance processes. This would appear to be the more traditional nursing triage guideline model. Sense.ly supply the West Midlands 'Ask NHS' app and Northern Ireland's 'NHS 111' system and utilises a clinical decision support system software platform called Odyssey owned by UK based provider Advanced. Odyssey uses a probability score system based on Bayesian logic rather than the standard decision-making tree (Armstrong, 2018). This logic employs inferential statistics that deals with probability inference which uses prior knowledge to predict future events (Rouse, 2006). Alex Yeates, the medical director from Advanced, explains Odyssey is not like most decision tree algorithms where once the consumer is on a certain branch (of the decision-making tree) it is difficult to shift to another branch. By using Odyssey's probability scores, questions can instead be asked in any order (Armstrong, 2018). The Expert-24 program used in the Suffolk pilot also uses probability scores based on Bayesian logic.
Babylon Health provides the fourth NHS online app in North West London and their system utilises AI algorithms and machine learning. Babylon Health's symptom checker chatbot app has recently come under scrutiny as it appeared patients were manipulating the system to secure appointments with a GP expediently (Fraser, Coiera, & Wong, 2018). No independent peer-reviewed clinical data has been published to support their claim that their symptom checker 'Babylon Check' outperforms physicians and nurses in diagnosing and triaging clinical vignettes (McCartney, 2017). As previously discussed, according to two articles posted on arXiv in 2016 and 2018, Babylon Check gave congruent triage advice in 88% of vignettes, versus 76% for doctors and 74% for nurses in a role-play scenario with the identical vignettes, and the symptom checker gave advice that was clinically safe in 100% of scenarios (Middleton et al., 2016). The sensitivity (or recall) of the symptom checker in providing diagnostic advise was 80% versus the physician's average of ~84% accuracy. They define sensitivity as "the proportion of relevant diseases that are included in the differential" (Razzaki et al., 2018, p. 3). This implies there is not one correct answer to the vignette (as in this research) however, there is no list of vignettes and 'acceptable' differential

diagnoses attached. The number of differentials offered by the program or physicians are not specified in the article, nor is it clear whether the appropriate disease needed to be listed first, or simply appear somewhere in the differential list. Precision is then defined as "the proportion of the diseases in the differential that are relevant. A precision of one hundred percent would be achieved if the differential diagnosis contained only the disease modelled by the vignette" (Razzaki et al., 2018, p. 3). The symptom checker's precision for diagnosis was 0.8% higher (44%) than the physicians (Razzaki et al., 2018). Babylon Health's research determined that their program provided triage dispositions, or 'consultations' more expediently than humans did. The symptom checker 'consultations' took an average of 1.07 minutes to complete, which was significantly faster than doctors' consultations who averaged 3.12 minutes (p = <0.001) and nurses' consultations which averaged 2.27 minutes (p = <0.001). Babylon Health has promoted this as providing a "practical route to cost savings in healthcare systems" (Middleton et al., 2016, p. 6; Razzaki et al., 2018). It is unclear from this article if physicians and nurses where purely providing a triage disposition during their 'consult'. The consideration here is what 'value-add' physicians and nurses might have provided during their consultations, explanations may have been provided or questions answered, that might not occur within a symptom checker program.

5.5 Emerging trends

5.5.1 AI decision-making

Alex Yeates from Advanced advises a problem with AI decision-making is that it is inscrutable, it is not possible to know how the computer program arrived at the decision it did, whereas with Odyssey he can explain exactly how a decision was made (Armstrong, 2018). AI limited memory programs are dependent on feedback mechanisms to 'learn' and improve the answers or choices the program makes (Mathur, 2018). By comparing thousands of similar medical cases and diagnostic decisions, an AI symptom checker can rapidly produce an unbiased and impassionate decision based on extensive statistical information stored in its databank. The question remains as to where these feedback mechanisms originate from. If the program links to formal medical records with healthcare centres or hospitals, the feedback regarding diagnosis and triage will increase accuracy and appropriateness. At the end of Buoy Health's diagnostic process, they ask, "Were your results reasonable?" and show the three differential diagnoses given for that transaction, with thumbs up and thumbs down symbols next to each differential

diagnosis. The consumer is then able to select the thumbs up or down option for each potential diagnosis. Bouy Health (2018) then inquire "Which do you think is the right answer?*"* with the three differentials shown, as well as 'something else' and 'unsure' as additional options. Again, the consumer has the choice to select the answer they felt was correct for their situation. If this is a feedback mechanism for their AI program, this is concerning, as any response a consumer provides at this point is undiagnosed by a physician and merely a consumer's own opinion. Also, there may be an inherent bias when it comes to diagnosis that a consumer might prefer. Research indicates mental health diagnoses often expose consumers to stigma and may cause an individual to reject a diagnosis (Perkins et al., 2018; Watts, 2018). There needs to be transparency in how AI symptom checkers are formulating their diagnostic and triage information, so the consumer and healthcare professionals can understand where decisions and advice given by these programs are derived from (Australian Medical Association, 2014; Bates, Landman, & Levine, 2018). Research regarding Buoy Health suggests consumers may choose different care pathways post encounter with their symptom checker than was originally intended, however, without linking the online tool to a patient's healthcare record (or linking via other means), it cannot be determined if consumer behaviour is actually being influenced by online advice (Winn, Somai, Fergestrom, & Crotty, 2019). Given Buoy Health (2018) provides three differential diagnoses and three corresponding triage suggestions which can vary considerably regarding care urgency, the appropriateness of the triage advice selected by the consumer remains unassessed.

5.5.2 'Crowd diagnosis'

Recently a study was undertaken using public posts on social media site Reddit (Nobles et al., 2019). Reddit runs more than 200 health forums, one of which is a subreddit where consumers can discuss issues, concerns and stories pertaining to sexually transmitted infections (STIs). Analysis of a random 500 posts found that 58% of posts were seeking a diagnosis on social media, otherwise known as 'crowd diagnosis'. Furthermore, 20% of these diagnostic requests were seeking a second opinion following receipt of a formal diagnosis from a healthcare professional. The benefits for using such sites for 'crowd diagnosis' is the ability to gain multiple opinions quickly, while the consumer remains relatively anonymous. The concern with using social media for diagnosis is that there is no guarantee the respondents to such requests have any evidence-based knowledge or medical training to

support their opinions. A suggestion from Nobles et al. (2019) was to couple such sites with appropriate health care professionals to leverage off the benefits of social media whilst ensuring consumer safety (Nobles et al., 2019). This also exemplifies the necessity for oversight and regulation of health-related websites, and not just symptom checkers; where does the line get drawn between 'health information' versus 'diagnosis', and where is the accountability? Crowd diagnosis could have serious consequences for an untreated STI, including cervical cancer, pelvic inflammatory disease and infertility (Porth & Gaspard, 2015).

However, the ability to connect consumers through social media world-wide is a valuable asset. People with rare and unusual conditions can connect with others similarly diagnosed almost anywhere on the planet, providing moral support and the ability to compare management and treatment of their condition. Rare becomes less rare when viewed thorough a global lens. On Facebook alone, there are thirteen groups relating to multiple sclerosis ("Facebook Multiple Sclerosis Groups," n.d.). Social media sites and telecommunication networks tend to be dynamic, with constant shifts in the members partaking in discourse and flow of conversations, however often organised by a 'core' group of people (Sekara, Stopczynski, & Lehmann, 2016). According to Sekara et al., (2016, p. 9977) "core [individuals] exhibit a pattern of recurring meetings across weeks and months, each with varying degrees of regularity." So even within the apparent chaos of social interaction there is a degree of pattern or order to events over time. It has been found that the number of 'friends' a consumer has on Facebook is associated with perceived social support, which in turn was associated with lower levels of stress, less illness and improved wellbeing (Nabi, Prestin, & So, 2013). Social media and electronic interactions are therefore affecting how well supported individuals feel and influencing their feelings of psychological wellbeing.

5.5.3 Contemporary patient-centred care and use of information technology

Medicine is shifting to more a personalised and patient-centred care (Delaney, 2018). Current computing power allows for rapid exchange of scientific data and up-to-date evidence-based 'best practice' being shared nationally and internationally, benefiting the advancement of health science immensely (Fieschi, 2002; Kankanhalli, Hahn, Tan, & Gao, 2016). By way of an example, there is a movement towards genomic medicine to be integrated into everyday healthcare transforming the delivery of medicine to Australians (Australian Genomic Health Alliance,

n.d.). Genomic medicine is an emerging discipline which uses an individual's genomic information as part of their clinical care. For instance, specific mutations in BRCA1 and BRCA2 genes may increase a person's risk of breast and ovarian cancer relative to the general population. The BRCA testing is a predictive test, not a diagnostic one. However, having this knowledge can potentially improve the long-term outcomes for the consumer by ensuring they follow appropriate risk management and prevention strategies like undertaking annual mammograms (Pruthi, Gostout, & Lindor, 2010). Issues concerning 'big data' (a large volume of complex information) such as storage, privacy, ethics, patient access and national practices being consistent are some of challenges that the Australian Genomic Health Alliance face. It is expected genomic medicine will improve patient outcomes via expeditious diagnosis, prevention, early intervention and targeted therapy (Australian Genomic Health Alliance, n.d.).

If large organisations and Governments can utilise information technology and international networking to improve healthcare, it is not such a surprise that individuals are attempting to do the same thing. There is perhaps potential to tap into the dynamics of social media, combined with the knowledge base of certified health care clinicians to allow for 'crowd sourced diagnosis'. Imagine the potential of an AI symptom checker which runs like Wikipedia, where individuals updated the symptom checker following clinical diagnosis. The feedback loops informing the AI system would then match consumer symptoms with an accurate diagnosis.

5.5.4 Additional functionality within symptom checker programs

The NHS (2017) states all four symptom checker systems trialled allow online booking of appointments with after-hours medical centres and consumers can receive a call back from a 111 clinician where required. Ultimately consumers will be able to access their medical records and order prescriptions via online symptom checkers, integrating with other parts of the healthcare system (Limb, 2016). Healthdirect Australia's (2018c) symptom checker provides printable copies of their online evaluation for the consumer, while their after hours GP Helpline allows the online GP to email a summary of the call to the consumers regular GP if the consumer desires. Symcat (2019) offers pricing guides for the various triage options suggested (both with and without medical insurance) and has a link to book a medical appointment via 'ZocDoc' at a clinic local to the consumer. This kind of collaboration with the wider health industry is an aspect of symptom checkers the AMA also recommends; the ability to provide

continuity of care with the consumer's regular GP by providing an updated event summary following receiving digital advice (Australian Medical Association, 2014).

5.6 Regulation of symptom checkers and health-related sites

5.6.1 Guidelines and frameworks

As previously discussed, symptom checkers are not covered by the TGA in Australia nor the FDA in the USA as they are not classified as a medical device. In the UK, the Medicines and Healthcare products Regulatory Agency has analogous regulations pertaining to medical apps which provide 'general recommendations to seek further advice,' and therefore are not subject to certification there either (Burgess, 2017). Healthdirect Australia complies with standards suggested by the Health on the Net (HON) foundation (Health on the Net, 2019), which aims to promote transparent and reliable health information, and their endorsement is in small print at the foot of Healthdirect Australia's home page. The HON foundation has links to the World Health Organization, and over 8000 sites have been given certification worldwide. The FDA have also recently introduced voluntary software pre-certification in an effort to provide consumers with a level of security and safety in the health information received from digital technologies (Department of Health and Ageing, 2011, 2013; US Food and Drug Administration, 2019), however approval can be a time-consuming and expensive task for a company to undertake (Bates et al., 2018; Thompson, 2019) which would deter many companies when it is not mandatory to obtain pre-certification. Additionally, the FDA only has oversight for software products that affect American citizens, not other countries. Essentially anyone can design, build and publish a medical website or app with no relevant credentials. Before a medication is released to the marketplace, rigorous testing for patient safety and effectiveness is undertaken, with FDA/TGA approval being mandatory (Bates et al., 2018) however many healthcare websites and mobile applications undertake no scrutiny before being released to the market. The eHealth marketplace is truly global, with products being marketed in one country and the 'back end logic' such as IP addresses operating elsewhere in the world. For example, Symptomate's (2019a, 2019b, 2019c) symptom checker is available in English, Spanish, French, Dutch, Polish, Russian and Chinese, with the technology provided by Infermedica, a Polish company. Kotlo, Muragundi, Ligade, & Udupa (2015) observed that Indian software developers should target local and international markets (particularly European markets) for mobile medical apps and suggest regulation of such

products in India is necessary for consumer safety. Furthermore, it was observed many developers designing such programs have no medical background, necessitating appropriate guidelines. General frameworks for computerised diagnostic decision support programs exist (Fraser et al., 2018) however without regulatory oversight, developers have little incentive to follow these guidelines. Fraser et al. (2018) suggests regulatory guidelines should follow the development of such programs from inception through to operational use, as depicted in Table 5.3 below. This multistage process should evaluate symptom checkers as they are increasingly exposed to the marketplace; testing such factors as consumer useability, effectiveness and safety. Particularly, real-world outcomes like how consumers interact with the system, how they interpret information given by the program and act upon its advice should be subject to ongoing evaluation. This last step in the process appears to be under researched which is why regulation is essential, to ensure consumer safety, effectiveness of the programs and determine the cost to the healthcare system.

Table 5.3: Guidelines for evaluating symptom checkers based on frameworks in health informatics and mobile health

Please refer to the following reference to review the chart for Table 5.3:

Fraser, H., Coiera, E., & Wong, D. (2018). Safety of patient-facing digital symptom checkers. *The Lancet*, *392*(10161), 2263–2264. doi:10.1016/S0140-6736(18)32819-8, p.2263.

(Fraser et al., 2018, p. 2263).

Sites such as Reddit and online blogs pose a conundrum for the medical and health industry, as these sites are largely unregulated and potentially pose a safety risk for consumers, not too dissimilar to symptom checkers. There

is a need for further research in this area to understand how to adequately protect consumers without diluting the benefits of social media and the promotion of health information.

5.6.2 Policy considerations

Bates et al. (2018) suggests four policy issues need addressing concerning medical apps: firstly, the safety of apps must be established to ensure consumers are protected; second, a directory of evidence should be accessible to clinicians and consumers to determine the effectiveness of the app; third, apps should be connected to electronic health records to allow transfer of information and lastly, policies should be introduced to encourage apps to be developed that improve care and value, especially in areas where it is most desirable, for instance with chronic diseases and high-cost care patients. Currently apps have no 'rating' from a reputable source (i.e.: government agency), and instead rely on a 'five-star' rating determined by consumer feedback (Bates et al., 2018; Khalid, Nagappan, & Hassan, 2015). Bates et al. (2018) advises apps are largely designed for the 'worried well' and not for consumers with health issues, high-cost care or poor health literacy. Presently, there is a gap between what is actually required by the healthcare industry in terms of meaningful apps which make a difference to individual's health and what is being supplied.

