Can generic health website quality assessment tools identify the 'best' websites for a specific health topic?: A study of orthognathic treatment websites

'Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Dental Science (Orthodontics)'

Ayeh Mahdmina

September 2015

Table of Contents

List of Tables

List of Figures

i. Acknowledgement

I would like to express my deep gratitude to my supervisors Professor Rebecca Harris and Dr. Norah Flannigan for their help and guidance throughout this entire research project. They supported me by sharing their knowledge and expertise at every step of this study while still allowing me to express my ideas. Without their encouragement and enlightenment, I would not have been able to complete this project.

I would also like to thank Dr. Girvan Burnside for his help and support with the statistical analysis of the study results.

Finally, I would like to thank my family for their support and encouragement in my education especially since moving to the UK. Your endless input of love, confidence and encouragement and enduring great hardships have helped me achieve my dreams and complete such a research project. Mum and Dad thank you for all you have done and I am forever indebted to you.

ii. Structured abstract

Introduction: Health website quality assessment tools are designed to guide patients to high quality websites and away from those of poor quality. However, the ability of such tools to do this successfully is unclear. **Aim:** To assess the applicability of three recognised generic health website quality assessment tools (LIDA, DISCERN and HWAT 3.0) to a specific health topic by ascertaining whether they can identify the highest quality orthognathic websites. **Method:** A cross-sectional study was carried out to assess the correlation between the quality scorings of 100 orthognathic websites produced from a Google search, when the websites were examined using these three tools. The rankings of the 100 websites produced with each tool were also compared with quality scores from a newly developed (AM) checklist which had been specifically developed for the assessment of orthognathic context. The AM checklist was used as a standard to assess if orthognathic websites contained all of the information perceived by clinicians as necessary for their patients. **Results:** DISCERN was the only tool to identify correctly the highest quality orthognathic websites. A strong correlation was found between DISCERN and the AM checklist ($r= 0.816$ $p<0.01$) but no other correlations were found. The AM checklist and LIDA both had good inter- and intra-examiner reliability. However, the reliability of the HWAT 3.0 and DISCERN tools was moderate. The AM checklist and DISCERN indicated that UK websites were of significantly better quality than US websites. No significant differences were found between the quality of websites according to the Google ranking and suffix. Blogs had the highest overall scores and were comparable to websites of hospitals and professional organisations. **Conclusions:** Generic health website quality assessment tools require further development to increase their reliability and ability to identify high quality websites. UK orthognathic websites can be used as a marker of high quality compared with US websites but the Google ranking and suffix are poor quality markers. Blogs should be investigated further as a possible useful resource of information for orthognathic patients.

1. Introduction

The Internet is a relatively new information medium and is unique in that it is directly accessible to a large proportion of the population. It is estimated that 2500 million individuals have access to the Internet worldwide (Fox and Duggan, 2013) so with such ready access to this information, it is not surprising that there has been a significant increase in the use of the Internet for health related purposes (Powell and Clark, 2002). A large number of health websites exist on the Internet and there is increasing use of these websites by patients. However, the quality of information can vary widely between websites. This is of concern particularly in a health context because poor quality information could be misleading and potentially harmful to patients. In response to this concern, different methods have been developed to assess the quality of health websites.

There are numerous studies in the literature which have applied 'Health website quality assessment tools' (HWQATs) to medical websites to ascertain the quality of websites for particular topics such as cosmetic surgery and orthodontic pain (Parikh et al., 2006, Livas et al., 2013). HWQATs are designed to identify the 'best' website from a list of websites for any health topic. Their purpose is to guide patients towards high quality websites and away from those of poor quality by giving each website a numerical score.

Most studies merely use HWQATs to merely produce a list of websites with different qualities at a single time point. Such studies can quickly become outdated. There are fewer studies which have investigated these tools against a standard to see if the tools actually serve their purpose by identifying the 'best' websites i.e. the websites which contained the highest quality information for patients (Hsu and Bath, 2008, Khazaal et al., 2012). This is important because if HQWATs do not identify the best website for a topic correctly, this could lead to incorrect or misleading information being accessed by patients.

Every effort should be made to give patients high quality information so that they are able to make well informed treatment decisions. This is particularly important in elective interventions such as orthognathic treatment where there are numerous risks, including some which are permanent, and where the treatment is long and complex.

This study investigated the ability of HWQATs to serve their purpose in identifying the highest quality of orthognathic websites. The outcome of this investigation was an objective assessment of three chosen HWQATs and to identify needed improvements.

2. Literature review

2.1 The use of the Internet is increasing

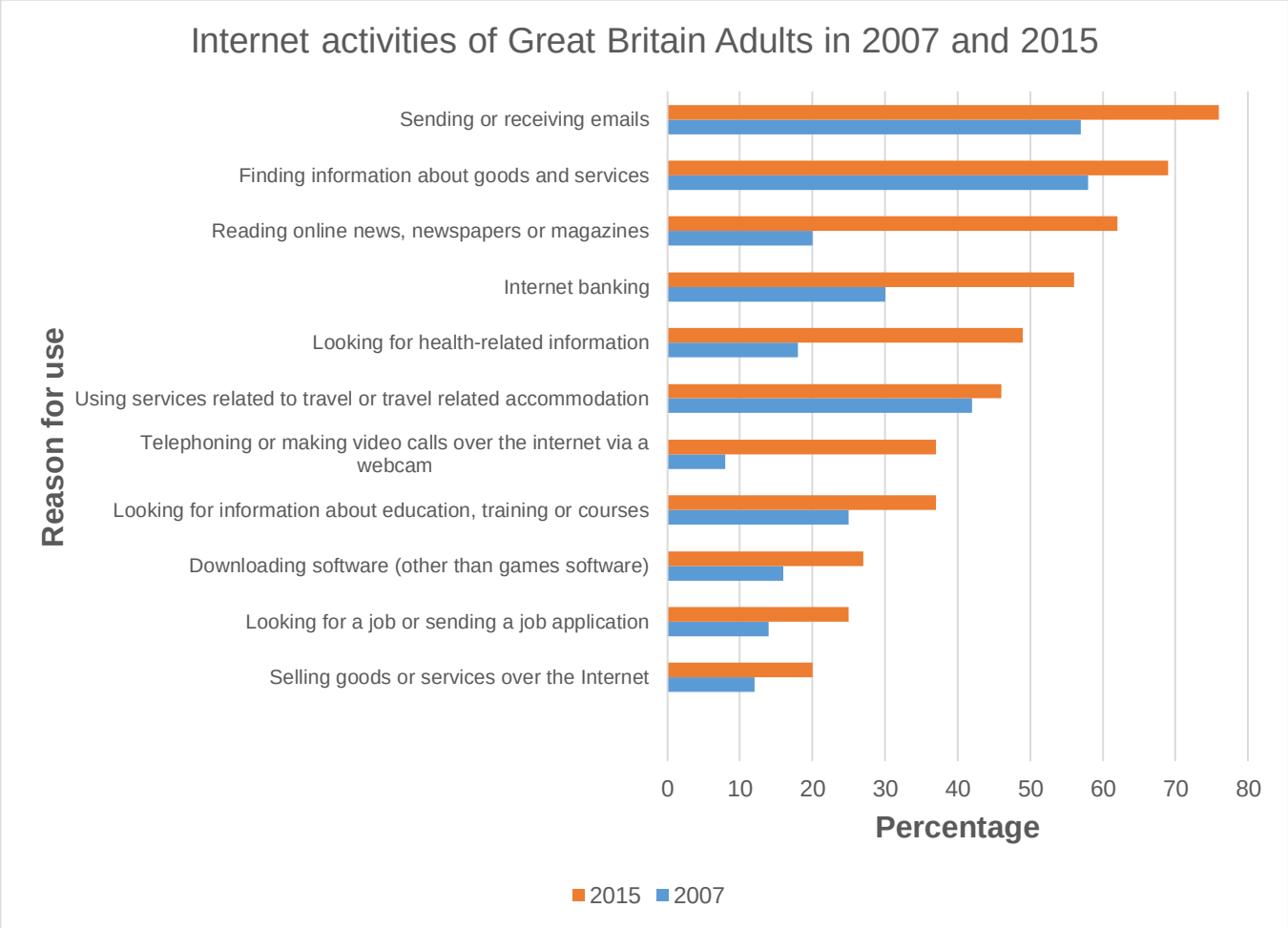
Since its advent, the Internet has become the single largest source of information available to the general population (Fahy et al., 2014). The Internet is an appealing way to find information because of its speed, anonymous nature, attractive visual presentation (Bohacek et al., 2003), and ease of information availability compared to other information sources. The Internet is also appealing because the ease of access to this information is constantly improving with the use of mobile Internet devices such as smartphones and portable computers. A recent report by the Office for National Statistics (Office for National statistics [ONS] 'Internet Access' report 2015) highlighted that 96% of 16-24 year old adults in Great Britain had used the Internet 'on the go'. This easy access may explain why there has been such a huge increase in Internet use in the last few years. The same report showed a huge increase in Internet use over the last decade, with 78% of adults in Great Britain accessing the Internet every day, compared to 35% in 2006.