The US Department of Health and Human Services' Office of Disease Prevention and Health Promotion have introduced a set of objectives to improve the health of Americans over the next decade, known as 'Healthy People'. Included in these objectives are two measures pertaining to website quality: the first relates to the improvement of information reliability and the second for improved website usability (Devine, Broderick, Harris, Wu, & Hilfiker, 2016). Analysis of US health websites includes (but is not limited to) criteria such as content design and development, privacy, content updating, site design, and information architecture. A 2015 review determined 58% of US websites evaluated met three out of six criteria for information reliability. Only two sites out of 100 met all six criteria, whilst one site met none of the criteria at all. Regarding website usability, 42% of websites reviewed met ten out of 19 usability principles. No sites met more than 14 of the principles, and all sites met at least one of the principles. Whilst 'Healthy People' now provide a standardised approach to measuring quality web-based health information for US sites, it does not appear to enforce or regulate these guidelines. Like many such guidelines and

recommendations, these criteria are limited to sites operating from the US and do not account for the vast volume of sites operating from foreign locations. Their results demonstrate an improvement is necessary as there were significant deficiencies in multiple criteria under both website usability and reliability categories, including disclosure of authorship and funding; differentiation between original site content and advertising; provision of search and print functionality; and minimising medical and technical jargon (Devine et al., 2016). Additionally, Devine et al. (2016) observe that 90% of Americans struggle with health literacy, affecting consumer's understanding of web-based information and services.

5.6.3 Governance of eHealth and mHealth information

Google has recognised there is a necessity for oversight of their eHealth information and are endeavouring to correct misinformation posted on their search engine (Rege, 2019) except without an official 'gateway' process being instigated by search engines or app stores, this will continue to be an ongoing concern. The NHS has initiated an NHS Apps Library (Gann, 2012; NHS England, n.d.-b, 2017) where links to sites supported and approved by the NHS can be located. There are a set of standards that must be met in order to appear in the NHS Apps Library which gives consumers some degree of assurance when selecting from these sites that the information is from reputable sources. However, there is clearly a disclaimer from the NHS advising the app developers are solely responsible for the 'advertisement, compliance and fitness for purpose' of their product and the NHS is not liable for their use (NHS England, n.d.-b).

In line with other research (Lupton & Jutel, 2015; Semigran et al., 2015), some of the symptom checkers reviewed were found to lack concrete information pertaining to editors, authors or validity of information (Hamilton & Brady, 2012; Jutel & Lupton, 2015; Lupton & Jutel, 2015; Rosser & Eccleston, 2011). All symptom checkers tested in this research had a caveat on their website or app that specified use of the symptom checker is for informational purposes only and does not replace the need to consult a physician, or wording to similar effect. Buoy Health state they have used thousands of clinical papers to inform their AI program (Baum & News, 2017), whilst others—What's My Diagnosis (n.d.) is an example—give no explanation where their clinical information has been derived, how often this information is updated, nor give any indication that the authors or editors having any medical

background. Online eHealth products should be coupled with credentialed experts to ensure validity and currency of health information and advice (Nundy, Montgomery, & Wachter, 2019). Boon-itt (2019) determined details such as reliable content, contemporary information, facts regarding authors and user-friendliness all contribute towards the perceived quality of a health-related website from a user perspective. The perceived quality of a website in turn influences the consumers intention to trust the technology and use the site to seek health information. The question remains, is consumer perceived quality a reliable indicator of genuine quality? Accordingly, a suitably qualified and independent body needs to endorse online products to ensure they meet predetermined criteria before being released to the general public.

5.7 Other considerations

5.7.1 Health literacy

As alluded to previously, the ability for consumers to read and accurately interpret health information is becoming a necessary skill in today's society (Cheng & Dunn, 2015). The World Health Organization defines health literacy as "…the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health" (Nutbeam, 1998, p. 357). Australia has an ageing population and more chronic diseases are being self-managed (Cheng & Dunn, 2015). For consumers to appropriately manage their health and treatments options, a certain level of health literacy is essential.

Individuals with poor health literacy may be less discerning of where health information is sourced from and be more inclined to trust the internet than those individuals with stronger health literacy (Diviani, van den Putte, Giani, & van Weert, 2015). An Australian study found that there were few easily understood online health materials, which could lead to consumers misunderstanding information and making inappropriate healthcare choices (Cheng & Dunn, 2015). Accordingly, internet sources should be easy to locate, be evidence based, be edited and authored by appropriately qualified personnel, and be written to be understandable by the bulk of the population. This is critical as the internet has become a popular source of health information.

5.7.2 The holistic approach

One of the main differences between a computer program and a medical practitioner is the medical practitioner's ability to view the patient holistically, to consider the patient history, comorbidities, medications currently and previously taken and the patient's social circumstances (Murtagh & Rosenblatt, 2015, p. 3). Additionally, a physician is aware of what services and resources are available locally and their hours of operation. If a physician should see a patient late in the day, they may refer the patient to the ED for ultrasound or x-ray if local services have closed for the day, effectively 'over-triaging', as this will address the patient's needs most expediently. A symptom checker does not typically provide this level of detail in their assessment and responses. A computer program is only as good as the programming and the quality of information input into the system. Given the case described previously when a patient was convinced he had renal calculi (Avery, Ghandi, & Keating, 2012), a consumer's own 'self-diagnosis' could skew information provided to a symptom checker. However, a physician can explore a patient's signs and symptoms in-depth especially when something does not appear to add up. If a patient believed they were suffering from anxiety for instance, this could colour the answers given to a symptom checker and potentially lead down a mental health pathway rather than an algorithm for a physiological disorder; for example, an underlying heart condition. The resultant triage disposition and acuity for anxiety may be quite different than one for a cardiac disorder. How a patient narrates their signs and symptoms and provides a chronological order of events allows the physician to infer potential differential diagnoses from their evidence-based knowledge and potentially apply some specific testing to narrow the diagnostic possibilities. The diagnostic process is more than a listing of symptoms (Baerheim, 2001). The example given previously concerning Mayo Clinic's (2018) responses to the malaria vignette demonstrates how even the selection of a chief complaint can significantly change a symptom checker's diagnostic output. This suggests that currently symptom checker algorithms cannot adapt to the complexity of real-life patients and fall short of the 'gold standard' of physician diagnosis, but in comparison to a generalised web-based search for information, symptom checkers appear superior (Berry et al., 2019).

Nonetheless, technology can aid a clinical practice and should most certainly not be completely disregarded. My Health Record is a secure online repository for an individual's health information established by the Australian Government in 2016 (Australian Digital Health Agency, n.d.). This allows medical practitioners to view a patient's

medical history even if they are not regularly treating this person. For instance, if someone was involved in a motor vehicle accident an ED physician would be able to access the patient's My Health Record and be aware of their existing medical conditions and allergies, medications and current treatments. This information helps inform the ED physician concerning treatment and management of their patient. It may be useful in the future to be able to link consumers' electronic health records to symptom checker outputs. How consumers interpret and act upon online information is an essential component of research lacking in many symptom checkers exposed to the marketplace. Real world outcomes and impacts from online advice is poorly understood to-date and data linkage with consumer health records would provide the data on consumer behaviours and will provide the understanding of how they are being influenced by online tools. Such linkage could (with permission from the consumer) also allow clinicians such as GPs to see the symptoms that have been entered by the patient into the symptom checkers. The difficulty with this concept is the likelihood of an overwhelming (and possibly unnecessary) amount of information kept on electronic health records. GPs would not have the time to review every output provided by symptom checkers for all their consumers. At this stage, Healthdirect Australia's (2018d, 2018b) symptom checker provides an option for consumers to print out their transaction to take to a clinician's appointment, which appears an appropriate solution currently.

5.7.3 Trust in eHealth

An essential component to the functioning of a healthcare system is trust, and reciprocal trust between a patient and physician aids treatment adherence, improvements in self-reported health and provides an improved patient experience (Nundy et al., 2019). There are three components to trust; positive relationships, good judgement/expertise and consistency (Zenger, Jack; Folkman, 2019). For an online tool a 'positive relationship' might equate to the consumer's experience of the tool; for example, how easy it was to use or navigate the program, how informative the site was, the technical issues which might arise when the product is used, or how the service might link to other aspects of the healthcare system. 'Expertise' might represent the credibility of editors and authors, confirmation of evidence-based information, accurate and timely advice or suggesting useful and relevant healthcare pathways for consumers. 'Consistency' in a program might require uniformity in the responses provided for the identical scenario, and some similarity to a physician's 'gold standard' diagnosis or disposition.

The results for this research suggest there is room for improvement across all three components of trust before healthcare professionals will have much faith in these online products, with perhaps valid concern for patient safety given some serious instances of under-triage in this evaluation (Table 4.8). Boon-itt (2019) suggests consumer's interaction with symptom checkers is partially based on perception: "perceived information quality can motivate users to have trust in order to use a health website as their source of information" (Boon-itt, 2019, p. 15). The level of trust consumers place in such online tools may be more than is necessarily prudent, as the surface impressions, or perceptions, of a site does not necessarily mean that the site is a quality program or information is from a quality source. Medical apps aimed at consumers may be viewed as an authoritative source and given the wide variation in accuracy of information from these programs, this is potentially an issue. Whereas programs aimed at physicians are part of the clinician's 'tool kit', which helps inform a clinician's decision. Symptom checkers may pose challenges for the traditional patient-physician relationship, as consumers may accept the authority of the internet over a physician's, particularly when a physician's opinion may differ from the consumer's own (Lupton & Jutel, 2015). Conversely, other research suggests more positive behaviours might evolve from using online resources, as outlined below.

5.7.4 Positive health behaviours

Research conducted on Microsoft's extended campus in Washington found internet behaviour was affected by consumers searching the web for health information on undiagnosed conditions (White & Horvitz, 2009). Just under 40% of participants agreed their behaviour had changed; increasing searches on the internet for potentially serious conditions; increased viewing of webpages which described their perceived condition; and engagement with physicians and specialists had increased (White & Horvitz, 2009). It could be argued that internet searches may convince patients to present to their GP earlier than they would otherwise if they had not researched health information online, hastening a diagnosis that may have been delayed. Equally, it could be argued that more cases might present to GPs unnecessarily due to cyberchondria, and more research would be required to determine which is the case. The same project identified that participants also used the web following receipt of a diagnosis for reassurance and to help them understand the medical terminology used by their physician (White & Horvitz, 2009). Thus, this would work to improve health literacy and could facilitate stronger patient-physician relationships

rather than erode them. Consumers becoming more educated and empowered about their own health (Lupton & Jutel, 2015)(Healthdirect Australia, 2016) and being directed to the appropriate healthcare service for their needs positively impacts the healthcare system.

5.7.5 Benefits of symptom checkers and public health

There are many other advantages. Epidemiological information can be tracked via symptom checker websites providing valuable timely data to healthcare professionals. WebMD (2017) runs weekly analysis on the cold and flu season in the USA, which is updated every Sunday night at midnight. The information on WebMD's searches had a 95% correlation to the Centres for Disease Control and Prevention's (CDC) data. WebMD also powers an interactive Cold and Flu map showing real time information, tracking cold and flu symptoms by the entire country, by state, or can drill down to a county level (PR Newswire, 2018). In the UK a retrospective observational study was undertaken, where it was determined that online symptom checkers seem superior to call centres in providing early warning on public health syndromic indicators and especially useful for cold and flu symptoms, breathing difficulties and eye conditions (Berry et al., 2019; Elliot et al., 2015). Via syndromic surveillance, diseases can be monitored for variations from the normal deviations of epidemiological patterns allowing healthcare services to respond to outbreaks promptly and allocate additional resources to minimise potential outbreaks (Berry et al., 2019).

A further benefit is that resource poor nations and underserved rural regions can utilise online tools to bolster their limited resources as health professionals are sparse in these areas (Kotlo et al., 2015; Morita, Rahman, Hasegawa, Ozaki, & Tanimoto, 2017). Consumers would be able to access health information and receive advice 24/7, potentially improving health outcomes and diverting patients to appropriate local services. In a similar vein, certain demographic populations like the elderly and chronically ill consumer could be targeted, where there is a greater medical need (Powell, Landman, & Bates, 2014). Some symptom checkers can link directly to healthcare services to facilitate access to appropriate facilities, with Symcat (Symcat, 2019) and the UK already pioneering these options. Healthdirect Australia (Healthdirect Australia, 2018c) currently provides an extensive database of services which includes locations and hours of operation for pharmacies, medical centres, hospitals and EDs (Taylor, 2014).

Over time, these digital resources will be increasing utilised, and it would be beneficial for appropriate checks and balances to be in situ for all programs containing medical content.

5.7.6 Future considerations for symptom checkers

In 2015, Semigran et al. (2015) tested the 'first generation' of symptom checkers, finding many limitations with such tools. They have been touted as reducing traffic in both primary care and EDs, but this has not been apparent in regard to ED presentations in Australia (Ng et al., 2012). In the subsequent four years AI technology has introduced a new generation of programs, and these products did perform better than their counterparts in this evaluation. Of particular interest, are the 'feedback mechanisms' used to inform the AI program, helping them improve their performances over time. It is unclear where these AI programs obtain their feedback from (consumers, healthcare professionals or elsewhere). If from the consumer, this is a dubious source of authority, as they have already been unsure enough about their condition to consult an online tool, so would not be considered a medical expert. Their assessment is also likely to be based on other factors which might be unrelated to the accuracy of diagnosis. This research determined the same limitations Semigran et al. (2015) found still typically apply. Symptom checkers are still risk-averse, and they vary widely with regard to diagnostic accuracy. With nearly two-thirds of the first listed diagnoses incorrect, and only 58% accuracy within the top 10 reported options, their specificity is no better than random chance. An incorrect diagnosis could cause unnecessary appointments with healthcare providers or potentially delay treatment if a consumer was advised their symptoms were self-resolving.

While some over-triage might be expected, especially regarding borderline cases, ideally self-care should be recommended when it is appropriate, as should primary care pathways instead of the ED. There will always be marginal instances where pneumonia may be considered an emergency rather than urgent care, especially with at risk populations such as the elderly, however not all cases of pneumonia should be automatically sent to the ED. A symptom checker needs to have the algorithms incorporated to consider all the variables for each consumer and give the appropriate care advice. There are considerable cost savings in online symptom checkers, but the economies from promoting online tools would be counter-productive if consistent over-triage caused increases in

unnecessary visits to the GP or ED. However, over-triage may be infinitely preferable to under-triage, due to the potentially serious consequences of not seeking care soon enough. Optimally a balance would be established, rather than having all symptom checkers perceived as being risk-averse, contributing to further burden on an already overburdened healthcare system (Toloo et al., 2011).

While it is not known if consumers would obtain the same results (since these evaluations were done under 'laboratory conditions'), the health and technological literacy required to navigate these programs may provide significant challenges for some consumers causing further variability with results. Based on research into consumer adherance with nursing triage call centre advice, it is possible consumers may not follow advice given by a symptom checker either, if it differs from their own beliefs.

5.8 Brief recap on the comparisons to Harvard Medical School's research

The 36-symptom checkers evaluated provided overall diagnostic accuracy of 36% and triage accuracy of 49%. These results are closely aligned to previous research conducted by Semigran et al. (2015), indicating the medical terminology they used may not have affected their results appreciably. However, evaluating the programs using layperson's language was an important decision, as this is likely to reflect how the general public would use these products. The medical terminology used by some symptom checkers for questioning and provision of health information can potentially be challenging for consumers with poor health literacy. It is well documented that health information should be targeted to a Grade 8 level; however, multiple studies have found online sites to read two to four grades higher than this. This has implications for how consumers interpret and comprehend health advice, and further research should be conducted to investigate this subject. The triage categories in this research were expanded to four (emergency, urgent, non-urgent and self-care), and there was some reclassification of vignettes to better reflect the Australian healthcare system. The vignettes also included some newly created scenarios describing diseases endemic to Australia. Another noteworthy difference between Semigran et al.'s (2015) research and this evaluation was how many differential diagnoses were reviewed before an 'incorrect' diagnosis would be assigned. Previous research allowed the top 20 differentials to be considered, whereas this research reduced this number to the top ten differentials, based on the large number of differential diagnoses offered by

some websites. Furthermore, this research allowed more than one 'correct' diagnosis in some vignettes due to the similarity of symptoms with multiple conditions. Nevertheless, the results from prior research and the results from this evaluation are notably consistent.

5.9 Strengths and limitations of this research

There are some limitations to this research. It is conceivable not every symptom checker available to the Australian general public was identified during the search process, however the exhaustive systematic approach to identifying symptom checkers covering several popular search engines should have captured well those most commonly utilised. The fourth and fifth search engines did not identify any new symptom checkers, indicating saturation had likely been achieved.

The vignettes used were primarily simple scenarios with few comorbidities for the patient included in the script. Mostly, these vignettes contained no or minimal prior medical history, and current or previous medication use was rare, not necessarily reflecting the complexity of actual patients. The simplicity of these vignettes may have inflated the accuracy of the symptom checkers reviewed, as actual patients may suffer from comorbidities and take multiple medications, making any diagnostic process more complex. Incorporating many of the same vignettes used by Semigran et al. (2015) allows a degree of comparison with previous research, whilst the inclusion of new 'Australianised' vignettes makes this research relevant for the Australian community. Australia has distinctive flora and fauna, as well as some unique diseases not found in other parts of the world. Therefore, it is worthwhile testing international symptom checker programs to determine their utility for the Australian general public.

Another limitation was that no gold standard was robustly established during this auditing process. Previous research (Semigran et al., 2015) suggested nursing triage call centres as a comparator, however research by Ng et al. (2012) indicates the appropriateness of nursing triage call centre referrals to ED are similar to patient self-referrals, while GPs provided the most appropriate referrals to ED. As the algorithms for nursing call centres are similar to many online symptom checkers, this may not be providing a 'gold standard' comparison. The argument to compare symptom checkers to nursing call centres is becoming obsolete with the introduction of AI symptom

checker programs which learn as they go, no longer adhering to strict decision tree logic that nursing call centres utilise. Also, given Ng et al. (2012) research which indicates nursing call centre's appropriateness of referrals to the ED is similar to the general public's, the bar is not appreciably raised by comparison with call centres. The true gold standard would be comparison to a physician's diagnosis or disposition, as this would be the most accurate assessment process at this point in time. Semigran et al. (2016) subsequently conducted a research project assessing medical practitioners' accuracy with diagnosis using the same vignettes they had on the symptom checkers. Their results determined that physician's overall accuracy in listing the correct diagnosis first was 72% and listing the correct diagnosis within the top three options was 84%. Further research would be required to determine the accuracy of medical practitioners with this collection of vignettes, creating a gold standard to compare this evaluation of symptom checker outcomes.

Testing was conducted under 'laboratory conditions' with only one-person inputting data into the symptom checkers, which is both a limitation and a strength, the latter because this ensured consistency with data entry between programs, guaranteeing symptoms and subsequent questioning was kept as uniform as possible. While the researcher was not familiar with the specific websites and apps, a degree of proficiency in navigating various online programs was gained during this process, therefore it is also a limitation of this study design. It is possible the general public may have greater difficulty navigating certain programs than the researcher did, if they are not sufficiently computer literate. The researcher was also familiar with the clinical vignettes, which may introduce a certain user bias. This research was unable to predict how consumers would use the tool, especially given the variation of situations and health literacy in the real world. However, it demonstrates that even when there are standardised scenarios and a single user of the tools, there is enormous variation in the results obtained. It is therefore likely that when consumers use the tool in the real world, the results could be even more varied and unreliable. The vignettes were categorised into common and uncommon for this very reason.

A further limitation is that the process used to determine whether symptom checkers used AI or not was imprecise. The information was taken from what was available for a consumer to read on each website or app. It was beyond the scope of this study to contact site owners to investigate exactly what type of AI their product was (if they would

even divulge this level of information), and then interpret whether this process actually met the definition of AI or not. We chose to limit our comparisons to those sites clearly claiming to make use of AI versus those that did not.

A strength of this research design was the use of layperson's language to test the programs. These programs are designed to be used by the general public, not medical personnel, therefore should be tested using layperson's language rather than medical terminology. Given that health literacy is cited as being lower than online health information is being written (Cheng & Dunn, 2015), it is vital to an individual's comprehension and adherence with advice that online tools pitch their information at an appropriate level and use appropriate language. This had been identified as a limitation with Semigran et al.'s (2015) research which has been addressed in this research.

5.10 References

Aaronson, M. (2011). *Common things are common, except when the diagnosis is rare.* Retrieved November 26, 2019, from https://www.kevinmd.com/blog/2011/01/common-common-diagnosis-rare.html

AARP Health Tools. (2018). *Health Tools / Symptoms*. Retrieved February 7, 2019, from https://healthtools.aarp.org/symptomsearch#

Abimanyi-Ochom, J., Bohingamu Mudiyanselage, S., Catchpool, M., Firipis, M., Wanni Arachchige Dona, S., & Watts, J. J. (2019). Strategies to reduce diagnostic errors: a systematic review. *BMC Medical Informatics and Decision Making, 19*(1), 1–1. https://doi.org/10.1186/s12911-019-0901-1 https://ecu.on.worldcat.org/oclc/8217796233

Armstrong, S. (2018). The apps attempting to transfer NHS 111 online. *BMJ*, *360,* 156. doi:10.1136/bmj.k156. Retrieved from http://www.bmj.com/content/360/bmj.k156.abstract

Ashwath, M. L., & Gandhi, S. (2018). *ST-elevation myocardial infarction.* Retrieved November 20, 2019, from https://bestpractice-bmj-com.ezproxy.ecu.edu.au/topics/en-gb/150/pdf/150.pdf

Asthma and Allergy Foundation of America. (2019). *Asthma facts and figures.* Retrieved November 05, 2019, from https://www.aafa.org/asthma-facts/

Australian Digital Health Agency. (n.d.). *My Health Record.* Retrieved January 30, 2020, from https://www.myhealthrecord.gov.au/

Australian Genomic Health Alliance. (n.d.). *Integrating genomics into healthcare.* Retrieved January 30, 2020, from https://www.australiangenomics.org.au/about-us/australian-genomics/

Australian Medical Association. (2014). *Call Centre Triage and Advice Services - 2004.* Revised 2014. Retrieved October 9, 2019, from https://ama.com.au/position-statement/call-centre-triage-and-advice-services-2004-revised-2014

Avery, N., Ghandi, J., & Keating, J. (2012). The 'Dr Google'phenomenon–missed appendicitis. *NZ Med J*, *125*(1367), 135–137.

Baerheim, A. (2001). The diagnostic process in general practice: has it a two-phase structure? *Family Practice, 18*(3), 243–245.

Balogh, E., Miller, B. T., & Ball, J. (2015). *Improving diagnosis in health care*. National Academies Press Washington, DC. Retrieved from https://ebookcentral.proquest.com/lib/ecu/reader.action?docID=4393813

Bates, D. W., Landman, A., & Levine, D. M. (2018). Health apps and health policy: what is needed? *Jama*,

320(19), 1975–1976. doi:10.1001/jama.2018.14378

Baum, S., & News, M. (2017). *Buoy health raises $6.7M for smarter triage tool aimed at hospitals and payers*. Retrieved April 14, 2018, from https://medcitynews.com/2017/08/buoy-health-raises-6-7m-smarter-triage-tool-aimed-hospitals-payers/

Berry, A. C., Cash, B. D., Wang, B., Mulekar, M. S., Van Haneghan, A. B., Yuquimpo, K., … Green, W. K. (2019). Online symptom checker diagnostic and triage accuracy for HIV and hepatitis C. *Epidemiology & Infection*, *147*, 104. doi:10.1017/S0950268819000268

Boon-itt, S. (2019). Quality of health websites and their influence on perceived usefulness, trust and intention to use: an analysis from Thailand. *Journal of Innovation and Entrepreneurship*, *8*(1), 4.

Brain Foundation. (2017). *Stroke.* Retrieved November 20, 2019, from https://brainfoundation.org.au/disorders/stroke/

Breivik, H., Borchgrevink, P. C., Allen, S. M., Rosseland, L. A., Romundstad, L., Breivik Hals, E. K., … Stubhaug, A. (2008). Assessment of pain. *BJA: British Journal of Anaesthesia*, *101*(1), 17–24.

Buckby, E. (2016). *How to confuse a foreigner: American vs. British English.* Retrieved November 20, 2019, from https://www.communicaid.com/business-language-courses/blog/uk-usa-two-countries-divided-common-language/

Buoy Health. (2018). *Buoy*. Retrieved December 15, 2018, from https://www.buoyhealth.com/

Burgess, M. (2017). *Can you really trust the medical apps on your phone?* Retrieved October 9, 2019, from https://www.wired.co.uk/article/health-apps-test-ada-yourmd-babylon-accuracy

Canto, A. J., Kiefe, C. I., Goldberg, R. J., Rogers, W. J., Peterson, E. D., Wenger, N. K., … Canto, J. G. (2012). Differences in symptom presentation and hospital mortality according to type of acute myocardial infarction. *The American Heart Journal*, *163*(4), 572–579. https://doi.org/10.1016/j.ahj.2012.01.020 https://ecu.on.worldcat.org/oclc/5568692978

Cheng, C., & Dunn, M. (2015). Health literacy and the Internet: a study on the readability of Australian online health information. *Australian and New Zealand Journal of Public Health*, *39*(4), 309–314. https://doi.org/10.1111/1753-6405.12341

Cooke, G., Valenti, L., Glasziou, P., & Britt, H. (2013). Common general practice presentations and publication frequency. *Australian Family Physician*, *42*, 65–68.

Delaney, L. J. (2018). Patient-centred care as an approach to improving health care in Australia. *Collegian*, *25*(1), 119–123. doi:10.1016/j.colegn.2017.02.005

Department of Health and Ageing. (2011). *Australian regulatory guidelines for medical devices (ARGMD*). Retrieved April 13, 2018, from https://www.tga.gov.au/sites/default/files/devices-argmd-01.pdf

Department of Health and Ageing. (2013). *Regulation of medical software and mobile medical 'apps'*. Retrieved April 13, 2018, from https://www.tga.gov.au/regulation-medical-software-and-mobile-medical-apps

Devine, T., Broderick, J., Harris, L. M., Wu, H., & Hilfiker, S. W. (2016). Making quality health websites a national public health priority: toward quality standards. *Journal of Medical Internet Research*, *18*(8), e211. doi:10.2196/jmir.599

Diviani, N., van den Putte, B., Giani, S., & van Weert, J. C. M. (2015). Low Health Literacy and Evaluation of Online Health Information: A Systematic Review of the Literature. *Journal of Medical Internet Research*, *17*(5), e112. https://doi.org/10.2196/jmir.4018

Doctor Diagnose Symptom. (2018). *Doctor Diagnose Symptoms Check.* (Version 1.0.5) [Mobile application software]. Retrieved January 17, 2019, from https://play.google.com/store/apps/details?id=com.appcolliders.doctordiagnose

Doherty-Torstrick, E. R., Walton, K. E., & Fallon, B. A. (2016). Cyberchondria: parsing health anxiety from online behavior. *Psychosomatics*, *57*(4), 390–400. doi:10.1016/j.psym.2016.02.002

Drugs.com. (2019a). Symptom Checker. (Version 2.9.3) [Mobile application software]. Retrieved January 5, 2019, from https://play.google.com/store/apps/details?id=com.drugscom.app

Drugs.com. (2019b). *Symptom Checker.* Retrieved January 5, 2019, from https://www.drugs.com/symptom-checker/

Elliot, A. J., Kara, E. O., Loveridge, P., Bawa, Z., Morbey, R. A., Moth, M., … Smith, G. E. (2015). Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system. *Epidemiology & Infection*, *143*(16), 3416–3422. doi:10.1017/S0950268815000503

ePain Assist. (2018). *Symptom Checker.* (Version 1.5.5) [Mobile application software]. Retrieved December 7, 2018, from https://play.google.com/store/apps/details?id=com.epainassist.symptomchecker

Everyday Health. (2019). *Symptom Checker.* Retrieved January 8, 2019, from https://www.everydayhealth.com/symptom-checker/

Facebook. (n.d.). *Facebook Multiple Sclerosis Groups.* Retrieved January 30, 2020, from https://www.facebook.com/search/top/?q=multiple sclerosis&epa=SEARCH_BOX

Family Doctor. (2018). *Symptom Checker.* Retrieved November 24, 2019, from https://familydoctor.org/your-health-resources/health-tools/symptom-checker

Ferrini, R., Mannino, E., Ramsdell, E., & Hill, L. (1996). Screening mammography for breast cancer: American College of Preventive Medicine practice policy statement. *American Journal of Preventive Medicine*, *12*(5), 340–341.

Fieschi, M. (2002). Information technology is changing the way society sees health care delivery. *International Journal of Medical Informatics*, *66*(1–3), 85–93.

Fraser, H., Coiera, E., & Wong, D. (2018). Safety of patient-facing digital symptom checkers. *The Lancet*, *392*(10161), 2263–2264. doi:10.1016/S0140-6736(18)32819-8

Gann, B. (2012). Giving patients choice and control: health informatics on the patient journey. *Yearbook of Medical Informatics*, *21*(01), 70–73.

Gardiner, L., Cullen, M., Karabatsos, G., & St George, I. (n.d.). Universal telenursing triage in Australia and New Zealand: A new primary health service. *Australian Family Physician, 37*(6), 476–479.

Gibson, A., Randall, D., Tran, D. T., Byrne, M., Lawler, A., Havard, A., … Jorm, L. R. (2018). Emergency Department Attendance after Telephone Triage: A Population-Based Data Linkage Study. *Health Services Research*, *53*(2), 1137–1162. doi:10.1111/1475-6773.12692

Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety, 22,* 27. doi:10.1136/bmjqs-2012-001615

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, *165*(13), 1493–1499.

Gregory, P. Bs., Ward, A., & Sanders, M. J. (2010). *Sanders' paramedic textbook.* London: Wolfe Medical Publications.