2.2 Use of the Internet for medical and dental information

Previous research has shown that a significant proportion of Internet searches are carried out on health-related topics (McMullan, 2006). In Great Britain, health information was the fifth most common reason for Internet use in 2015 (Figure 1). By comparison, in 2007 looking for health information was the seventh most common reason for Internet use demonstrating that more individuals are currently using the Internet to find health related information.

The use of the Internet for seeking health information is also common in other parts of the world. For example Andreassen et al. (2007) reported in a study of European citizens that 71% had used the Internet for health purposes. Furthermore, a USA based survey showed that 54% of American adults had sought health information online (Fox and Duggan, 2013).

Figure 1: The percentage of 16-24 year old Great Britain adults' reasons for use of the Internet in 2007 and 2015



Source: Office for National Statistics

The growing significance of the Internet in health communication is as relevant to dentistry as it is to other healthcare areas. However, there are very few recent studies in the literature which investigate how much dental patients use the Internet for online dental information. A study by Riodan and McCreary in 2009 reported that 17% of UK dental patients used the Internet to look for treatment options, and a further 15% had reported its use for 'acceptance of their condition'. Although 17% may appear to be a small proportion compared to some of the other studies, this study was carried out 6 years ago and this percentage is therefore now likely to have changed. Also, in 2009 only 70% of households had Internet access (ONS

'Internet Access' report 2009) compared to 86% in 2015 (ONS 'Internet Access' report 2015) so there may now be more dental patients accessing online information.

2.3 Trust in online health related information

Members of the public are not only increasingly using the Internet to access health information but they also place a great deal of trust in the information they read online. A study of US patients (Diaz et al., 2002) reported that 60% of patients rated the information gained from the Internet to be the "same as" or "better than" the information provided by their doctors. Furthermore, 59% of patients who found information online did not feel the need to discuss the subject further with their doctors. Similar attitudes were found in an Israeli study (Russ et al., 2011). In this study, 81% of patients stated that they did not share the information they found online with their doctors. Furthermore, in another UK-based study investigating reasons for Internet use, 32% of patients were found to use online information instead of seeking professional advice (Charnock and Shepperd, 2004).

2.4 Impact of the Internet on patients' relationships with clinicians in medicine and dentistry

The increase in the use of the Internet and the level of trust placed in online health information by patients can impact on the clinician patient relationship. This relationship may be affected in a positive way as shown by a study by Russ et al. (2011). This study showed that of the patients who did discuss their findings with their doctor, 87% reported that this action had a positive impact on the relationship with their doctor. However, some clinicians do not appear to share this viewpoint. For example a survey of UK Obstetricians and Gynaecologists (Nwosu and Cox, 2000) demonstrated changes in perception with an increase in Internet use from a clinician's viewpoint. This study indicated that clinicians accepted that their patients might be better informed than the clinicians themselves because of their use of the Internet. This study also reported that 40% of the clinicians surveyed felt that the Internet might damage the doctor-patient relationship (Nwosu and Cox, 2000).

Again, there are few studies exploring the impact of the Internet on the dentist patient relationship. One such study was carried out by Chestnutt and Reynolds (2006) who conducted a questionnaire based study of 457 UK general dental practitioners (GDPs) to ascertain how their 'dental treatment delivery' had changed since the advent of the Internet. The findings of this study showed that a third of patients asked for more complex procedures from their GDPs after reading online information. Ten per cent of the GDPs reported a perception that the Internet was a "threat" to their relationship with patients.

Thus patients appear to view information acquired online as comparable to, if not superior to the information they receive from their clinicians. This highlights the need for clinicians to be aware of the information that their patients are coming across online.

2.5 Benefits of online health information for patients

The Internet can potentially be a very useful source of health information both from a patient and clinician's perspective. Patients have reported many advantages of Internet usage, for example, using the Internet to prepare for a visit to their physician, sharing information and avoiding stigma by seeking information anonymously for certain types of drug information (Pohjanoksa-Mantyla et al., 2009).

Reading high quality information on the Internet can potentially help patients in understanding their treatment options as well as the implications of treatment in order to make well informed treatment decisions. Previous studies have demonstrated the importance of patients being well informed before the start of treatment. For example, a systematic review of terminal cancer patients showed that almost all patients preferred to know as much information as possible about their disease and treatment before making a decision (Gaston and Mitchell, 2005). Other studies have also shown that well informed patients were less anxious (Fallowfield et al., 1990, Wiles et al., 1998) and better prepared for health care consultations (Meredith et al., 1995). Patients' participation in healthcare decision- making has also been related to treatment outcome in that patients who had

greater participation in their own care had better clinical outcomes (Greenfield et al., 1985; Kaplan, 1989). Even those who did not take an active role in their healthcare were shown to benefit from further information (Coulter, 1997). The Internet allows patients to become further involved in their care, which in turn helps with their treatment decision making (Coulter, 1998; Goldsmith, 2000). The Internet can also potentially overcome communication problems between patients and professionals due to time constraints, the use of technical terms, and a high volume of information provided during a relatively short consultation (Farrell et al., 2006; Thompson and Graydon, 2009), which potentially improves the patient-professional relationship (Murray et al., 2003; Diaz et al., 2005). It may also help increase compliance with treatment (Thompson and Graydon, 2009).

From a clinician's point of view, the information available on the Internet can potentially increase the public's access to health related information (Cline and Hayes, 2001; Eysenbach and Kohler, 2002) and remove barriers to information access (McLellan, 1998; McLeod, 1998; Edejer, 2000).

However, these benefits can only be true if the information provided on health websites is of high quality and the information factually correct. This is important because as discussed in section 2.3, patients place a great deal of trust in the information they read online and therefore, if the information they read is of poor quality or factually incorrect, this may pose a risk to them.

2.6 Risks of online health information

Due to its unregulated and open nature, a major disadvantage of the Internet is that the quality of available health websites cannot easily be controlled, allowing patients to have access to health websites of diverse quality. For a clinician who has had training in obtaining high quality information and evidence, peer-reviewed written articles are available. These have usually gone through a rigorous process of quality assessment, ensuring that information is supported by scientific evidence and is as free from bias as possible.

However, patients usually have little medical knowledge or the ability to assess the quality of the material they are reading. Hence, the Internet has the potential to be misleading and harmful (Hainer et al., 2000; Diaz et al., 2002; Eysenbach et al., 2002; Schmidt and Ernst, 2004).

Websites that appear educational may in fact be promotional (Huang et al., 2005; Jain and Barbeiri, 2005) and the information provided can be incomplete, conflicting and out of date (Lewiecki et al., 2006). A study reported that the quality of health website was below optimal levels in 70% of websites (Eysenbach et al., 2002). This highlights the importance of a means to assess the quality of medical information online.

Another disadvantage is that static information such as a list of websites can quickly become outdated as certain websites become unavailable (Bohacek et al., 2003). Therefore, it is important to establish other means of health website quality assessment which can be used at any time.

2.7 Methods of quality assessment of health websites

2.7.1 How patients assess the quality of online health information

Previous research has shown that patients tend to rely on pre-selected websites from an authoritative source such as a hospital or professional organisation to assess the quality of online information (Marshall and Williams, 2006), or select a website on the basis of its ranking in the search engine used (Eysenbach and Kohler, 2002). Several search engines exist, with Google the most popular of these, according to Internet statistics websites (http://www.alexa.com/siteinfo/https://www.google.com), although such statistical information is not frequently reported in scientific literature.