Hamilton, A. D., & Brady, R. R. W. (2012). Medical professional involvement in smartphone 'apps' in dermatology. *British Journal of Dermatology*, *167*(1), 220–221. doi:10.1111/j.1365-2133.2012.10844.x

Hauer, S. (2018, March 10). *Symptom checker uses artificial intelligence.* Retrieved March 20, 2019, from http://ezproxy.ecu.edu.au/login?url=https://search-proquest-com.ezproxy.ecu.edu.au/docview/2012399945?accountid=10675

Health on the Net. (2019). *Health on the Net.* Retrieved February 10, 2019, from https://www.hon.ch/home.html

Healthdirect Australia. (2016). *About Healthdirect Symptom Checker.* Retrieved February 2, 2019, from https://www.healthdirect.gov.au/about-healthdirect-symptom-checker

Healthdirect Australia. (2017). *Annual Report 2016 – 2017 Healthdirect Australia*. Retrieved April 14, 2018, from

https://media.healthdirect.org.au/publications/Annual_Reports_Business_Highlights_2016-2017.pdf

Healthdirect Australia. (2018a). *Asthma statistics*. Retrieved November 5, 2019, from https://www.healthdirect.gov.au/asthma-statistics

Healthdirect Australia. (2018b). *Check Your Symptoms.* (Version 2.2.1) [Mobile application software]. Retrieved from https://apps.apple.com/au/app/healthdirect/id1021494621

Healthdirect Australia. (2018c). *Healthdirect Australia Annual Report Business Highlights 2017-2018*. Retrieved from http://media.healthdirect.org.au/publications/healthdirect-australia-annual-report_17-18_business-highlights.pdf

Healthdirect Australia. (2018d). *Symptom Checker.* Retrieved December 18, 2018, from https://www.healthdirect.gov.au/symptom-checker

Healthdirect Australia. (2019). *Flu trends in Australia*. Retrieved December 1, 2019, from https://www.healthdirect.gov.au/flu-trends-in-australia

Healthline. (2018). *Symptom Checker.* Retrieved November 22, 2018, from https://www.healthline.com/symptom-checker

Heneghan, C., Glasziou, P., Thompson, M., Rose, P., Balla, J., Lasserson, D., … Perera, R. (2009). Diagnostic strategies used in primary care. *Bmj*, *338*, b946.

Isabel Healthcare. (2018). *Isabel.* Retrieved November 1, 2018, from https://symptomchecker.isabelhealthcare.com/suggest_diagnoses_advanced/landing_page

Jutel, A., & Lupton, D. (2015). Digitizing diagnosis: a review of mobile applications in the diagnostic process. *Diagnosis, 2*(2), 89–96 https://doi.org/10.1515/dx-2014-0068

Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: introduction to the special section. *Information Systems Frontiers*, *18*(2), 233–235. doi:10.1007/s10796-016-9641-2

Khalid, H., Nagappan, M., & Hassan, A. E. (2015). Examining the relationship between FindBugs warnings and app ratings. *IEEE Software*, *33*(4), 34–39.

Kotlo, A., Muragundi, P. M., Ligade, V. S., & Udupa, N. (2015). Regulation of the mobile medical apps in india: Need of the hour. *Current Pharma Research*, *5*(4), 1600.

Lanseng, E. J., & Andreassen, T. W. (2007). Electronic healthcare: a study of people's readiness and attitude toward performing self-diagnosis. *International Journal of Service Industry Management*, *18*(4), 394–417.

Limb, M. (2016). *NHS announces online symptom checker. BMJ (clinical research ed.) 354.*4905. doi:101136/bmj.i4905

Lind, S. (2018). *Increasing number of 111 calls handled by clinicians, says NHS England.* Retrieved January 10, 2020, from http://www.pulsetoday.co.uk/news/commissioning/commissioning-topics/urgent-care/increasing-number-of-111-calls-handled-by-clinicians-says-nhs-england/20037359.article

Lovette, L. (2018). *AI triage chatbots trekking toward a standard of care despite criticism*. Retrieved October 20, 2019, from https://www.mobihealthnews.com/content/ai-triage-chatbots-trekking-toward-standard-care-despite-criticism

Lupton, D., & Jutel, A. (2015). 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Social Science & Medicine*, *133*, 128–135. https://doi.org/10.1016/j.socscimed.2015.04.004

Marklund, B., Ström, M., Månsson, J., Borgquist, L., Baigi, A., & Fridlund, B. (2007). Computer-supported telephone nurse triage: an evaluation of medical quality and costs. *Journal of Nursing Management*, *15*(2), 180–187. doi:10.1111/j.1365-2834.2007.00659.x

Mathur, A. (2018). *Create an AI feedback loop with Continuous Relevancy Training in Watson Discovery*. Retrieved October 10, 2019, from https://developer.ibm.com/blogs/create-an-ai-feedback-loop-with-watson-discovery/

Mayo Clinic. (2018). *Symptom Checker*. Retrieved November 15, 2018, from https://www.mayoclinic.org/symptom-checker/select-symptom/itt-20009075

McCartney, M. (2017). Margaret McCartney: Innovation without sufficient evidence is a disservice to all. *Bmj*, *358*, j3980. doi:10.1136/bmj.j3980

MedicineNet. (2018). *Symptom Checker:* Symptoms & Signs A-Z. Retrieved November 28, 2018, from https://www.medicinenet.com/symptoms_and_signs/symptomchecker.htm#introView

Michie, S., Yardley, L., West, R., Patrick, K., & Greaves, F. (2017). Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *Journal of Medical Internet Research*, *19*(6), 232. doi:10.2196/jmir.7126

Middleton, K., Butt, M., Hammerla, N., Hamblin, S., Mehta, K., & Parsa, A. (2016). Sorting out symptoms: design and evaluation of the "babylon check" automated triage system*. arXiv preprint arXiv:1606.02041.* Retrieved April 04, 2018, from https://arxiv.org/abs/1606.02041

Morita, T., Rahman, A., Hasegawa, T., Ozaki, A., & Tanimoto, T. (2017). The potential possibility of symptom checker. *International Journal of Health Policy and Management*, *6*(10), 615–616. doi:10.15171/ijhpm.2017.41

Mosby. (2016). *Mosby's Medical Dictionary.* (10th ed.). St Louis: Elsevier Health Sciences. Retrieved from http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4747147

Murtagh, J., & Rosenblatt, J. (2015). *John Murtagh's general practice* (Sixth edit). North Ryde, NSW : McGraw-Hill Education. Retrieved from https://murtagh.mhmedical.com/book.aspx?bookid=1522

My Health Alberta. (2018). *Check Your Symptoms.* Retrieved December 12, 2018, from https://myhealth.alberta.ca/Health/pages/symptom-checker.aspx

Nabi, R. L., Prestin, A., & So, J. (2013). Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being. *Cyberpsychology, Behavior, and Social Networking*, *16*(10), 721–727. doi:10.1089/cyber.2012.0521

New South Wales Government. (2005). *Pain matters: Red and yellow flags.* Retrieved November 20, 2019, from http://www.hnehealth.nsw.gov.au/Pain/Documents/red and yellow flags.pdf

Ng, J. Y., Fatovich, D. M., Turner, V. F., Wurmel, J. A., Skevington, S. A., & Phillips, M. R. (2012). Appropriateness of healthdirect referrals to the emergency department compared with self-referrals and GP referrals. *Medical Journal of Australia*, *197*(9), 498–502. doi:10.5694/mja12.10689

NHS England. (n.d.-a). *NHS 111 online.* Retrieved December 1, 2018, from www.nhs.uk/Conditions/Pages/hub.aspx

NHS England. (n.d.-b). *NHS Apps Library.* Retrieved October 26, 2019, from https://www.nhs.uk/apps-library/?page=1

NHS England. (2017). *NHS111 Online evaluation.* Retrieved October 9, 2019, from https://askmygp.uk/wp-content/uploads/111-Online-Evaluation-DRAFT_.pdf

Nobles, A. L., Leas, E. C., Althouse, B. M., Dredze, M., Longhurst, C. A., Smith, D. M., & Ayers, J. W. (2019). Requests for Diagnoses of Sexually Transmitted Diseases on a Social Media Platform. *JAMA*, *322*(17), 1712. https://doi.org/10.1001/jama.2019.14390

North, F., Varkey, P., Laing, B., Cha, S. S., & Tulledge-Scheitel, S. (2011). Are e-health Web users looking for different symptom information than callers to triage centers? *Telemedicine and E-Health: The Official Journal of the American Telemedicine Association,*, *17*(1), 19–24. doi:10.1089/tmj.2010.0120.

Ntaios, G. (2018). *Ischaemic Stroke.* Retrieved November 20, 2019, from https://bestpractice-bmj-com.ezproxy.ecu.edu.au/topics/en-gb/1078/differentials

Nundy, S., Montgomery, T., & Wachter, R. M. (2019). Promoting trust between patients and physicians in the era of artificial intelligence. *Jama*, *322*(6), 497–498. doi:10.1001/jama.2018.20563

Nutbeam, D. (1998). Health promotion glossary. *Health Promotion International*, *13*(4), 349–364.

Patient.info. (n.d.). *Symptom Checker*. Retrieved November 23, 2018, from https://patient.info/symptom-checker

Perkins, A., Ridler, J., Browes, D., Peryer, G., Notley, C., & Hackmann, C. (2018). Experiencing mental health diagnosis: a systematic review of service user, clinician, and carer perspectives across clinical settings. *The Lancet Psychiatry*, *5*(9), 747–764. doi:10.1016/S2215-0366(18)30095-6

Picardo, J. (2012). *Why students need a global awareness and understanding of other cultures*. Retrieved November 20, 2019, from https://www.theguardian.com/teacher-network/2012/sep/25/students-global-awareness-other-cultures

Porth, C., & Gaspard, K. (2015). *Essentials of pathophysiology: Concepts of altered health states* (Fourth Edi). Philadelphia: Wolters Kluwer Health.

Powell, A. C., Landman, A. B., & Bates, D. W. (2014). In search of a few good apps. *Jama*, *311*(18), 1851–1852. doi:10.1001/jama.2014.2564

PR Newswire. (2018). *Searches for Cold and Flu Symptoms Up Sharply on WebMD's Symptom Checker™ As Consumers Look for Guidance; Most Recent Data Suggest Flu Rates Are on the Decline.* Retrieved March 20, 2019, from https://www.prnewswire.com/news-releases/searches-for-cold-and-flu-symptoms-up-sharply-on-webmds-symptom-checker-as-consumers-look-for-guidance-most-recent-data-suggest-flu-rates-are-on-the-decline-300605273.html

Pruthi, S., Gostout, B. S., & Lindor, N. M. (2010). Identification and Management of Women With BRCA Mutations or Hereditary Predisposition for Breast and Ovarian Cancer. *Mayo Clinic Proceedings*, *85*(12), 1111–1120.

Ragoschke-Schumm, A., Walter, S., Haass, A., Balucani, C., Lesmeister, M., Nasreldein, A., … Grunwald, I. Q. (2014). Translation of the 'time is brain'concept into clinical practice: focus on prehospital stroke management. *International Journal of Stroke*, *9*(3), 333–340.

Rare Voices Australia. (n.d.). Uniting Austalians living with rare disease. Retrieved January 30, 2020, from https://www.rarevoices.org.au/

Razzaki, S., Baker, A., Perov, Y., Middleton, K., Baxter, J., Mullarkey, D., … Majeed, A. (2018). A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *ArXiv Preprint ArXiv:1806.10698*.

Rege, A. (2019). *Dr. David Feinberg: Google's plan to fix fake news in healthcare.* Retrieved September 10, 2019, from https://www.beckershospitalreview.com/hospital-management-administration/dr-david-feinberg-google-s-plan-to-fix-fake-news-in-healthcare.html

Richter, T., Nestler-Parr, S., Babela, R., Khan, Z. M., Tesoro, T., Molsen, E., & Hughes, D. A. (2015). Rare disease terminology and definitions—a systematic global review: report of the ISPOR rare disease special interest group. *Value in Health*, *18*(6), 906–914. doi:10.1016/j.jval.2015.05.008

Right Diagnosis. (2011). *Symptom Checker.* Retrieved December 20, 2018, from https://symptoms.rightdiagnosis.com/

Rosser, B. A., & Eccleston, C. (2011). Smartphone applications for pain management. *Journal of Telemedicine and Telecare*, *17*(6), 308–312. doi:10.1258/jtt.2011.101102

Rouse, M. (2006). *Bayesian logic.* Retrieved from https://whatis.techtarget.com/definition/Bayesian-logic

Sanders, M. J., Lewis, L. M., & Quick, G. (2012). *Mosby's paramedic textbook*. Jones & Bartlett Publishers.

Saver, J. L. (2006). Time is brain—quantified. *Stroke*, *37*(1), 263–266.

Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, *113*(36), 9977–9982.

Semigran, H. L., Levine, D. M., Nundy, S., & Mehrotra, A. (2016). Comparison of physician and computer diagnostic accuracy. *JAMA Internal Medicine*, *176*(12), 1860–1861.

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ (clinical research ed.)*, *351,* 3480. doi:10.1136/bmj.h3480

Simone Baldovino, M. D., Domenica Taruscio, M. D., & Dario Roccatello, M. D. (2016). Rare diseases in Europe: from a wide to a local perspective. *Israel Medical Association Journal*, *18*(6), 359–363.

Singh, H., Meyer, A. N. D., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Quality & Safety*, *23*(9), 727–731. doi:10.1136/bmjqs-2013-002627

Smith, C. J. (2012). Diagnostic tests (1)–sensitivity and specificity. *BMJ: British Medical Journal, 308*(6943), 1552–1552.

Symcat. (2019). *What is bothering you today?* Retrieved January 14, 2019, from www.symcat.com

Symptify. (2018). *How are you feeling today?* (Version 2) [Mobile application software]. Retrieved January 3, 2019, from https://play.google.com/store

Symptomate. (2019a). Symptomate. (Version 2.1) [Mobile application software]. Retrieved January 8, 2019, from

https://apps.apple.com/au/app/symptomate/id837725433

Symptomate. (2019b). *Symptomate.* Retrieved January 8, 2019, from https://symptomate.com/diagnosis/

Taylor, H. (2014). *SOA Software Enables Healthdirect Australia to Deliver Anywhere Access to Health Data.* Retrieved March 10, 2018, from http://www.qlifepro.com/press/20140826-20-15920/ca-soa-software/

Thompson, B. M. (2019). *How many companies stand to benefit from the FDA Pre-Cert program? Fewer than you might think.* Retrieved September 1, 2019, from https://www.mobihealthnews.com/content/how-many-companies-stand-benefit-fda-pre-cert-program-fewer-you-might-think

Toloo, S., FitzGerald, G., Aitken, P., Ting, J., Tippett, V., & Chu, K. (2011). Emergency health services: demand and service delivery models. Monograph 1: literature review and activity trends.

U.S. Department of Health & Human Services. (2018). *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement.* Retrieved from https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata1.html

University of Michigan. (2019). *Symptom Checker.* Retrieved January 4, 2019, from http://www.uofmhealth.org/health-library/sx

US Food and Drug Administration. (2019). *Developing a software precertification program: A working model.* Retrieved May 14, 2019, from https://www.fda.gov/media/119722/download

Watts, J. (2018). *Mental health labels can save lives. But they can also destroy them*. Retrieved February 8, 2020, from https://www.theguardian.com/commentisfree/2018/apr/24/mental-health-labels-diagnosis-study-psychiatrists

WebMD. (2017). *WebMD Symptom Checker*. Retrieved November 14, 2018, from https://symptoms.webmd.com/default.htm

What's My Diagnosis. (n.d.). *What's My Diagnosis.* Retrieved January 9, 2019, from www.whatsmydiagnosis.com

White, R. W., & Horvitz, E. (2009). Experiences with web search on medical concerns and self diagnosis. In *AMIA annual symposium proceedings* (Vol. 2009, p. 696). American Medical Informatics Association.

Winn, A. N., Somai, M., Fergestrom, N., & Crotty, B. H. (2019). Association of Use of Online Symptom Checkers With Patients' Plans for Seeking Care. *JAMA Network Open*, *2*(12), 1918561. doi:10.1001/jamanetworkopen.2019.18561

Wolf, J. A., Moreau, J. F., Akilov, O., Patton, T., English, J. C., Ho, J., & Ferris, L. K. (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatology*, *149*(4), 422–426.

Zenger, Jack; Folkman, J. (2019). *The 3 elements of trust.* Retrieved October 17, 2019 from https://hbr.org/2019/02/the-3-elements-of-trust

CHAPTER 6

CONCLUSION

Thesis Map
Chapter One – Introduction
Chapter Two – Literature Review
Chapter Three – Methodology
Chapter Four – Results
Chapter Five – Discussion
➡ Chapter Six – Conclusion

6.1 Research questions reviewed

In Chapter One, the following research questions were posed:

1) Are free online English language symptom checkers available to the Australian general public clinically accurate? Is the correct diagnosis listed as the first diagnosis, within the top three diagnoses, or within the top ten diagnoses?
2) Are free online English language symptom checkers which are currently available to the Australian general public providing accurate triage advice?
3) How do international English websites and apps most readily available to the Australian public perform in comparison to the Australian Government funded Healthdirect Australia symptom checker?

In response to the first question, as reported in Chapter Four, the correct diagnosis was listed first in 36% of SCVTs, within the top three in 52% of cases and within the top ten in 58% of instances. Individual symptom checker performance for the first listed diagnosis varied widely, from 12% to 61%. These figures suggest diagnostic symptom checkers should be used by the Australian general public with caution, as their clinical accuracy is inferior to diagnostic advice provided by a healthcare professional. There was no Australian diagnostic symptom checker, and international programs were found to not provide information on Australian specific conditions.

Secondly, in terms of triage advice, overall accuracy was 49% from all SCVTs. Vignettes which represented emergency scenarios were the most concordant with our own assessment, at 63%, followed by urgent scenarios at 56%. This fell to 30% with respect to non-urgent vignettes and rose again to 40% for self-care SCVTs. Again,

the triage advice offered from free online symptom checkers should be viewed with caution by the Australian general public. International symptom checkers sometimes suggested triage options not available within the Australian healthcare system, affecting their utility for this market. Additionally, some symptom checkers did not provide a timeframe in which treatment should be sought. Triage is not only about where to seek assistance; it also encompasses how urgently assistance is required.

The last question addressed how well international symptom checkers performed against Australia's only symptom checker, Healthdirect Australia's triage website and app. Healthdirect Australia's symptom checkers were the leading programs on both platforms, with 61% concordance with our own triage assessments. Other individual symptom checkers performance varied from 17% to 58% accuracy. Healthdirect Australia is funded by the Australian Government, with information and advice offered being relevant to the Australian community.

6.2 Conclusion

The Australian general public are actively engaged with electronic media, seeking healthcare advice from multiple online organisations. Symptom checker websites and apps are a frequent source of healthcare guidance providing differential diagnosis and triage advice for health consumers. According to Healthdirect Australia approximately 80% of Australians seek healthcare information online with half of those investigating self-treatment options. This represents approximately 20.3 million Australians investigating their healthcare alternatives on the internet. Healthdirect Australia reported 23.6 million visits to their combined websites for the 2016/2017 year, indicating their popularity with the Australian public. It is, however, unknown how many international symptom checker websites Australians visit per annum and the quality of the information gleaned.

The results from this evaluation strongly align to prior research conducted by the Harvard Medical School (Semigran et al., 2015), reinforcing their findings that symptom checkers tend to be risk-averse and may not be directing consumers to the correct healthcare pathways as a result. The inclusion of Australian specific vignettes provided an interesting insight, in that no international symptom checkers offered Australian diseases in their differentials list (for example Ross River Virus), indicating they may not provide information and advice completely

relevant for Australian consumers. In terms of triage, international symptom checkers occasionally recommended options not available to Australian consumers under our Medicare Scheme. The introduction of AI programming algorithms into symptom checkers has meant direct comparisons to nursing triage call centres may no longer be appropriate. Many AI programs learn 'on the job' and do not adhere to strict decision tree guidelines that call centres follow. This has created a certain degree of unease in some quarters, as it may be more difficult to rationalise a decision made by an AI program; whereas a program which follows a decision tree can easily have the output explained. In this evaluation, AI programs outperformed non-AI symptom checkers under all diagnostic classifications (first listed diagnosis, top three and top ten listed differentials). For triage, there was no significant differences found. Whilst determination of AI programs was imprecise in this evaluation, AI programs may be worthy of further research, particularly in regard to their feedback mechanisms which help to inform their output. AI is a burgeoning business, and it is likely to increase penetration into the global digital market. As shown in the UK, online health tools are cheaper to operate than nursing triage call centres, which has considerable financial implications for the NHS. This is why the UK has a purposeful channel-shift policy towards online technology. The NHS anticipates overworked parts of the healthcare system such as the 999-call centre will be liberated by online symptom checkers and consumers will be diverted to more appropriate services for their needs.

While the usability of the symptom checkers was not specifically tested, certain concerns came to light during evaluation. Some symptom checkers failed to 'red flag' concerning symptoms, which poses serious risks for the consumer. Other sites used UK or USA spelling, Fahrenheit instead of Celsius, recognise different influenza seasons, or used unfamiliar drug names affecting the usability of such systems for Australian consumers. Many programs do not use pain scales and do not ask questions regarding demographics, prior medical history and medication use, which are standard questions for any health professional during an initial consultation. Such issues with digital tools have led to growing interest in health regulation and guidelines for website useability and information reliability of online health sites. Unregulated information is a serious concern for the health industry and information consumers receive needs to be relevant and appropriate for the Australian population and healthcare system. While symptom checkers may not fall under FDA/TGA guideline for diagnostic tools, other criteria should be established globally to ensure published information is current and evidence based, and from suitably qualified authors.

There are potential advantages to a well-designed symptom checker, however. Resource poor nations and remote rural regions can gain instant 24/7 access to diagnostic and triage advice and appropriate health information. Some programs currently integrate with other parts of the healthcare system, where consumers can book appointments online through the symptom checker. These programs present both an opportunity to enhance or diminish the traditional patient-physician relationship, depending on the quality of information and educational material furnished by the provider. With consumers becoming more educated, active and informed regarding their healthcare, online tools will become an increasingly popular resource. Consumers in Australia and the UK are using symptom checkers in greater numbers than previously to gain diagnostic and triage information, which may not be accurate or appropriate. However, symptom checkers are likely to provide more accurate advice than a general Google search. The anonymity and convenience of these digital tools may in fact encourage some consumers to seek care earlier than they might have otherwise.

6.3 Recommendations

Recommendations for future development and maintenance of symptom checkers would include:

1) the establishment of international standards and guidelines for online health tools. While the development of compulsory international standards may be years away, Governments can ensure certain standards are met nationally. Regulatory framework can be formulated from existing guidelines such as those depicted in Table 5.1, which would provide consistency and reliability with symptom checkers at a domestic level.
2) Adherence with HON standards should not be voluntary and Government policy needs to reinforce this, to ensure healthcare advice is to an acceptable minimum standard.
3) Online products such as symptom checkers should have consumer input, as suggested by previous research. A symptom checker informed by consumer feedback would have care for the medical language used, aid health literacy and have an interface that was simple and intuitive to use.
4) All symptom checkers should be required to alert consumers to concerning 'red flag' symptoms and direct them to an appropriate care pathway, irrespective of whether the symptom checker provides

triage and/or a diagnosis. Consumer safety is a priority and should form part of these regulatory guidelines.

5) Transparency of how the product is developed and informed is also required. Editors and authors should be clearly identifiable with qualifications relevant to medical practice. AI programs which rely on feedback mechanisms to inform their program need to explain where they obtain their feedback from.

6.4 Concluding remarks

The results between this evaluation and Harvard Medical School's investigation are remarkably resilient. Despite methodological differences between Harvard Medical School's previous study and this research, the findings are very consistent between investigations. Diagnostic and triage advice suffer deficiencies when compared to the current 'gold standard' of a physician's disposition or diagnosis. Certain symptom checkers provide greater diagnostic and triage accuracy than others. More research is required to understand the differences between these programs and determine why some symptom checkers are superior. Ideally, all symptom checkers should perform optimally and be evaluated periodically to ensure 'best practice' is being provided to the consumer. The general public should be informed which online sites provide contemporary evidence-based information and appropriate advice, suitable for the Australian community.

Ongoing research and evaluation into symptom checkers will be required to determine how consumers actually use these programs and what impact such products have on the management of their healthcare concerns. Symptom checkers need to be from quality sources and incorporate algorithms that provide accurate diagnostic and/or triage advice. With an already over-taxed healthcare system and EDs frequently running at capacity, it is vital online tools like symptom checkers direct consumers to an appropriate point of care in a suitable timeframe, and information is suitable to the Australian community.

APPENDICES

Appendix A: Symptom Checkers Included in Study

Internet Based	Symptom Checker (citation)	Description	Included in Semigran et al. study	Maximum number of diagnoses	Triage options available
	AARP Health Tools (AARP-American Association of Retired Persons) (AARP Health Tools, 2018)	Non-profit, nonpartisan American organisation empowering people to choose how they live as they age. SC powered by Talix.	No	10	N/A
	APWUHP (American Postal Workers Union Health Plan) (APWU Health Plan, 2018)	The American Postal Workers Health Plan website. The SC redirects you to the AARP Health Tools site which is powered by Talix. Between testing the AARP site and APWUHP site, Health Tools had updated their SC. Results will therefore reflect the difference between old and new versions of the SC.	No	90	N/A
	Buoy Health (Buoy Health, 2018)	An American online SC using an artificial intelligence algorithm, backed by medical data.	No	3	N/A
	CHW.org (Children's Hospital of Wisconsin) (Children's Wisconsin, 2019)	Paediatrics only site. (Healthy Children; Hopkins; St Luke's; CHW identical; however, Children's Hospital of Wisconsin does not credit the AAPP for powering their site and has a slightly different interface to Healthy Children). American IP address.	No	N/A	- Call 911 now - Go to ER now - Call Dr. or seek care now - Call Dr. within 24 hours - Call Dr. during office hours - Self-care at home
	Drugs.com (Drugs.com, 2019a)	Online resource for drug and related health information. Information for the SC is provided by Harvard Health Publications. (American)	Yes	5	- Call 911 immediately - Emergency care now - Call for medical help immediately; you may need to be examined today - Call your doctor; contact your doctor or go to a local emergency room right away for evaluation; you need immediate medical evaluation - You should arrange an evaluation with your doctor - Self-treatment
	Everyday Health (Everyday Health, 2019)	An American lifestyle and medical health information site. Adults only site. SC powered by Infermedica in Poland. Uses artificial intelligence.	No	8	- Immediate medical attention recommended - Prompt medical attention recommended - May require medical attention
	Family Doctor (Family Doctor, 2018)	Powered by AAFP (American Academy of Family Physicians). A physician-reviewed SC.	Yes	5	- Emergency; call your doctor right away or go to the hospital – Urgent; call your doctor right away - Self-care

	Healthdirect Australia (Healthdirect, 2018b)	A national, Australian government owned not for profit organisation. Uses NHS algorithms.	No	N/A	- Call an ambulance - Go to the emergency department - See your GP today - Call a nurse immediately - See your GP in the next few days - Speak to a nurse - Speak to a nurse about your child - See your GP in the next few days - See your GP routinely - See an optometrist or GP soon - Speak to your pharmacist - Look after yourself at home
	Healthline (Healthline, 2018)	An American health information website.	Yes	172	N/A
	Healthwise (myhealth.alberta.ca) (My Health Alberta, 2018)	Nonprofit organisation in Canada which provides health education, IT solutions guidance to other organisations. SC accessed via Myhealth Alberta website. Also powers HealthLinkBC and Uofmhealth.	Yes	N/A	- Call 911 now - Seek care now - Seek care today - Make an appointment -Try home treatment
	Health Link BC (HealhLinkBC, 2018)	Provides access to non-emergency health information and advice in British Columbia, Canada. SC powered by Healthwise.	No	N/A	- Call 911 now - Seek care now - Seek care today - Make an appointment - Try home treatment
	Health Status (Health Status, n.d.)	Privately held American company, using the SC algorithms from Isabel, therefore uses artificial intelligence. General health and well-being site.	No	33	N/A
	Healthy Children (Healthy Children, 2018)	From the American Academy of Pediatrics. Paediatrics only site. (Healthy Children; Hopkins; St Lukes; CHW identical)	Yes	N/A	- Call 911 now - Go to ER now - Call Dr. or seek care now - Call Dr. within 24 hours - Call Dr. during office hours - Self-care at home
	Hopkins all children's (John Hopkins Children's Hospital) (Johns Hopkins All Chilren's Hospital, 2019)	Paediatrics only site. (Healthy Children; Hopkins; St Lukes; CHW identical; however, John Hopkins Children's Hospital does not credit the AAPP for powering their site and has a slightly different interface to Healthy Children). American IP address.	No	N/A	- Call 911 now - Go to ER now - Call Dr. or seek care now - Call Dr. within 24 hours - Call Dr. during office hours - Self-care at home
	Isabel Health Care (Isabel Healthcare, 2018)	SC is built using Artificial Intelligence searching technology. An American website. Provides Isabel SC to Health Status and Patient.Info.	Yes	10	- Walk in clinic / telemedicine - Family physician/ urgent care clinic/ minor injuries

					- Emergency medicine
	Mayo Clinic (Mayo Clinic, 2018)	Not for profit clinic in Minnesota USA. Integrates clinical practice, research and education.	Yes	15	N/A
	MedicineNet (MedicineNet, 2018)	Owned and operated by WebMD. Is an American online healthcare media publishing company.	No	110	N/A
	Patient.info (Patient.info, n.d.)	A UK based resource for medical staff and patients; evidence based clinical information for a global audience. Uses computerised algorithms. SC powered by Isabel, therefore uses artificial intelligence.	No	17	N/A
	Right Diagnosis (Right Diagnosis, 2011)	Independent online American medical health information website. SC powered by Health Grades Inc.	No	676	N/A
	RxList (RxList, 2018)	Founded by pharmacists, RxList is an American online medical resource. Acquired by WebMD in 2004.	No	110	N/A
	St Lukes online (St Luke's Online, n.d.)	Paediatrics only site. (Healthy Children; Hopkins; St Luke's; CHW identical; however, St Luke's does not credit the AAPP for powering their site and has a slightly different interface to Healthy Children). American IP address.	No	N/A	- Call 911 now - Go to ER now - Call Dr. or seek care now - Call Dr. within 24 hours - Call Dr. during office hours - Self-care at home
	Symcat.com (Symcat, 2019)	An American based triage tool. Symcat = System-based, Computer Assisted Triage; based on patient data records.	No	6	- Emergency Room - Urgent care - Primary care - Retail clinic
	Symptomate (Symptomate, 2019a)	An online health-oriented service developed by Infermedica in Poland. Adults only site. Uses artificial intelligence. (IP address is based in America however Symptomate is copyright of Infermedica).	Yes	8	- Immediate medical attention recommended, contacting a medical professional or emergency services is advised - Prompt medical attention recommended, consulting a medical professional is strongly advised - May require medical attention, symptoms you've reported are rarely sign of a severe condition. However, if you are concerned about your health or your symptoms persist or worsen, we advise to contact a medical professional
	Uofmhealth (University of Michigan) (University of Michigan, 2019)	On the University of Michigan (USA) webpage; there is a SC powered by Healthwise.	No	N/A	- Call 911 now - Seek care now - Seek care today - Make an appointment - Try home treatment