The Google ranking, however, has been previously shown to be a poor marker of quality (Bohacek et al,. 2003; Perez-Lopez, 2004). This might be due to the method by which Google ranks websites – a sophisticated technique to search for relevant terms in all parts of a website (including other pages linked from a website) that also uses a complex algorithm.

The algorithm calculates a *PageRank* score and the pages with the highest *PageRank* score are ranked towards the top of the search results (Brin and Page, 2012). The algorithm that calculates the *PageRank* function is based on diverse factors but places great emphasis on the number of times a particular website is cited, because the developers believe that the number of citations and links from other websites is an approximation of website quality. Although this method may produce relevant websites, they might not necessarily have the highest quality. Therefore, it can be appreciated that assessing the quality of online health information can be complex.

2.7.2 Methods of assessing the quality of health information described in the literature

The lack of regulation of online information and clinicians' awareness of websites with poor quality information has led to the development of various methods for assessing the 'quality' of health websites (Silberg et al., 1997; Doupi and Van Der Lei, 1999; Gagliardi and Jadad, 2002; Wilson, 2002; Charnock and Shepperd, 2004) to help distinguish websites with useful and a high standard of information. These can be divided into six areas:

1. Comparison with an existing gold standard. The gold standard would be that all required information reaches the patient (Satterlund et al., 2003). Unfortunately no gold standard methods exist for dental websites.
2. 'Codes of conduct' are a set of quality criteria recommended for websites. They set out ideas of good practice from the perspective of consumers and professionals to enable website developers to self-assess their website and consequently improve standards. If such a code is provided by a larger organisation such as a national or specialty medical association, the association may take on a further enforcement role to ensure that members comply with the codes and discipline those who do not. An example was the Health Summit working group (this no longer exists).
3. Self-applied quality labels. These logos or symbols are displayed by a website certifying their commitment to adhere to a code of conduct after following a process application and approval. The certifying organisation may continue to check the

information provided by the websites and users may have the opportunity to report misuse e.g. Health on the Net (www.healthonnet.org).

4. HWQATs: This is the most common method reported in the literature. Initially this method was developed for patients' use, for it has been shown that HWQATs' use increases awareness of signs of poorer quality websites among patients (Charnock and Shepperd, 2004). However, the usefulness of these systems depends on the user's awareness of the HWQATs' existence and whether the user is familiar with them (Berland et al., 2001). HWQATs comprise several questions with possible answers that have a numerical value, usually 0, 1 or 2. The answers are tallied and an overall score is given to the website. This is a particularly useful method because it provides a means to rank websites in order of quality once they have been assessed. Examples of HWQATs include DISCERN, LIDA and QUICK tools, which will be discussed further in section 2.8.4.
5. Filtering. This is a concept of a establishing a 'gateway' to health websites. A set of approved websites is stored within a database and the user can search for a particular topic. The websites are selected for their precision and relevance after application of a manual or automatic filter. Unfortunately, there are currently only filtering systems for the use by professionals and not by patients.
6. Third party accreditation logos. This is the most advanced form of accreditation which can be awarded. It is attained after a rigorous assessment of quality by a third party, usually for a fee. A logo is then displayed to confirm that not only has the website been investigated, but it also continues to adhere to the stated quality criteria. However, it must be noted that it is possible that fraudulent logos can also be displayed and that the health website developers may not be aware of the proper procedures for accreditation (Delamothe, 2000). Also, some health website developers may not be able to afford the fee to apply for accreditation. An example of an accreditation organisation includes URAC (Utilization Review Accreditation

Commission) which has its own set of criteria that websites must follow to be accredited (http://www.urac.org/prog_accred_HWS_po.asp?).

2.7.3 Other quality indicators for medical websites

Previous studies have tried to identify simple ways to identify health websites of higher quality. An established finding from the medical literature is that websites with a '.gov' suffix rank the highest for quality, as assessed using HWQATs (Bohacek et al., 2003; Ansani et al., 2005; Dash et al., 2012) and '.com' websites usually rank lowest (Bohacek et al., 2003) in comparison with other suffixes.

Although there has also been some research into assessing the quality of websites according to country of origin and types of websites, (Perez-Lopez., 2004; Ogunwale et al., 2009; Thompson and Graydon, 2009; Best et al., 2014) findings are contradictory and inconclusive. Research in this area however, is in its infancy and so there is no work as yet which investigates website quality in the field of orthodontic care.

2.8 Health websites quality assessment tools

2.8.1 Aim of the tools

As discussed in section 2.7.2 HWQATs are the most common method used to assess the quality of health websites. Earlier tools were developed to be used by patients in order to help them find high quality websites. However, patients have little awareness of their existence (Berland, 2001) and these tools are now frequently used by professional (e.g. clinicians) and non professional (e.g. informatics specialist) individuals instead to assess the quality of websites. These high quality websites can then be recommended to patients.

2.8.2 Method of use of the tools

These tools aim to identify high quality medical websites by giving a numerical score according to various parameters that are then used as quality markers. The use of such quality indicators allows the tools to be applicable to a wide variety of health websites (Breckons et al., 2008). These parameters include the title, the update date and the

presence of a reference list. For example, if the topic of the website is clearly displayed in it, then the title parameter is scored 1 instead of 0. If there is no reference list then a score of 0 is given. These numerical scores are then summed to give an overall score that enables users to rank the website, with the highest ranking websites identified as being the 'best' website i.e. having the highest quality.

2.8.3 Critique of health websites quality assessment tools

Although many tools have been developed and applied to various health topics to ascertain quality within that field, relatively little literature examines and compares the performance of the various website evaluation tools themselves. Many aspects of a tool must be tested to determine whether the tool is fit for its purpose. These aspects are discussed in sections 2.8.3.1 to 2.8.3.4.

2.8.3.1 Tool reliability

Reliability can be divided into two types: inter- and intra-examiner reliability. Inter-examiner reliability is the agreement between results produced by two or more investigators and intra-examiner reliability is the agreement between the results produced from the same investigator on two or more occasions. Ideally, a tool or test should produce the same result every time it is used regardless of the investigator or the time when it is used.

The inter- and intra-examiner reliability of HWQATs has been assessed in previous research (Ademiluyi et al., 2003; Lewiecki et al., 2006; Downing et al., 2011). However, results are contradictory and therefore the reliability of the tools must be investigated further.

2.8.3.2 Tool validity

Validity is defined as whether a test measures what it is supposed to measure. A HWQAT is considered valid if it was shown that patients felt they had obtained the information they needed from a website which also had a high score on a HWQAT- in other words where patients agreed that the highest ranking website according to the HWQAT gave them the best information.

However, to date, none of the range of available quality ranking tools has been assessed for validity in this regard (Jadad and Gagliardi, 1998; Berland et al., 2001; Gagliardi and Jadad, 2002; Lewiecki et al., 2006). Ideally, any HWQAT should be validated before use.

2.8.3.3 Agreement between the health website quality assessment tools

The agreement between the HWQATs is contradictory in the literature, with some studies showing a good level of agreement between the tools (Hsu and Bath, 2008) and some reporting poorer agreement (Breckons et al., 2008). It is unclear whether different HWQATs can identify the same website as having the best quality and this must be investigated further.

2.8.3.4 Assessment of website content

Most HWQATs focus mainly on assessing general quality indicators such as the presence of the website author's name, a clear indication of the date that the website was last updated, and a source to contact for further information. These indicators have been identified on the assumption that adherence to them is likely to ensure that the website content will be of reasonable quality (Walji et al., 2004; Breckons et al., 2008). The content (i.e. the factual information) of the website is however, assessed to a lesser degree in most tools (Table 1). There are only a few studies in the literature (Hsu and Bath, 2008, Prusti et al., 2012) which have assessed whether tools that rate websites as having the best quality actually identified those with the best content; i.e. did the general quality indicators correctly identify websites with comprehensive and accurate content. These studies have contradicting results, however, and therefore further investigation in this important area is needed.

Multiple authors have demonstrated that well written information with good readability does not necessarily mean that the content is of good quality (Charnock et al., 1999, Lewiecki et al., 2006). This is important as it determines whether these tools are useful for the assessment of website content quality in the first place. Previous research has shown that

content assessment is required to assess the overall quality of health literacy in print (Tones, 2002) therefore the contents of health websites also needs to be assessed.