	WebMD (WebMD, 2017)	An American online information site pertaining to health and well-being. Owns and operates RxList and MedicineNet.	Yes	30	N/A
	What's My Diagnosis (What's My Diagnosis, n.d.)	An analytically driven medical diagnostic tool. Only 100 symptoms and diseases included. No information provided about the company or where it is based. IP address shows California, USA.	No	33	N/A
(n = 26)					
Mobile Apps	**Symptom Checker**	**Description**	**Included in Semigran et al. study**	**Maximum number of diagnoses**	**Triage options available**
	Doctor Diagnose Symptom Google Play (Doctor Diagnose Symptom, 2018)	Appcolliders built the Doctor Diagnose Symptoms Check; there is no information on who has designed and contributed to the App. The email address is appcolliders@gmail.com. There is a web page for Appcolliders, and that IP address is based in Isreal.	Yes	3	- Seek out medical assistance immediately - Consult your doctor immediately - Call a doctor immediately - Consult your doctor
	Drugs.com Google Play (Drugs.com, 2019b)	As above	Yes	5	- Call 911 immediately - Emergency care now - Call for medical help immediately; you may need to be examined today - Call your doctor - Contact your doctor or go to a local emergency room right away for evaluation - You need immediate medical evaluation; you should arrange an evaluation with your doctor - Self-treatment
	ePain Assist Symptom Checker Apple App (ePain Assist, 2018a)	General health information and pain App. The site gives no information on where it is based or who contributes to the page. Assume it is American based as they have an epainassist.com Facebook page which supplied no further information. Their webpage displays a contact address as St. Petersburg, USA.	No	13	N/A
	ePain Assist Symptom Checker Google Play (ePain Assist, 2018b)	General health information and pain App. The site gives no information on where it is based or who contributes to the page. Assume it is American based as they have an epainassist.com Facebook page which supplied no further information. Their webpage displays a contact address as St. Petersburg, USA.	No	13	N/A
	Healthdirect Apple App (Healthdirect, 2018a)	As above	No	N/A	- Call an ambulance - Go to the emergency department - See your GP today - Call a nurse immediately - See your GP in the next few days - Speak to a nurse

					- Speak to a nurse about your child - See your GP in the next few days - See your GP routinely - See an optometrist or GP soon - Speak to your pharmacist - Look after yourself at home
	Symptify Google Play (Symptify, 2018)	An American online health assessment tool using a customized algorithmic engine. Provides medical education, information, advice and facility locator. Can pre-notify care facilities of users impending arrival.	Yes	10	N/A
	Symptomate Apple App (Symptomate, 2019b)	As above	Yes	8	- Immediate medical attention recommended, contacting a medical professional or emergency services is advised - Prompt medical attention recommended, consulting a medical professional is strongly advised - May require medical attention, symptoms you've reported are rarely sign of a severe condition. However, if you are concerned about your health or your symptoms persist or worsen, we advise to contact a medical professional
	Symptomate Google Play (Symptomate, 2019c)	As above	Yes	8	- Immediate medical attention recommended, contacting a medical professional or emergency services is advised - Prompt medical attention recommended, consulting a medical professional is strongly advised - May require medical attention, symptoms you've reported are rarely sign of a severe condition. However, if you are concerned about your health or your symptoms persist or worsen, we advise to contact a medical professional
	WebMD Apple App (WebMD, 2018a)	As above	Yes	30	N/A
	WebMD Google Play (WebMD, 2018b)	As above	Yes	30	N/A
(n = 10)					

Appendix B: Clinical Vignettes

Diagnosis	Semigran et al.'s original vignette (citation)	Vignette for current research using layman's terminology (newly created scenarios for this research have no vignette, only signs and symptoms)	Common (Yes/No)	Is it likely the diagnosis will appear in the top 3 conditions? (Yes/No)
Requires emergency care (n=13)				
Acute liver failure (Acetaminophen poisoning acceptable). Triage could be emergent or urgent.	A 48-year-old woman with a history of migraine headaches presents to the emergency room with altered mental status over the last several hours. She was found by her husband, earlier in the day, to be acutely disoriented and increasingly somnolent. On physical examination, she has scleral icterus, mild right upper quadrant tenderness, and asterixis. Preliminary laboratory studies are notable for a serum ALT of 6498 units/L, total bilirubin of 5.6 mg/dL, and INR of 6.8. Her husband reports that she has consistently been taking pain medications and started taking additional 500 mg acetaminophen pills several days ago for lower back pain. Further history reveals a medication list with multiple acetaminophen-containing preparations. (Australian Institute of Health and Welfare, 2017; Semigran, Linder, Gidengil, & Mehrotra, 2015)	48-year-old female with history of migraines. Confused. Disoriented. Increasingly sleepy. Mild right upper abdomen tenderness. Has been on various pain medications and has been taking more paracetamol over the last few days because of lower back pain. Chief complaint (c/c) = confusion	Yes	No
Appendicitis	A 12-year-old girl presents with sudden-onset severe generalized abdominal pain associated with nausea, vomiting, and diarrhoea. On exam she appears ill and has a temperature of 104°F (40°C). Her abdomen is tense with generalised tenderness and guarding. No bowel sounds are present. (Australian Institute of Health and Welfare, 2017; Semigran et al., 2015)	12-year-old female. Sudden, severe abdominal pain. Feels sick (nausea). Had vomiting and diarrhoea. Temperature 40°C. c/c = abdominal pain	Yes	Yes
Asthma (Or acute asthma, status asthmaticus, exacerbation of asthma)	A 27-year-old woman with a history of moderate persistent asthma presents to the emergency room with progressive worsening of shortness of breath, wheezing, and cough over 3 days. She reports prior exposure to a person who had a runny nose and a hacking cough. She did not receive significant relief from her rescue inhaler with worsening symptoms, despite increased use. She has been compliant with her maintenance asthma regimen, which consists of an inhaled corticosteroid and a leukotriene receptor antagonist for maintenance therapy and albuterol as rescue therapy. Her cough is disrupting her sleep pattern and as a consequence she is experiencing daytime somnolence, which is affecting her job performance. (Cooke, Valenti, Glasziou, & Britt, 2013; Semigran et al., 2015)	27-year-old female. History of asthma. Progressively worsening shortness of breath, wheezing and coughing over 3 days. Not responding to inhalers. Cough disturbing her sleep and is sleepy in the daytime. c/c = shortness of breath/ breathing problem	Yes	Yes
COPD exacerbation (severe)	A 67-year-old woman with a history of COPD presents with 3 days of worsening dyspnea and increased frequency of coughing. Her cough is now	67-year-old female. History of Chronic Obstructive Airways Disease (COPD). 3	Yes	Yes

(Diagnosis may include acute bronchitis) Triage could be emergent or urgent.	productive of green, purulent sputum. The patient has a 100-pack-year history of smoking. She has had intermittent, low-grade fever of 100°F (37.7°C) for the past 3 days and her appetite is poor. She has required increased use of rescue bronchodilator therapy in addition to her maintenance medications to control symptoms. (Britt et al., 2016; Semigran et al., 2015)	days of worsening shortness of breath and cough with green phlegm. Mild fever. (37.7°C) Poor appetite. Increased use of bronchodilator/Salbutamol in addition to normal treatment. Smoker (if asked). c/c = shortness of breath/ breathing problem		
Haemolytic uremic syndrome	A 4-year-old boy presents with a 7-day history of abdominal pain and watery diarrhoea that became bloody after the first day. Three days before the onset of symptoms, he had visited the county fair with his family and had eaten a hamburger. Physical examination reveals a mild anemia (Semigran et al., 2015; Victoria State Government, 2014)	4-year-old male. 7 days of stomach pain. Blood in diarrhoea. (Only if asked) - Ate a hamburger at a fair 3 days before stomach-ache began. c/c = Diarrhoea	No	No
Myocardial infarction (Heart Attack)	Mr. Y is a 64-year-old Chinese male who presents with chest pain for 24 hours. One day prior to presentation, the patient began to experience 8/10, non-radiating substernal chest pressure associated with diaphoresis and shortness of breath. The pain initially improved with Tylenol, however over the following 24 hours, his symptoms worsened. The patient went to his primary physician, where an EKG was performed which showed ST elevation in leads V2-V6. (Britt et al., 2016; Semigran et al., 2015)	64-year-old male. 1 day of chest pain. (8/10 pain). Pain does not move elsewhere. Sweating. Breathless. Feels tightness in mid chest (include if asked, as this was an added symptom for Semigran et al.'s research). c/c = Chest pain	Yes	Yes
Ureteric stones (Kidney stones)	A 45-year-old white man presents to the emergency department with a 1-hour history of sudden onset of left-sided flank pain radiating down toward his groin. The patient is writhing in pain, which is unrelieved by position. He also complains of nausea and vomiting. (Semigran et al., 2015; Victoria State Government, 2018)	45-year-old male. 1 hour of severe left sided lower back pain radiating to groin. Feels sick. Vomiting. Pain unrelieved by position. c/c = Back pain	Yes	Yes
Malaria	A 28-year-old man presents to his physician with a 5-day history of fever, chills, and rigors, not improving with acetaminophen (paracetamol), along with diarrhea. He had been traveling in Central America for 3 months, returning 8 weeks ago. He had been bitten by mosquitoes on multiple occasions, and although he initially took malaria prophylaxis, he discontinued it due to mild nausea. He does not know the specifics of his prophylactic therapy. On examination he has a temperature of 100.4°F (38°C) and is mildly tachycardic with a BP of 126/82 mmHg. The remainder of the examination is normal. (Semigran et al., 2015; Victoria State Government, 2015)	28-year-old male. 5 days of chills, shivering, diarrhea. (Only if asked by symptom checker - Recently been overseas to New Guinea. Was bitten by mosquitos. Didn't regularly take anti-malaria medication). c/c = fever	No	No
Meningitis	An 18-year-old male student presents with severe headache and fever that he has had for 3 days. Examination reveals fever, photophobia, and neck stiffness. (Menigicoccal Australia, n.d.; Semigran et al., 2015)	18-year-old male. 3 days of severe headache. Fever, sore neck and light sensitivity.	Yes	Yes

		c/c = headache		
Pulmonary embolism	A 65-year-old man presents to the emergency department with acute onset of SOB of 30 minutes' duration. Initially, he felt faint but did not lose consciousness. He is complaining of left-sided chest pain that worsens on deep inspiration. He has no history of cardiopulmonary disease. A week ago, he underwent a total left hip replacement and, following discharge, was on bed rest for 3 days due to poorly controlled pain. He subsequently noticed swelling in his left calf, which is tender on examination. His current vital signs reveal a fever of 100.4°F (38.0°C), heart rate 112 bpm, BP 95/65, and an O2 saturation on room air of 91%. (Australian Institute of Health and Welfare, 2017; Semigran et al., 2015)	65-year-old male. Breathless for last 30 minutes. Chest pain on left side which is worse when breathing in. Had recent surgery, with bedrest. Has a swollen left calf which is painful. Has a fever. No history of heart disease. c/c = shortness of breath/ breathing problem	Yes	Yes
Stroke	A 70-year-old man with a history of chronic HTN and atrial fibrillation is witnessed by a family member to have nausea, vomiting, and right-sided weakness, as well as difficulty speaking and comprehending language. The symptoms started with only mild slurred speech before progressing over several minutes to severe aphasia and right arm paralysis. The patient is taking warfarin. (Cooke et al., 2013; Semigran et al., 2015)	70-year-old male. History of high blood pressure and atrial fibrillation. Feels sick. Vomiting. Weak down right side. Right arm paralysed. Has trouble speaking and is confused. Takes warfarin. c/c = weak right arm	Yes	Yes
Tetanus	A 63-year-old man sustained a cut on his hand while gardening. His immunization history is significant for not having received a complete tetanus immunization schedule. He presents with signs of generalized tetanus with trismus ("lock jaw"), which results in a grimace described as "risus sardonicus" (sardonic smile). Intermittent tonic contraction of his skeletal muscles causes intensely painful spasms, which last for minutes, during which he retains consciousness. The spasms are triggered by external (noise, light, drafts, physical contact) or internal stimuli, and as a result he is at the risk of sustaining fractures or developing rhabdomyolysis. The tetanic spasms also produce opisthotonus, board-like abdominal wall rigidity, dysphagia, and apneic periods due to contraction of the thoracic muscles and/or glottal or pharyngeal muscles. During a generalized spasm the patient arches his back, extends his legs, flexes his arms in abduction, and clenches his fists. Apnea results during some of the spasms. Autonomic overactivity initially manifests as irritability, restlessness, sweating, and tachycardia. Several days later this may present as hyperpyrexia, cardiac arrhythmias, labile hypertension, or hypotension. (Semigran et al., 2015; Victoria State Government, 2017)	65-year-old male. Cannot open mouth. Muscles tightened, causing painful spasms for minutes at a time. Sweating. Fast heart rate. Has cut his hand while gardening and did not get a tetanus injection. c/c = muscle spasms	No	Yes
Hendra Virus	(Queensland Government, 2017; Young, Selvey, & Symons, 2011)	29-year-old female. High fever, cough, sore throat, headache and tiredness. Becoming increasingly drowsy. Finding it difficult to breathe. Lives in Queensland and works in stables.	No	No

		c/c = fever		
Requires Urgent Care (n=14)				
Deep vein thrombosis	A 65-year-old woman presents with unilateral leg pain and swelling of 5 days' duration. There is a history of hypertension, mild CHF, and recent hospitalization for pneumonia. She had been recuperating at home but on beginning to mobilize and walk, the right leg became painful, tender, and swollen. On examination, the right calf is 4 cm greater in circumference than the left when measured 10 cm below the tibial tuberosity. Superficial veins in the leg are more dilated on the right foot and the right leg is slightly redder than the left. There is some tenderness on palpation in the popliteal fossa behind the knee. (Australian Institute of Health and Welfare, 2017; Kruger, Eikelboom, Douketis, & Hankey, 2019; Semigran et al., 2015)	65-year-old female. History of high blood pressure and mild congestive heart failure. 5 days of swelling and pain in 1 leg. Recent hospitalisation. Leg is painful, sore, red, swollen. c/c = Leg pain	Yes	Yes
Pneumonia (Could be called chest infection)	A 65-year-old man with hypertension and degenerative joint disease presents to the emergency department with a three-day history of a productive cough and fever. He has a temperature of 38.3°C (101°F), a blood pressure of 144/92 mm Hg, a respiratory rate of 22 breaths per minute, a heart rate of 90 beats per minute, and oxygen saturation of 92 percent while breathing room air. Physical examination reveals only crackles and egophony in the right lower lung field. The white-cell count is 14,000 per cubic millimeter, and the results of routine chemical tests are normal. A chest radiograph shows an infiltrate in the right lower lobe. (Australian Institute of Health and Welfare, 2017; Semigran et al., 2015)	65-year-old male. History of high blood pressure and degenerative joint disease. Had 3 days of coughing up phlegm with a fever. Temperature 38.3°C. c/c = cough	Yes	Yes
Acute bronchitis (Or Bronchitis, URTI, are also acceptable)	A 34-year-old woman with no known underlying lung disease 12-day history of cough. She initially had nasal congestion and a mild sore throat, but now her symptoms are all related to a productive cough without paroxysms. She denies any sick contacts. On physical examination she is not in respiratory distress and is afebrile with normal vital signs. No signs of URI are noted. Scattered wheezes are present diffusely on lung auscultation. (Cooke et al., 2013; Semigran et al., 2015)	34-year-old female. 12 days of coughing. Initially had blocked nose and sore throat. Now has cough which brings up phlegm, but no fever. c/c = cough	Yes	Yes
Acute otitis media (Or URTI, viral infection, are also acceptable)	An 18-month-old toddler presents with 1 week of rhinorrhea, cough, and congestion. Her parents report she is irritable, sleeping restlessly, and not eating well. Overnight she developed a fever. She attends day care and both parents smoke. On examination signs are found consistent with a viral respiratory infection including rhinorrhea and congestion. The toddler appears irritable and apprehensive and has a fever. Otoscopy reveals a bulging, erythematous tympanic membrane and absent landmarks. (Britt et al., 2016; Semigran et al., 2015)	18-month-old female. 1 week of runny nose, cough, feels chesty, irritable. No appetite. Fever. Goes to daycare. c/c = runny nose	Yes	Yes