If HWQATs only utilise general quality indicators rather than thoroughly checking the website content, then it is questionable whether a website can be deemed to be of high quality when the veracity of the information it contains has not been evaluated. Since generic HWQATs are not condition- or treatment-specific, they may be an inadequate means of identifying information quality when considered from the perspective of a specific health topic such as orthognathic treatment.

Very few studies in medicine have assessed the content of the websites. This is also the case in dentistry, with only one published study by Livas et al. (2013) where the quality of the websites content is investigated.

2.8.4 Examples of tools in the literature

There are a large number of tools available in the literature. The main HWQATs available are summarised in Table 1. This table lists who developed the tool (usually clinicians but sometimes informatics specialists); the type of questions included in the tool (general questions or content- specific questions); whether the tool was developed to assess a content- specific or general health topic and studies which have applied these tools to assess the quality of a chosen health topic. Examples of such studies are described in section 2.8.5.

Table 1: Examples of health websites quality assessment tools and their features: number of questions, details of developer and health topic developed to assess

Name of Tool	Developer	Developer background	Health topic intended for	Number of questions	Percentage of general parameter questions (%)	Percentage of content-specific questions (%)	Patients involved in development?	Studies tool used by
LIDA	Tomlin and Badenoch (www.minervation.com/lida-tool)	Information Science	Any	41	98	2	No	1) Patel and Cobourne, 2011 2) Best et al.,, 2014 3) Livas et al., 2013 4) Prusti, et al., 2012 5) Leira- Feijoo., 2014 6) Neumark, et al., 2012 7) Park et al., 2012
DISCERN	Charnock et al., 1999	Health Science	Any	16	70	30	Yes	1) Aldairy et al., 2012 2) Khazaal et al., 2012 3) Best, 2014 4) Downing et al., 2011 5) Parikh et al, 2006 6)Cerminara et al, 2014, 7)Kumar et al., 2014 8)Zahedi et al., 2013
HWAT 3.0	Lewiecki, 2006	Clinician (Physician)	Breast cancer	13	100	0	Yes	None
DARTS	Narhi et al., 2008	Finland Pharmaceutical forum	Depressions	5	100	0	No	Prusti et al,. 2012
OncoRx-IQ	Yap et al., 2010	Clinician (Pharmacist)	Anticancer drug interactions	25	80	20	No	None
American Medical Association criteria	Ambre et al., 1997	Clinician (Physician)	Any	23	100	0	No	1)Khazaal et al., 2012 2) Park et al., 2012 3) Lopez- Jornet et al., 2010 4) Chang et al., 2010

Table 1 represents just a small number of available tools which have been described in the literature. It shows that the majority of the main HWQATs available have been developed by clinicians and have not involved patients during their development process. Therefore, they cannot be considered to be validated (see section 2.8.3.2). Most of these HWQATs also contain general questions rather than content- specific questions. These issues highlight further the shortfalls of these tools described in section 2.8.3.

The tools which were chosen for this study were the LIDA, DISCERN and HWAT 3.0 tool for reasons discussed further in section 3.4.5.1. These HWQATs are discussed further in the following sections.

2.8.4.1 Minervation LIDA tool (www.minervation.com/lida-tool)

This tool was created by two information scientists working in evidence-based health care. Three key qualities of health care websites; accessibility, usability and reliability, are the basis for this 41 item tool. The accessibility of the website assesses whether the website functions (loads) and is user- friendly for disabled persons. The accessibility is calculated by placing the URL of the website being tested into a section on the LIDA website which generates a score. The usability aspect aims to assess whether users can gain the information they need from the website and is assessed by answering questions such as 'can the user make an effective judgement of whether the site applies to them?'. The reliability of the website is assessed to ascertain whether the author of the website is clear about how they have produced their content, and is also assessed by answering questions such as 'is site content updated at an appropriate interval'.

Information on the use of the LIDA tool is available on the LIDA website (www.minervation.com/lida-tool). The LIDA website has a shortened version of the full tool with a drop down menu for the answering a question. The full version is provided in a handbook which is complemented with additional guidance below each question. The short version of the questions is outlined in Appendix 1 and the full 41 items can be found in

Appendix 2. Each question in both the short and long form are scored 0= Never, 1= Sometimes, 2= Mostly or 3= Always. The full version was used in this study.

The authors claim that the LIDA tool has been validated with surveys and evaluations by independent organisations, although details of the evaluation methods employed were not described fully (www.minervation.com/lida-tool). The reliability of the LIDA tool was tested by two assessors independently, rating 40 prostate cancer websites. Correlation rankings for each examiner were compared and a Spearman's rank correlation coefficient of 0.611 was found. This measure however, reflects the degree of inter- examiner reliability arising from the use of the tool rather than measuring validity in its true sense even though the authors purport their investigation was one of validity testing.

2.8.4.2 DISCERN

The DISCERN tool was developed for written information on treatment choices (Charnock et al.,1999). The tool was developed in recognition that few methods existed to assess the quality of written information which would be useful for a range of health care professionals and patients. The DISCERN tool may not appear to be an obvious HWQAT for health related websites because it was designed for use on written information such as patient information leaflets. However, its authors later evaluated the performance of their tool for use with online information with good results (Charnock and Shepperd, 2004).

The DISCERN tool was developed by a panel of experts which consisted of medical specialists as well as self-help group members and a lay medical publisher. Twenty- seven themes related to treatment options of three medical conditions (myocardial infarction, endometriosis and chronic fatigue syndrome) were identified and these were eventually developed into 16 questions. A Likert scale of 1–5 was used for responses and each question was supplemented with a guide to the user in choosing their score:

1- Are the aims clear?
2- Does it achieve its aims?
3- Is it relevant?
4- Is it clear what sources of information were used to compile the publication (other than the author or producer)?
5- Is it clear when the information used or reported in the publication was produced?
6- Is it balanced and unbiased?
7- Does it provide details of addition source of information?
8- Does it refer to areas of uncertainty?
9- Does it describe how each treatment works?
10- Does it describe the benefits of each treatment?
11- Does it describe the risks of each treatment?
12- Does it describe what would happen if the treatment is used?
13- Does it describe how the treatment choices affect the overall quality of life?
14- Is it clear that there may be more than one possible treatment choice?
15- Does it provide support for shared decision making?
16- Based on the answers to all the above questions, rate the overall quality of the publication as a source of information about treatment choices.

The functionality of the DISCERN tool was assessed by its authors by carrying out a national pilot of the tool which recruited health information providers (described as those giving health advice to patients or dealing with health information) and self-help group members. This pilot assessed inter- examiner reliability and also asked participants for feedback of the tool itself. The results of this pilot showed inter- examiner reliability to be higher for health information providers than self-help group members. From these findings the authors concluded DISCERN to be a valid and reliable tool for judging the quality of written consumer information (Charnock et al., 1999).

In 2004, the authors carried out another study (Charnock and Shepperd, 2004) in which professionals and consumers were recruited and asked to assess breast cancer websites using the DISCERN tool. Feedback from users was positive and so the authors concluded that the DISCERN tool could also be successfully applied to health information website too (Charnock and Shepperd, 2004).

2.8.4.3 Health Website Assessment Tool 3.0 (HWAT 3.0)

HWAT 3.0 was developed from quality assessments of osteoporosis information on the Internet (Lewiecki et al., 2006). Initially, scoring elements of website quality from previous studies were identified and weighted according to the developer's perception of their importance. The first version of the tool (HWAT 1.0) was tested by osteoporosis nurses and physician experts to assess the quality of a range of poor to good quality websites. Findings were used to modify the tool (HWAT 2.0) and the process was repeated to develop the final version (HWAT 3.0). The categories and quality indicators and their respective weightings in the tool can be found in Appendix 3. The tool scores web information in five domains; content, credibility, navigability, currency and readability. The content is assessed by answering questions such as 'whether the originating person/ organisation is identified?' The credibility is also assessed by answering questions such as 'is the source of information credible?'. The navigability assessment includes questions concerned mainly with the functionality of the website such as 'do the interwebsite links function?'. The currency is assessed using a single question which asks if there is a date of when the website was last updated. The readability domain is assessed by using the Flesch- Kincaid grade.