Acute sinusitis (Or sinusitis, chronic sinusitis, acute on chronic sinusitis)	Mrs. S is a 35-year-old woman who presents with 15 days of nasal congestion. She has had facial pain and green nasal discharge for the last 12 days. She has had no fever. On physical examination, she has no fever and the only abnormal finding is maxillary tenderness on palpation. She is otherwise healthy, except for mild obesity. She is on no medications, except for an over-the-counter decongestant. She has no drug allergies (Cooke et al., 2013; Semigran et al., 2015)	35-year-old female. Unwell for 15 days. Has blocked nose with green mucous and facial pain for last 12 days. No fever. c/c = nasal congestion	Yes	Yes
Back strain (Or acute back sprain, acute lumbar sprain, lumbar sprain, lumbar injury, disc prolapse, sciatica)	Consider a 35-year-old man who developed low back pain after shovelling snow 3 weeks ago. He presents to the office for an evaluation. On examination there is a new left foot drop. In study 82% physicians recommend MRI (sciatica/sprain) (Cooke et al., 2013; Semigran et al., 2015)	35-year-old male. Back pain following shovelling 3 weeks ago. (Only if asked - Left foot has gone numb and is weak. This was an added symptom for Semigran et al.'s research). c/c = back pain	Yes	No
Bowel cancer (Potential diagnoses are bowel cancer, haemorrhoids, diverticular disease, other inflammatory bowel disease)	(Australian Institute of Health and Welfare, 2017; Murtagh & Rosenblatt, 2015)	56-year-old female with bleeding noticed when opening bowels. Has had some diarrhoea and constipation in the past 6 months. c/c = bleeding from the bowels	Yes	Yes
Cellulitis	A 45-year-old man presents with acute onset of pain and redness of the skin of his lower extremity. Low-grade fever is present, and the pretibial area is erythematous, edematous, and tender. (Britt et al., 2016; Semigran et al., 2015)	45-year-old male. Pain, swollen, sore and redness of skin in lower leg. Mild fever. c/c = leg pain	Yes	No
Infectious Mononucleosis (Glandular fever) (Also: Urti, tonsillitis, pharyngitis, viral infection – any of these are okay, IMN too specific)	A 16-year-old female high school student presents with complaints of fever, sore throat, and fatigue. She started feeling sick 1 week ago. Her symptoms are gradually getting worse, and she has difficulty swallowing. She has had a fever every day, and she could hardly get out of bed this morning. She does not remember being exposed to anybody with a similar illness recently. On physical examination she is febrile and looks sick. Enlarged cervical lymph nodes, exudative pharyngitis with soft palate petechiae and faint erythematous macular rash on the trunk and arms are found. (Cooke et al., 2013; Semigran et al., 2015)	16-year-old female. 1-week history of fever, sore throat, fatigue, difficulty swallowing, unable to get out of bed. c/c = fever	Yes	Yes
Migraine	(Britt et al., 2016; Kelman & Tanis, 2006)	44-year-old female. 2-day severe throbbing headache. Feels nauseated when she moves. Wants to stay in bed with curtains closed. Family has history of migraines.	Yes	Yes

		c/c = headache		
Queensland Tick Typhus	(Hore, 2001; McBride, Hanson, Miller, & Wenck, 2007; Wilson, Tierney, Lai, & Graves, 2013)	25-year-old female. Been unwell for 9 days. Fever (38.5°C), headache, dry cough, widespread rash. Muscle weakness. Painful upper left abdomen, with nausea and vomiting. Painful joints. Has been camping and gone on bush walks. c/c = fever	No	No
Ross River Virus	(Wertheim, Horby, & Woodall, 2012; Westhorpe, 2014)	19-year-old male. 2 weeks of fever with chills, muscle aches and joint pain with swelling and stiffness at joints. Rash. Fatigue. Swollen glands. Headache behind the eyes. c/c = fever	No	No
Shingles	A 77-year-old man reports a 5-day history of burning and aching pain on the right side of his chest. This is followed by the development of erythema and a maculopapular rash in this painful area, accompanied by headache and malaise. The rash progressed to develop clusters of clear vesicles for 3 to 5 days, evolving through stages of pustulation, ulceration, and crusting. (Cooke et al., 2013; Semigran et al., 2015)	77-year-old male. 5 days of burning and pain on right side of chest. Chest is red with a rash, some spots are clear raised bumps, while some are red and flat. Has a headache and feels tired and unwell. c/c = chest pain	Yes	No
Tonsillitis (Or Acute pharyngitis, throat infection, URTI)	A 7-year-old girl presents with abrupt onset of fever, nausea, vomiting, and sore throat. The child denies cough, rhinorrhea, or nasal congestion. On physical exam, oral temperature is 101°F (38.5°C) and there is an exudative pharyngitis, with enlarged cervical lymph nodes. A rapid antigen test is positive for group A Streptococcus (GAS). (Cooke et al., 2013; Semigran et al., 2015)	7-year-old female. Fever (38.5^{o}C). Feels sick. Vomiting. Sore throat, swollen neck glands. Tonsils have visible pus. No cough, no runny nose nor blocked nose. c/c = fever	Yes	Yes
Requires non-urgent care (n=11)				
Peptic Ulcer Disease (Or GORD, duodenal ulcer, gastric ulcer)	A 40-year-old man presents to his primary care physician with a 2-month history of intermittent upper abdominal pain. He describes the pain as a dull, gnawing ache. The pain sometimes wakes him at night, is relieved by food and drinking milk, and is helped partially by ranitidine. He had a similar but milder episode about 5 years ago, which was treated with omeprazole. Physical	40-year-old male. 2-month history of intermittent upper abdominal pain. (Dull and gnawing ache). Wakes at night and feels better with food, drink, milk or antacid/Gaviscon. Had a similar experience 5 years ago.	Yes	Yes

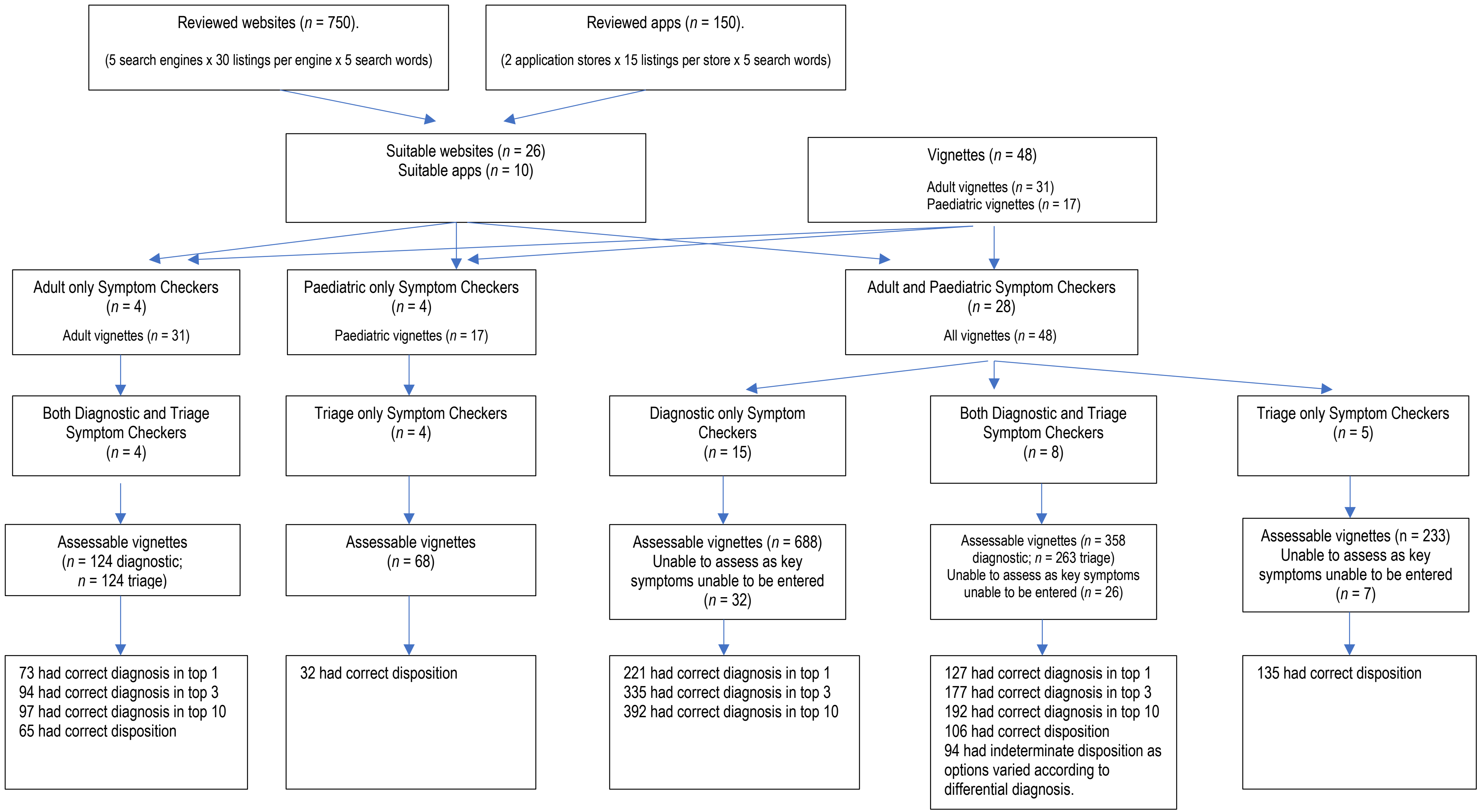
	examination reveals a fit, apparently healthy man in no distress. The only abnormal finding is mild epigastric tenderness on palpation of the abdomen. (Australian Institute of Health and Welfare, 2017; Semigran et al., 2015)	c/c = abdominal pain		
Allergic rhinitis	A 22-year-old student presents with a 5-year history of worsening nasal congestion, sneezing, and nasal itching. Symptoms are year-round but worse during the spring season. On further questioning it is revealed that he has significant eye itching, redness, and tearing as well as palate and throat itching during the spring season. He remembers that his mother told him at some point that he used to have eczema in infancy. (Britt et al., 2016; Semigran et al., 2015)	22-year-old male. 5-year history of blocked nose, sneezing, nasal itching which is worse in spring. Has itchy eyes which are red and watery. Throat and palate are also itchy. Has a history of eczema in early childhood. c/c = blocked nose	Yes	Yes
Apthous ulcers (Canker sore)	A 17-year-old male student presents with recurrent mouth ulceration since his early schooldays. He has no respiratory, anogenital, gastrointestinal, eye, or skin lesions. His mother had a similar history as a teenager. The social history includes no tobacco use and virtually no alcohol consumption. He has no history of recent drug or medication ingestion. Extraoral exam reveals no significant abnormalities and specifically no pyrexia; no cervical lymph node enlargement; nor cranial nerve, salivary, or temporomandibular joint abnormalities. Oral exam reveals a well-restored dentition and there is no clinical evidence of periodontal-attachment loss or pocketing. He has five 4 mm round ulcers with inflammatory haloes in his buccal mucosae. (Cooke et al., 2013; Semigran et al., 2015)	17-year-old male. Reoccurring mouth ulcers for a year. No respiratory, anal or genital, gastrointestinal, eye or skin lesions. Mother has similar history. No history of drugs or medications. c/c = mouth ulcers	Yes	Yes
Back pain (not a diagnosis)	A 38-year-old man with no significant history of back pain developed acute LBP when lifting boxes 2 weeks ago. The pain is aching in nature, located in the left lumbar area, and associated with spasms. He describes previous similar episodes several years ago, which resolved without seeing a doctor. He denies any leg pain or weakness. He also denies fevers, chills, weight loss, and recent infections. Over-the-counter ibuprofen has helped somewhat, but he has taken it only twice a day for the past 3 days because he does not want to become dependent on painkillers. On examination, there is decreased lumbar flexion and extension secondary to pain, but a neurologic exam is unremarkable. (Cooke et al., 2013; Semigran et al., 2015)	38-year-old male. Sudden lower back pain after lifting. No leg pain or weakness, no fevers or chills, no weight loss or recent infections. c/c = back pain	Yes	Yes
Bee sting without anaphylaxis	A 9-year-old boy is brought to the ER after being stung by a bee at a picnic. He is crying hysterically. After 15 minutes of calming him down, exam reveals a swollen tender upper lip but no tongue swelling, no drooling, no stridor, no rash, and no other complaints. (Britt et al., 2016; Semigran et al., 2015)	9-year-old male. Stung by a bee, swollen and sore upper lip. No tongue swelling, drooling, noisy breathing, rash or other complaints. c/c = bee sting	Yes	Yes

Bursitis (Trochanteric bursitis)[5]	(Murtagh & Rosenblatt, 2015)	60-year-old woman. Developed pain in L hip, hurts to sleep on that side. Otherwise well. c/c = Hip pain	Yes	Yes
Eczema	A 12-year-old female presents with dry, itchy skin that involves the flexures in front of her elbows, behind her knees, and in front of her ankles. Her cheeks also have patches of dry, scaly skin. She has symptoms of hay fever and has recently been diagnosed with egg and milk allergy. She has a brother with asthma and an uncle and several cousins who have been diagnosed with eczema. (Cooke et al., 2013; Semigran et al., 2015)	12-year-old female. Dry, itchy skin in front of elbows, in front of knees and cheeks have patches of dry, scaly skin. Symptoms of hay fever. Has egg and milk allergy. Brother has asthma and Uncle and cousins have eczema. c/c = Rash	Yes	Yes
Molluscum contagiosum	(Murtagh & Rosenblatt, 2015)	7-year-old male. Has small raised shiny pearly spots on his stomach and back with a dot in the middle. They are not painful. (If asked, no fever or other illness. Not itchy.) c/c = Rash	Yes	Yes
Solar keratosis	(Murtagh & Rosenblatt, 2015)	43-year-old female. She has fair skin. Has a small scab (3mm) on top of head which grows back if knocked off. Dry and rough to touch. c/c = skin lesion or skin mark	Yes	Yes
Sprained ankle	(Britt et al., 2016; Gregory & Mursell, 2010)	10-year-old female. Fell and twisted ankle. Slightly swollen and bruised ankle. Limping, and ankle is sore. No deformity, no sound of broken bones. c/c = Twisted ankle	Yes	Yes
Viral upper respiratory	Mr. R. is a 56-year-old man who presents to you with 6 days of non-productive cough, nasal congestion, and green nasal discharge. He has had intermittent fevers as high as 100.8. His physical examination is normal except for rhinorrhea. He is otherwise healthy, except for chronic osteoarthritis of the right knee. He has no drug allergies. (Cooke et al., 2013; Semigran et al., 2015)	56-year-old male, 6-day cough, nasal congestion, green nasal discharge. Fever (38.2°C) and runny nose. c/c = cough	Yes	Yes

Self-care appropriate (n=10)				
Blue bottle Jellyfish sting (Portuguese man of war) (*will accept jellyfish sting)	(Murtagh & Rosenblatt, 2015; St. John Ambulance WA, 2018)	16-year-old female. Swimming at beach. Stung by a jellyfish on arm. Immediate pain, slowly fading. Has some redness and swelling on arm, itchy. No trouble breathing, no abdominal pain. No nausea or vomiting. If asked- jellyfish was blue. c/c = jellyfish sting	Yes	Yes
Conjunctivitis	A 14-year-old boy with no significant past medical history presents 3 days after developing a red, irritated right eye that spread to the left eye today. He has watery discharge from both eyes, and they are stuck shut in the morning. He reports recent upper respiratory symptoms and that several children at his day camp recently had pink eye. He denies significant pain or light sensitivity and does not wear contact lenses. On examination, his pupils are equal and reactive, and he has a right-sided, tender preauricular lymph node. Penlight examination does not reveal any corneal opacity. (Semigran et al., 2015; Victoria State Government, 2016b)	14-year-old male. 3 days with red irritated eye (spread from right to left eye). Eyes have pus from ducts. Has recently had cold like symptoms. No pain or light sensitivity. c/c = red eye	Yes	Yes
Cradle cap	(Murtagh & Rosenblatt, 2015)	4-week-old male. Crusty yellowish skin appearing on head and behind ears, which looks waxy. c/c = rash	Yes	Yes
Dysmenorrhoea	(Murtagh & Rosenblatt, 2015)	15-year-old female. Monthly mid-abdominal pain appears with period. Pain radiates to back. Headache. c/c = period pain	Yes	Yes
Head lice	(Murtagh & Rosenblatt, 2015)	4-year-old female. Has very itchy head. Can see small white dots in her hair near scalp, especially behind ears. Some white dots are on hair strands and don't brush off. c/c = itchy head	Yes	Yes
Herpes simplex virus type 1 (Cold sore)	(Murtagh & Rosenblatt, 2015; Victoria State Government, 2016a)	17-year-old male. Blisters appearing on lips and just inside of mouth. Before	Yes	Yes

		blisters appeared, he felt tingling and was itchy. c/c = lip blisters		
Planter warts	(Murtagh & Rosenblatt, 2015)	20-year-old male. Hard skin-coloured lumps on the sole of the foot. Painful to walk on. Some have a black dot in the centre. c/c = skin lumps	Yes	Yes
Stinging nettle	(Ministry of Health, 2014)	25-year-old male has been bush walking. Brushed against a plant, and now has burning sensation on left arm and leg. Skin appears red, slightly swollen and itchy. (If asked, there are fine needle like hairs in the skin). c/c = burning skin	No	No
Threadworm (Strongyloides)	(Murtagh & Rosenblatt, 2015)	5-year-old male. Itchy bottom. White threads appearing in stools. Grumpy and tired. c/c = Itchy bottom.	Yes	Yes
Tinea pedis (Athlete's foot, fungal infection)	(Murtagh & Rosenblatt, 2015)	33-year-old male. Scaly skin between the toes. The skin is Itchy and has an odour. Skin goes soft and white when wet. c/c = rash	Yes	Yes

Appendix C: Flowchart for summary of diagnostic and triage advice for symptom checkers



References

AARP Health Tools. (2018). *Health Tools / Symptoms.* Retrieved February 7, 2019, from https://healthtools.aarp.org/symptomsearch#

APWU Health Plan. (2018). *Helpful tool: Symptom Checker.* Retrieved November 7, 2018, from https://www.apwuhp.com/members/for-all-members/symptom-checker/

Australian Institute of Health and Welfare. (2017). *Emergency department care 2016-17: Australian hospital statistics.* Retrieved from https://www.aihw.gov.au/reports/hospitals/ahs-2016-17-emergency-department-care/contents/table-of-contents

Britt, H., Miller, G. C., Bayram, C., Henderson, J., Valenti, L., Harrison, C., … Wong, C. (2016). *A Decade of Australian General Practice Activity 2006-07 to 2015-16 (No.41).* Sydney University Press. Retrieved from https://ses.library.usyd.edu.au/bitstream/handle/2123/15482/9781743325162_ONLINE.pdf;jsessionid=6DA6EB5836FBFB5D9DA29ACA4325A4D8?sequence=5

Buoy Health. (2018). *Buoy.* Retrieved December 15, 2018, from https://www.buoyhealth.com/

Children's Wisconsin. (2019). *Symptom Checker.* Retrieved January 7, 2019, from https://www.chw.org/health-information/symptom-checker

Cooke, G., Valenti, L., Glasziou, P., & Britt, H. (2013). Common general practice presentations and publication frequency. *Australian Family Physician, 42*(1/2), 65–68. Retrieved from http://www.racgp.org.au/afp/2013/januaryfebruary/common-general-practice-presentations/

Doctor Diagnose Symptom. (2018). *Doctor Diagnose Symptoms Check.* (Version 1.0.5) [Mobile application software]. Retrieved January 17, 2019, from https://play.google.com/store/apps/details?id=com.appcolliders.doctordiagnose

Drugs.com. (2019a). *Symptom Checker.* Retrieved January 5, 2019, from https://www.drugs.com/symptom-checker/

Drugs.com. (2019b). *Symptom Checker.* (Version 2.9.3) [Mobile application software]. Retrieved January 5, 2019, from https://play.google.com/store/apps/details?id=com.drugscom.app

ePain Assist. (2018a). *Symptom Checker.* (Version 1.5.5) [Mobile application software]. Retrieved December 7, 2018, from https://apple.com/au/ios/app-store

ePain Assist. (2018b). *Symptom Checker.* (Version 1.5.5) [Mobile application software]. Retrieved December 7, 2018, from https://play.google.com/store/apps/details?id=com.epainassist.symptomchecker

Everyday Health. (2019). *Symptom Checker.* Retrieved January 8, 2019, from https://www.everydayhealth.com/symptom-checker/

Family Doctor. (2018). *Symptom Checker.* Retrieved November 24, 2019, from https://familydoctor.org/your-health-resources/health-tools/symptom-checker

Gregory, P. Bs., & Mursell. (2010). *Manual of clinical paramedic procedures.* Chichester, U.K. ; Wiley-Blackwell.

HealhLinkBC. (2018). *Check Your Symptoms.* Retrieved November 11, 2018, from www.healthlinkbc.ca/health-topics/hwsxchk

Health Status. (n.d.). *Symptom Checker.* Retrieved November 9, 2018, from https://www.healthstatus.com/symptom-checker/

Healthdirect. (2018a). *Check Your Symptoms.* (Version 2.2.1) [Mobile application software]. Retrieved December 18, 2018, from https://apps.apple.com/au/app/healthdirect/id1021494621

Healthdirect. (2018b). *Symptom Checker*. Retrieved December 18, 2018, from https://www.healthdirect.gov.au/symptom-checker

Healthline. (2018). *Symptom Checker.* Retrieved November 22, 2018, from https://www.healthline.com/symptom-checker

Healthy Children. (2018). *KidsDoc Symptom Checker.* Retrieved November 25, 2018, from https://healthychildren.org/english/tips-tools/symptom-checker/Pages/default.aspx

Hore, C. (2001). Important unusual infections in Australia: a critical care perspective. *Critical Care and Resuscitation, 3*(4), 262–272.

Isabel Healthcare. (2018). *Isabel.* Retrieved November 1, 2018, from https://symptomchecker.isabelhealthcare.com/suggest_diagnoses_advanced/landing_page

Johns Hopkins All Chilren's Hospital. (2019). *Symptom Checker.* Retrieved January 6, 2019, from https://www.hopkinsallchildrens.org/Patients-Families/Patient-Family-Resources/Symptom-Checker#!/index/child/body

Kelman, L., & Tanis, D. (2006). The relationship between migraine pain and other associated symptoms. *Cephalalgia, 26*(5), 548–553. doi: 10.1111/j.1468-2982.2006.01075.x

Kruger, P. C., Eikelboom, J. W., Douketis, J. D., & Hankey, G. J. (2019). Deep vein thrombosis: update on diagnosis and management. *Medical Journal of Australia, 210*(11), 516–524. https://doi.org/10.5694/mja2.50201 https://ecu.on.worldcat.org/oclc/8145015391

Mayo Clinic. (2018). *Symptom Checker.* Retrieved November 15, 2018, from https://www.mayoclinic.org/symptom-checker/select-symptom/itt-20009075

McBride, W. J. H., Hanson, J. P., Miller, R., & Wenck, D. (2007). Severe Spotted Fever Group Rickettsiosis, Australia. *Emerging Infectious Diseases, 13*(11), 1742–1744. https://doi.org/10.3201/eid1311.070099

MedicineNet. (2018). *Symptom Checker: Symptoms & Signs A-Z.* Retrieved November 28, 2018, from https://www.medicinenet.com/symptoms_and_signs/symptomchecker.htm#introView

Menigicoccal Australia. (n.d.). *The Facts.* Albury: New South Wales, Australia. Retrieved July 1, 2018, from http://www.meningococcal.org.au/new-page-1/

Ministry of Health–Manatū Hauora. (2014). *Stinging nettles.* Retrieved July 1, 2018, from https://www.health.govt.nz/your-health/conditions-and-treatments/accidents-and-injuries/bites-and-stings/stinging-nettles

Murtagh, J., & Rosenblatt, J. (2015). *John Murtagh's general practice* (Sixth ed.). North Ryde, NSW : McGrw-Hill Education.

My Health Alberta. (2018). *Check Your Symptoms.* Retrieved December 12, 2018, from https://myhealth.alberta.ca/Health/pages/symptom-checker.aspx

Patient.info. (n.d.). *Symptom Checker*. Retrieved November 23, 2018, from https://patient.info/symptom-checker

Queensland Government. (2017). *Hendra virus infection.* Retrieved April 14, 2018, from http://conditions.health.qld.gov.au/HealthCondition/condition/14/217/363/Hendra-Virus-Infection

Right Diagnosis. (2011). *Symptom Checker.* Retrieved December 20, 2018, from https://symptoms.rightdiagnosis.com/

RxList. (2018). *Symptom checker*. Retrieved December 15, 2018, from https://rxlist.com/symptoms-checker/symptom.htm#introView

Semigran, H. L., Linder, J. A., Gidengil, C., & Mehrotra, A. (2015). Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ (clinical research ed.), 351*, 3480. doi: 10.1136/bmj.h3480

St. John Ambulance WA. (2018). *Non-tropical jellyfish stings.* Retrieved July 1, 2018, from https://www.stjohnambulance.com.au/docs/default-source/first-aid-compliance/non-tropicval-jellyfish-stings.pdf?sfvrsn=2

St Luke's Online. (n.d.). *Symptom Checker.* Retrieved January 12, 2019, from www.stlukesonline.org/symptom-checker

Symcat. (2019). *What is bothering you today?* Retrieved January 14, 2019, from www.symcat.com

Symptify. (2018). *How are you feeling today?* (Version 2) [Mobile application software]. Retrieved January 3, 2019, from https://play.google.com/store

Symptomate. (2019a). *Symptomate.* Retrieved January 8, 2019, from https://www.symptomate.com/diagnosis/

Symptomate. (2019b). *Symptomate.* (Version 2.1) [Mobile application software]. Retrieved January 8, 2019, from https://apps.apple.com/au/app/symptomate/id837725433

Symptomate. (2019c). *Symptomate.* (Version 2.0) [Mobile application software]. Retrieved January 8, 2019, from https://play.google.com/store/apps/details?id=com.symptomate.mobile

University of Michigan. (2019). *Symptom Checker*. Retrieved January 4, 2019, from http://www.uofmhealth.org/health-library/sx

Victoria State Government. (2014). *Better Health Channel - Haemolytic uremic syndrome.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/haemolytic-uraemic-syndrome

Victoria State Government. (2015). *Better Health Channel - Malaria.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/malaria

Victoria State Government. (2016a). *Better Health Channel - Cold Sores.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/cold-sores

Victoria State Government. (2016b). *Better Health Channel - Conjunctivitis.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/Conjunctivitis

Victoria State Government. (2017). *Better Health Channel - Tetanus.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/healthyliving/tetanus

Victoria State Government. (2018). *Better Health Channel - Kidney Stones.* Retrieved May 17, 2018, from https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/kidney-stones

WebMD. (2017). *WebMD Symptom Checker*. Retrieved November 14, 2018, from https://symptoms.webmd.com/default.htm

WebMD. (2018a). *WebMD Symptom Checker.* (Version 7.4) [Mobile application software]. Retrieved November 14, 2018, from https://apps.apple.com/au/app/webmd-symptoms-doctors-rx/id295076329

WebMD. (2018b). *WebMD Symptom Checker.* (Version 7.0) [Mobile application software]. Retrieved November 14, 2018, from https://play.google.com/store/apps/details?id=com.webmd.android

Wertheim, H. F. L., Horby, P., & Woodall, J. P. (2012). *Atlas of human infectious diseases.* John Wiley & Sons.

Westhorpe, C. (2014). Ross River virus. *Geodate, 27*(3), 4.

What's My Diagnosis. (n.d.). *What's My Diagnosis.* Retrieved January 9, 2019, from www.whatsmydiagnosis.com

Wilson, A., Tierney, L., Lai, K., & Graves, S. (2013). Queensland tick typhus: three cases with unusual clinical features. *Internal Medicine Journal, 43*(7), 823–825. https://doi.org/10.1111/imj.12184

Young, J., Selvey, C., & Symons, R. (2011). Hendra virus. *The Medical Journal of Australia, 195*(5), 250–251. https://doi.org/10.5694/mja11.10967