Each question is given a score of "0 or 1" and that score is multiplied by "4" if the quality indicator is helpful but not necessary, "6" if the quality indicator is desirable and "9" if the indicator is essential. The maximum possible score is 100.

Unlike most tools developed by clinicians and information specialists with little or no patient input in the development stage, HWAT 3.0 tool had patient involvement in its development.

Although patients were not involved in the initial tool development, developers used a group of patients to refine the tool (Lewiecki et al., 2006). This involved a simple Patient Evaluation Tool (PET) to evaluate 10 osteoporosis websites. The PET included five questions on each criteria to be answered as, “agree” for a score of 20, “not sure” for a score of 10 and “disagree” for a score of 0. The questions are listed below:

1) This website offers information about osteoporosis that is helpful to me
2) I believe the information provided on this website is scientifically correct
3) It is easy to find information I need on this website
4) The information on this website is up to date
5) The information on this website is easy to understand.

Scores from the PET were tallied to give an overall score and the 10 websites were ranked in order of quality according to the PET scores. Scores from the HWAT 3.0 tool as assessed by nurses were then also ranked and the two sets of rankings (from PET and HWAT 3.0) were compared. A 60% agreement between the two was found. Although this was a low score not ideal for a definitive tool, this is the closest attempt at validation so far.

Intra- examiner reliability was tested with nurses with 79% agreement and also with experts with 88% agreement but not between nurses and physician experts nor between patients (for inter- examiner reliability).

2.8.5 Use of HWQATs in the literature

There are a large number of studies which have assessed the quality of various health topics in medicine and orthodontics and examples of such studies are discussed in sections 2.8.5.1 and 2.8.5.2.

2.8.5.1 Application of tools to medical websites

There are numerous studies in the literature which have applied existing tools to evaluate the quality of medical websites on a specific health topic. For example, in a study by

Downing et al. (2011), the authors used the American Medical Association guidelines, HoN (Health on the Net quality criteria), DISCERN and LIDA to assess the quality of 89 websites giving information for asplenic patients. The tools indicated that these websites had diverse qualities. A particular strength of this study was that unlike many other studies, both professional and patient websites were assessed. This is important because patients have access to both. However, like many studies in the literature, this study simply identified the best asplenic information websites at the time of study based on the results of the tools used.

Another example of a study which simply produced a gateway to high quality websites was the study by Parikh et al., 2006. This study used DISCERN, HoN, and QUICK (Quality information checklist) tools to identify the best cosmetic surgery websites for the gateway. However, the authors of these studies did not assess objectively any of the tools before the start of the study in order to ascertain whether these tools would be able to identify correctly the best websites for their chosen topics of asplenic information and cosmetic surgery.

Other studies have developed a new tool more specific to the health topic they were interested in investigating. An example of this is the MWAT (medication website assessment tool) tool which was developed for the study of Methotrexate information websites. Using MWAT, these websites were shown to be missing some crucial information (Thompson and Graydon, 2009). The constant development of these new tools has little value because these tools cannot be applied to other health topics. Instead, it would have been more useful if the authors had compared the results of their MWAT tool to the results produced from existing tools such as LIDA and DISCERN which are not topic- specific tools. This would have shown whether these existing tools were able to fulfil their purpose and identify correctly the best Methotrexate information websites and therefore be deemed suitable for use on any given health topic.

Like most other studies, a study of antidepressant drug information websites also showed websites to have a wide range of qualities (Prusti et al., 2012). This study, however, also correlated the tools used (DISCERN and DARTS [Date, Author, Reference, Type, Sponsor]) with their own content guidelines. Although good correlation was reported, the small sample size of 22 websites meant that the findings of this study must be interpreted with caution. Therefore, this study did not clarify whether these tools could identify the website with the best contents as assessed by the investigators content guidelines.

2.8.5.2 Application of tools to orthodontic websites

Like most studies in the medical literature, there have been a lack of studies which assess the tools themselves in their ability to identify the best health website for a particular health topic. Most studies in orthodontics also simply apply the existing tools to assess the available websites, for a chosen orthodontic topic, and for quality. An example of this is a study of orthodontic extractions websites which used the LIDA tool and found the quality of these websites to be variable (Patel and Cobourne, 2011). Similarly the study by Aldairy et al. (2012) found there to be a wide range in the quality of orthognathic treatment websites, although two websites were identified as having the highest quality and therefore recommended for patient use. However, this study was only based on a small sample of 25 websites. None of these studies had investigated the tools they were using themselves and in fact claimed that the tools which they used were validated, but, as mentioned in section 2.8.3.2, no tools have yet been validated.

Like the study by Prusti et al. (2012) mentioned in section 2.8.5,1, a study by Livas et al. (2013) assessed the quality of the content of online information about pain during orthodontic treatment using their own developed method. To do this, the authors simply scored each orthodontic pain related website 1 to 5 for its content and also applied the LIDA tool. Findings showed that the content assessment and LIDA assessment were not correlated. This would have been useful for assessing whether LIDA had correctly identified the website with the best contents (as assessed with the authors content score).

As highlighted in sections 2.3 to 2.6 the information that patients come across can have a significant impact on them in many ways. Therefore, HWQATs should be able to identify correctly high and low quality websites for patients to be guided toward the best sources of information and away from the poor quality ones. It is unclear whether the currently available tools are able to do this, and this requires further investigation.

2.9 Using orthognathic websites to objectively assess health website quality assessment tools

As outlined in section 2.8.3, the literature shows that HWQATs require further investigation to ascertain whether these tools are able to fulfill their purpose of identifying correctly the best websites for any given health topic. Orthognathic treatment is a useful context in which to examine whether these tools can discriminate between websites containing high and low quality information for a number of reasons.

Firstly, orthognathic treatment is elective and requires the patient to understand the significant risks and implications on their lives so that they can give informed consent to participate. Orthodontics addresses problems with how teeth bite together or the malalignment of teeth, as well as improvement of facial appearance using fixed or removable appliances. This treatment takes around 18 to 24 months to complete (Proffit and Miguel, 1995). Orthognathic treatment, however, is a much more complicated treatment option that involves fixed appliances but also a major surgical procedure. This surgery is performed under general anaesthesia by a maxillofacial surgeon who corrects skeletal discrepancies. This is carried out by performing controlled breaks of the upper, lower or both jaws (for example if a patient's lower jaw is so prominent that fixed appliances alone could not improve their appearance or achieve a normal occlusal relationship). Orthognathic treatment takes a much longer period to complete (2.5 to 3 years) when compared to routine orthodontic treatment and as such requires a fully committed patient in order for treatment to be successful.

Treatment has a large impact on the patient's life because there is both a short- term physical impact on the patient's daily life as well as potentially long- term psychological effects (Hunt et al., 2001). Treatment also has the ability to dramatically change a patient's physical appearance. As well as the potential psychosocial impacts on the patient there are also serious risks associated with treatment such as pain, bruising, swelling, numbness of the lips, (which can be permanent), blindness and even mortality. It is therefore extremely important for patients to be as informed as possible about all aspects of treatment in order to decide whether they wish to accept the risks involved in order to gain the potential benefits. Information from the Internet can be used to help patients understand the risks and implications of treatment and help them in their treatment decision making process. However, this information needs to be of high quality and therefore having a HWQAT to assess the information correctly is important. Furthermore, accurate information has been shown to increase patient satisfaction with the overall outcome of orthognathic treatment (Cunningham et al, 1996).

Secondly, orthognathic patients are likely candidates to use the Internet during their treatment to seek further information. This is because orthognathic patients are usually young adults and likely to use the Internet at various treatment stages to acquire further knowledge on the treatment such as risks, benefits and the recovery process (Kruse et al., 2012). Also, orthognathic treatment is a common type of treatment in orthodontics making up 7% of all UK orthodontic treatment undertaken by Consultant Orthodontists (Luther et al., 2003) and more patients are undergoing this treatment. Not only have the outcomes improved due to advances in surgical techniques, but the procedure has also become more comfortable for the patient (Sarver and Johnston, 1993) making this a widely accepted and well-established procedure (Hunt et al., 2001). Therefore, it is likely that more patients will undergo this procedure in the future and with the rise in Internet use in the UK outlined in section 2.1, this group of patients might use the Internet more and more to seek information

about their treatment. Again, it is important for a high quality HWQAT to highlight high quality information for patients.

Finally, to be able to assess whether the HWQAT tool can separate successfully high and low quality websites for a given topic, the selection of websites needs to have a range of qualities (from high quality to low quality websites). This was shown to be the case in the study by Aldairy et al. (2012) in the case of orthognathic treatment websites. This study had applied the DISCERN tool to a small number of websites but found orthognathic treatment websites to have a wide range of qualities.

2.10 Rationale for study

Although many HWQATs have been developed and applied to various health topics to ascertain quality within a specific field, less attention has been paid to the assessment of the tools themselves in order to ascertain whether these existing HWQATs can correctly identify high quality websites. Moreover, very few studies have verified whether websites that have been ranked highly by an HWQAT identifying the best information for patients truly were the best website with the highest quality information. In order to accomplish this, an objective method to measure quality is necessary.

3. Purpose and design of study

3.1 Aim

The aim of this study was to assess the extent to which health website quality assessment tools reflect the quality of orthognathic websites. It investigated whether all assessment tools reported the same website to be the best and whether all websites ranked the same when different health website quality assessment tools were used.

3.2 Objectives

Primary objectives

1. To compare three chosen generic health website quality assessment tools (LIDA, DISCERN and HWAT 3.0) with a checklist specifically generated to assess orthognathic treatment websites (AM checklist).
2. To compare the ranking of orthognathic websites for quality using the three health website quality assessment tools LIDA, DISCERN and HWAT 3.0.

Secondary objectives

1. To assess whether the ranking in Google correlates with the overall quality rank.
2. To assess the inter- and intra-examiner reliability of the three health website quality assessment tools and the AM checklist.
3. To assess the difference in quality between sites according to suffix (e.g. '.gov,' and '.com')
4. To assess the difference in quality between sites according to country of origin.
5. To assess the difference in quality between different types of websites (e.g. Blogs and Hospital websites).
6. To assess whether sites with accreditation from the British Orthodontic Society ranked higher for overall quality.
7. To assess whether sponsored sites ranked higher for overall quality.

3.3 Null Hypotheses

1. There is no correlation between the orthognathic website quality rankings produced by LIDA, DISCERN and HWAT 3.0 and the AM checklist.
2. There is no difference in the orthognathic website quality ranking produced from LIDA, DISCERN and HWAT 3.0.
3. There is no correlation between the quality of orthognathic websites and Google ranking.
4. There is no difference in the inter- and intra-examiner reliability of the three health website quality assessment tools.
5. There is no correlation between the quality of orthognathic websites and suffix (e.g. '.gov,' and '.com')
6. There is no correlation between the quality of orthognathic websites and country of origin.
7. There is no correlation between the quality of orthognathic websites and website types (e.g. Blogs and Hospital websites).
8. There is no correlation between the quality of orthognathic websites and accreditation from the British Orthodontic Society.
9. There is no correlation between the quality of orthognathic websites and sponsorship.

3.4 Design

This was a cross-sectional study that analysed 100 websites at a single time point from a single search result using three HWQATs each against the developed standard, the AM checklist. This design was used because information on the Internet constantly changes and if the patient carries out a search, the patient is likely to consider the results of a single search.

3.5 Method

3.5.1 Development of the orthognathic specific checklist as a standard

3.4.1.1 Aim of the AM checklist

The AM checklist was developed to assess objectively whether health website quality assessment tools were able to identify 'correctly' the highest quality orthognathic websites. This checklist was composed of the information that clinicians involved in providing orthognathic treatment would like their patients to learn from an orthognathic treatment website and which would allow a focused examination of the content of each website. The AM checklist was developed in order to compare the results produced from the generic tools against this 'standard' which contained all the desired information from the clinician's point of view.

3.4.1.2 Method of development

In order to develop the checklist, the primary researcher attended and observed two multidisciplinary orthognathic treatment new patient consultation clinics at Liverpool University Dental Hospital. This was in order to help identify the information which is routinely discussed with patients before they start orthognathic treatment. This information was then used as a guide to what information should be present on an orthognathic treatment website. In total, two orthodontic and two oral and maxillofacial surgeon consultants were observed with nine patients attending the clinics. A list of the topics discussed with each of the patients who attended this clinic was compiled. This information was then categorised into checklist questions. These questions were then emailed to three different oral and maxillofacial surgeon consultants and three orthodontic consultants in the Mersey region for further feedback. Consultants involved in this second round of development were asked whether all the information they would want their patients to know rom an orthognathic website was covered. Feedback and comments from these consultants

was used to refine further the AM checklist. The checklist, at this stage of development, had seven topic areas covered within the tool:

1. Epidemiological information and frequency of the performance of procedure.

2. A description of procedure including an explanation of the surgery and the fixation of bones using plates and screws.

3. Multidisciplinary treatment involving both braces and surgery.

4. Indications/ contraindications.

   Indications include an extremely protrusive lower jaw (severe class III), extremely retrusive lower jaw (severe class II), an anterior open bite in which a large gap between the top and bottom teeth is evident, facial asymmetry, abnormality of the chin and craniofacial anomalies such as a cleft lip and palate.

   Contraindications may be due to medical or psychological reasons.

   Both indications and contraindications should be mentioned because this would inform the patient whether they are suitable for treatment.

5. Use of general anaesthesia.

   Orthognathic surgery is performed under a general anaesthesia and this is an important aspect of treatment which must be mentioned not only due to the potential risks (discussed under question 6) but also to highlight the commitment needed for an extensive treatment from the patient.

6. Risks, benefits and alternative treatment.

   Common risks include:

   a) Altered sensation of the lips with 12.8% experiencing this long- term (Colella et al., 2007)

b) Infection, which occurs in approximately 3.4 % of patients (Sousa and Turini, 2012)
c) Bleeding, which occurs in 1.4% of patients (Sousa and Turini, 2012)
d) Pain, bruising and swelling.

Rarer risks include:

a) Risk of mortality from the general anaesthetic thought to occur in 1 in 100,000 general anaesthetic procedures (Jenkins and Barker, 2003)
b) Blindness (V. Cruz and Santos, 2006).

Ideally all risks should be mentioned on an orthognathic treatment website.

7. Length of the overall treatment and recovery period.

A score of 1 was given for each question if the information was present or 0 if the information was not present (outlined in Figure 2). This very simple scoring system was used because the purpose of this checklist was merely to assess the presence of information rather than to create a new HWQAT, as such creation was not the aim of this study.

Figure 2: The AM checklist (Version 1)

Type of information	Score
1. Epidemiological information and frequency of performance of procedure	Present=1 Absent=0
2. Description of procedure including explanation of the surgery, fixation of bones using plates and screws	Present=1 Absent= 0
3. Multidisciplinary treatment involving both braces and surgery	Present=1 Absent=0
4. Indications/ contraindications	Present=1 Absent=0
5. Use of a general anaesthetic	Present=1 Absent=0
6. Risks, benefits and alternative treatment	Present=1 Absent=0
7. Length of overall treatment and recovery period	Present=1 Absent=0

3.4.1.3 Pilot test of the AM checklist

Before the start of the study a pilot test of the AM checklist was carried out in order to ensure the checklist had acceptable inter- and intra- examiner reliability and also to identify and address any potential issues before the start of the study. This pilot test was carried out by the application of the AM checklist to the top 10 orthognathic websites (Appendix 4) which had been saved offline from a Google search on one occasion. This was carried out by the primary researcher and a second examiner (S.H.) on two occasions- 6 weeks apart.

The results of the first reliability test (described in section 5.4) showed surprisingly lower reliability than expected. The inter-examiner reliability in particular was very poor. It was viewed as important that the reliability of the AM checklist was improved before the study continued, in order to compare the three existing tools to a ‘standard’, i.e. the AM checklist needed to be reliable. To identify the cause of the poor reliability, the questions which

produced the most disagreement in the reliability test were identified and discussed between the two raters. Several issues were identified and an attempt to improve the reliability was made through the following changes.

First, two questions which required multiple pieces of information were split into individual questions. A question about whether indications/ contraindications are present or absent was split into two separate questions. Another question about whether ‘risks, benefits and alternative treatment‘ were present on the website was also split into three questions asking separately whether each piece of information was present. Splitting the questions reflected the information on the website more accurately and eliminated confusion caused by an answer to one question of ‘no’ and to another of ‘yes’. For example, if risks were present but benefits were not, neither a score of 1 nor 0 could be assigned, but splitting the questions would facilitate scoring and the overall score for a website would reflect the information present.

Second, guidance notes were added for each question to make the questions more objective and less subjective. A discussion between the primary researcher and S.H. as well as the results of the inter-examiner reliability from both examiners showed that questions about topics such as ‘risks and use of a general anaesthetic’ might be answered differently. One examiner might assign a score of 1 (present) if a single risk was mentioned. However, another examiner might consider several risks of orthognathic treatment extremely important, so if only one risk was mentioned, that examiner might assign a score of 0 (absent). This is not necessarily accurate and reflective of the website content. To avoid personal opinion bias of the examiner influencing the score, more specific guidance notes were added for each question. For example for the question about risks, the guidance note indicated that ‘Numbness needs to be mentioned as well as swelling, bruising etc. ’ In order to emphasise the importance of numbness and that a mention of swelling and bruising alone is not sufficient for the assignment of a score of 1. The updated version of the AM checklist with the guidance notes added is outlined in Figure 3.

Figure 3: AM checklist (Version 2)

Checklist question	Score
1. Epidemiological information and frequency of performance of procedure Guide note: % occurrence of Cl II, Cl III, AOB, etc. OR % how often orthognathic surgery is carried out.	Present=1 Absent=0
2. Description of procedure including an explanation of the surgery and the fixation of bones using plates and screws Guide note: needs to mention cut/break/repositioning of bone AND screws/plates for fixation.	Present=1 Absent=0
3. Multidisciplinary treatment involving both braces and surgery Guide note: Needs to mention both orthodontics and surgery involved.	Present=1 Absent=0
4. Indications Guide note: If several indications are listed then assign a score of 1.	Present=1 Absent=0
5. Contraindications Guide note: Minimum ages before orthognathic Rx can be considered, medical conditions.	Present=1 Absent=0
6. Use of a general anaesthetic Guide note: Needs to specifically mention anaesthesia/ general anaesthetic. "Sleep" doesn't count.	Present=1 Absent=0
7. Risks Guide note: Numbness needs to be mentioned as well as swelling, bruising, etc.	Present=1 Absent=0
8. Benefits Guide note: Improved bite/function AND aesthetics both need to be mentioned.	Present=1 Absent=0
9. Alternative treatment Guide note: Needs to mention this is elective treatment.	Present=1 Absent=0
10. Length of overall treatment Guide note: Needs to mention that treatment duration is approximately 2–3 years.	Present=1 Absent=0
11. Recovery period Guide note: Needs to mention that the recovery period is approximately two weeks.	Present=1 Absent=0

Finally, it was noticed that the top 10 Google search results contained a video diary website which was updated very frequently. Unfortunately, because the primary researcher was unaware of this, the exact names and links to the videos were not initially saved when the website was saved offline. Therefore, it was possible that different videos were viewed during the intra- examiner and inter- examiner reliability tests.

Once these problems were identified and appropriate changes were made to the checklist, another reliability test was carried out using 10 different websites (Appendix 4) on two different occasions, 6 weeks apart by the primary researcher and S.H. Video diary websites were excluded for the reliability test to avoid encountering the problems found with the first reliability test. Video diary websites were only used for the main study.

After these changes were made, both the inter- and intra-examiner reliability tests were repeated (as outlined in section 5.4). Although the confidence intervals were wide which indicated lower precision of estimates, these results showed a significantly higher inter- and intra- examiner reliability than the initial reliability scores and the AM checklist was deemed acceptable to be used for the study.

3.5.2 Sampling

Google was used to identify the websites to be used in this study. As this was a cross sectional study and because websites are potentially updated constantly, the results from a Google search of the word ‘orthognathic’ were saved into an Excel document using the Google Chrome add on toolbar SEO QUAKE. The search was carried out on 28 November 2013, and the list of websites from the search was saved offline using the ‘Save offline’ function in Windows explorer for analysis later. The offline version of the website was used for analysis in all cases although there were 3 exceptions where the live version of the website had to be used. The live versions of the website had to be used for assessing video diary websites and also if the saved offline version failed to load. Furthermore the following

questions in the LIDA tool required the use of the live version of the website to be able to answer questions.

1.5 Browser Test

2.2.2 Do navigational links have a consistent function?

2.3.4 Does the site support the normal browser navigational tools?

After duplicates were automatically removed by Google 339 websites remained. The top 100 relevant websites were analysed from these according to the inclusion and exclusion criteria (section 3.4.3 and 3.4.4).

The 100 websites were saved offline so that the results from the search were from one occasion. Because patients are unlikely to do multiple searches on different occasions to find better websites (Eysenbach and Kohler, 2002), this strategy more closely reflected patients' experience of using the Internet for information.

The sample size of 100 websites was not based on a sample size calculation due to the lack of previous similar studies. The sample size was chosen based on a pilot study of orthognathic websites where the search term of 'orthognathic' identified 342 websites. Since one of the study objectives was to look at Google ranking as a measure of quality, a reasonable number of websites needed to be included within the study so that websites at the top and also nearer to the bottom of the Google ranking were analysed. One hundred websites out of the possible 339 was deemed to be an acceptable representation of website quality in order to address the study aims.

3.5.3 Inclusion criteria

- Websites disclosing information about orthognathic treatment were included whether they were aimed at patients or professionals. This inclusion criteria was chosen because patients have access to both patient and professional websites and could use either to gather the information they require.

3.5.4 Exclusion criteria

- Duplicate websites.
- Websites requiring subscriptions to view.
- Non-functional websites: Two further attempts were made to access non-functional websites at 2 week intervals after the initial attempt. If the website was still non-functional after the two attempts the next website from the saved rank of websites was used.
- Non-English language websites.
- Medical product supply websites

3.5.5 Health website quality assessment tools

3.4.5.1 Method of selection of tools

Three HWQATs, DISCERN, LIDA and HWAT 3.0, were chosen from the many available HWQATs in the literature using an electronic search in SCOPUS with the search term in Appendix 5. This search produced 2170 relevant articles. LIDA was the most commonly used tool in the dental literature with 28 out of 34 dental articles using LIDA. DISCERN was the most frequently used tool in the medical literature with 252 out of 2136 medical articles using DISCERN. The rest of the 1884 medical articles used different tools to assess quality. It was likely that future studies would also use LIDA and DISCERN therefore an objective assessment of these tools was important.

As discussed in section 2.8.3 most tools have shortfalls. One such shortfall is that most tools have not involved patients in their development and therefore cannot be deemed valid (section 2.8.3.2). The HWAT 3.0 tool was chosen because, as discussed in section 2.8.4.3, this tool had the most patient involvement during its development compared to the other tools screened for use in this study. Therefore, this tool was chosen for objective assessment in this study so that if good results were found, this tool could be recommended for use in future studies evaluating the quality of health websites.

3.4.5.2 Pilot test of the tools and amendments

Prior to the start of the study, two raters (the primary researcher and S.H.) applied the three tools (LIDA, DISCERN and HWAT 3.0) to the top 10 orthognathic websites from a Google search using the term 'orthognathic' in order to test the tools for ease of use and to identify any potential problems. The results of this pilot test were used to make several changes to facilitate the use of the tools before the start of the study.

Guidance notes for LIDA

Both examiners found some questions in the tools vague and subjective, making these questions likely to be prone to examiner bias. None of the existing tools were fully validated and in order to reduce the likelihood of examiner bias, the primary researcher developed some additional brief guidance notes to help assess and score the questions to increase reliability and facilitate scoring. These guidance notes were also used for the reliability tests. An example of a guidance note for LIDA question 2.1.6 'Is the colour scheme appropriate and engaging?' The guidance notes from the LIDA developers were as follows:

'Is it appropriate for the target audience?

'Is it tasteful?'

'Is it readable?'

'Print out a black and white page to see if there's enough contrast for color blind people.'

'Remember to check the colours of mouse-overs and previously-clicked links etc.'

Despite such guidance, scoring between 0 and 3 was difficult, so these additional guidance notes were added:

'If one color and pictures used score 2, if more than one color & pictures used score 3, if one or two color but no pictures used score 1.'

LIDA supplemental questions

Despite the large number of questions in LIDA, several other supplemental questions were available but not used in this study, because the LIDA authors had advised that these questions be used only if they could all be answered. Most of these supplemental questions required the website author to be contacted, for example, 'Is content reviewed by an independent expert or panel?' which would have been extremely time- consuming and difficult due to the large sample size of the study.

HWAT 3.0 readability assessment

As part of the health website assessment process, the authors of the HWAT 3.0 tool included one question that assessed website readability with the Flesch-Kincaid reading test. However, readability was not assessed separately or as part of the HWAT 3.0 assessment in this study as this was viewed as too time- consuming and out of the scope of the study resources.

Categorisation of websites by tools

The developers of each of the existing tools used in this study designated categories for the numerical scores produced from each tool, for example, 'high', 'moderate' and 'low' quality , to make quality judgement easier for the user. However, these parameters were different for each tool so it would have been challenging to directly compare the categories from each tool. Therefore only the numerical scores produced from each tool were used for analysis.

3.5.6 Recording of data and ranking

For each website the numerical score produced from the application of the three tools and the AM checklist were captured electronically using Google docs (an online survey facility) so that data was automatically uploaded to a spreadsheet to speed up the analytical process. The use of drop down menus minimised errors and increased the rate of the data-capturing process. To reduce the likelihood of fields being left blank, before moving on, the

number selection was verified after each question was answered, to ensure that the correct number was selected and the field was not left blank. Also a maximum of 10 websites were analysed at a time to reduce examiner fatigue.

Online access to Google docs facilitated data capturing and analysis and minimised errors because the data was stored in one place. Data was backed up weekly.

The numerical score from each of the three tools and the AM checklist for every website was then ranked in order to identify the highest quality websites at the top of the rank and lower quality websites towards the bottom of the rank. In total four ranking lists were therefore produced, 3 for the tools and one for the AM checklist.

3.5.7 Reliability

Intra-examiner reliability was tested by the primary researcher by repeating the scoring, using the three tools and AM checklist (discussed in section 3.4.1.3), of the top 10 websites from the Google search after 6 weeks.

Inter-examiner reliability was assessed by comparing the scores of the same 10 websites produced from the primary researcher with those produced when the second examiner (S.H.) also carried out the assessment. Only 10 websites were used in the reliability tests because of the length of time taken (around 30 minutes) to complete the evaluation of one website using the three tools and the AM checklist. The primary researcher did not give any verbal guidance on the use of the tools but provided S.H. with the written guidance from the developers of each of the tools (if this was provided) and the supplemental guidance notes described in sections 3.4.1.3 for the AM checklist and 3.4.5.2 for LIDA.

The best way to assess inter- examiner reliability would have involved using several other assessors to score four or five websites using each tool and the AM checklist. However, due to the time taken to assess each website it was decided that it was more practical for one other assessor to score the same 10 websites as the primary researcher.

3.5.8 Statistical analysis

3.4.8.1 Distribution of the data

To assess the distribution of the data, histograms (Figures 4- 7) of the frequency of scores from each tool and the AM checklist were compiled. These showed that the data was reasonably symmetrical and the level of skewness was not sufficient to invalidate the use of parametric tests. Therefore parametric tests were applied to assess the data.

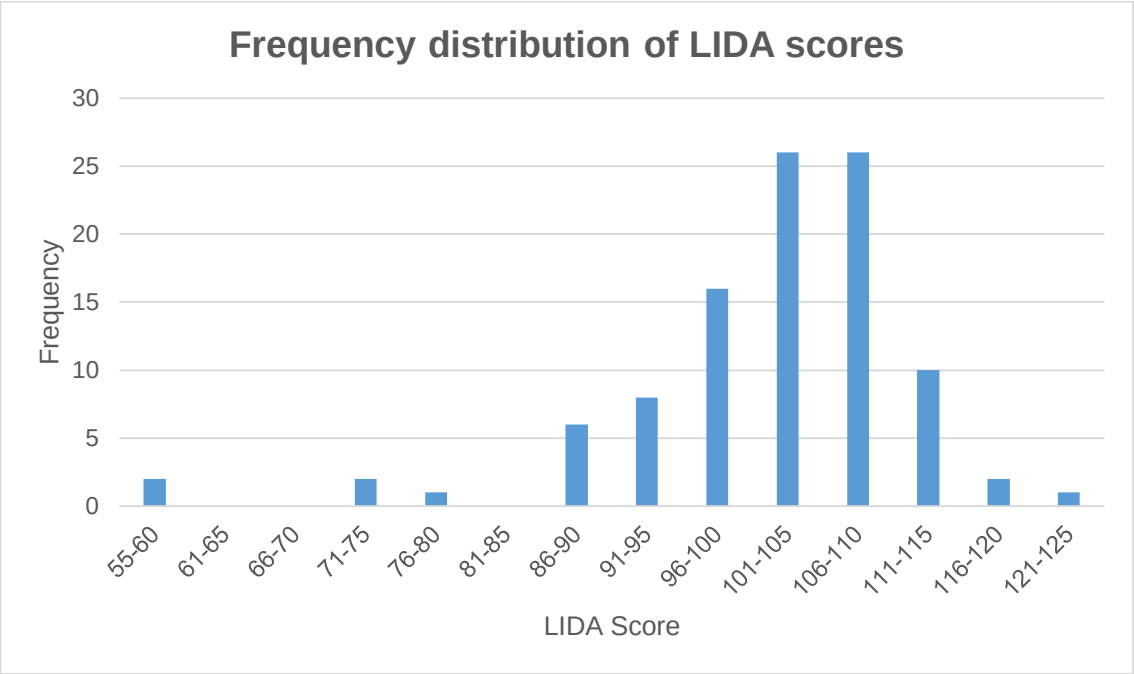


Figure 4: The frequency distribution of LIDA scores

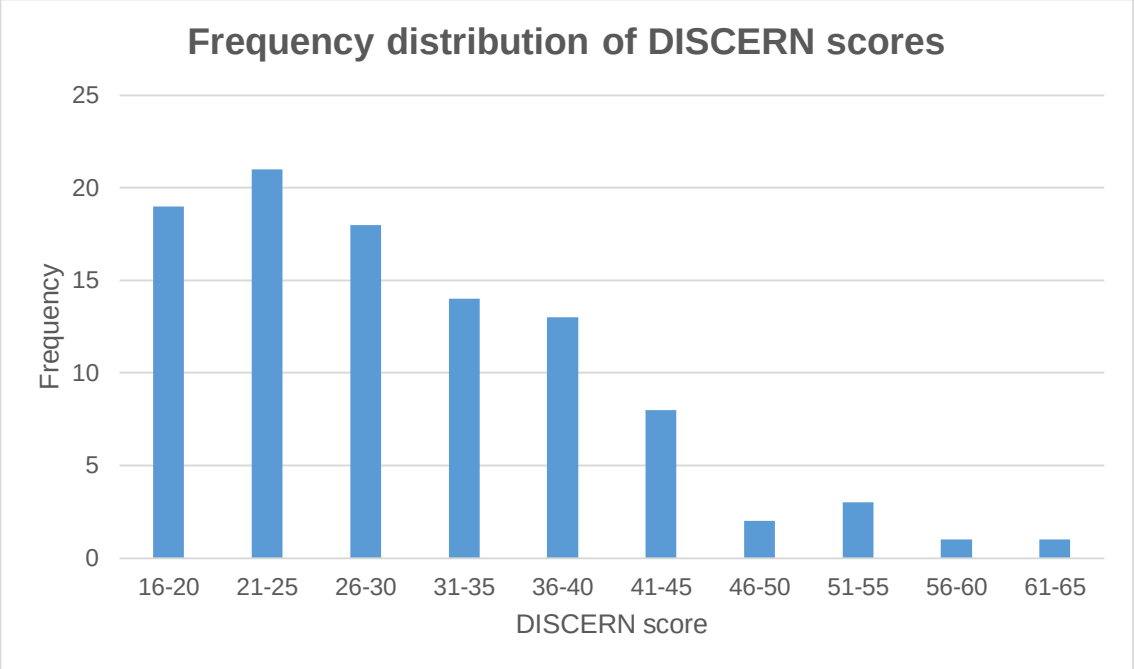


Figure 5: The frequency distribution of DISCERN scores

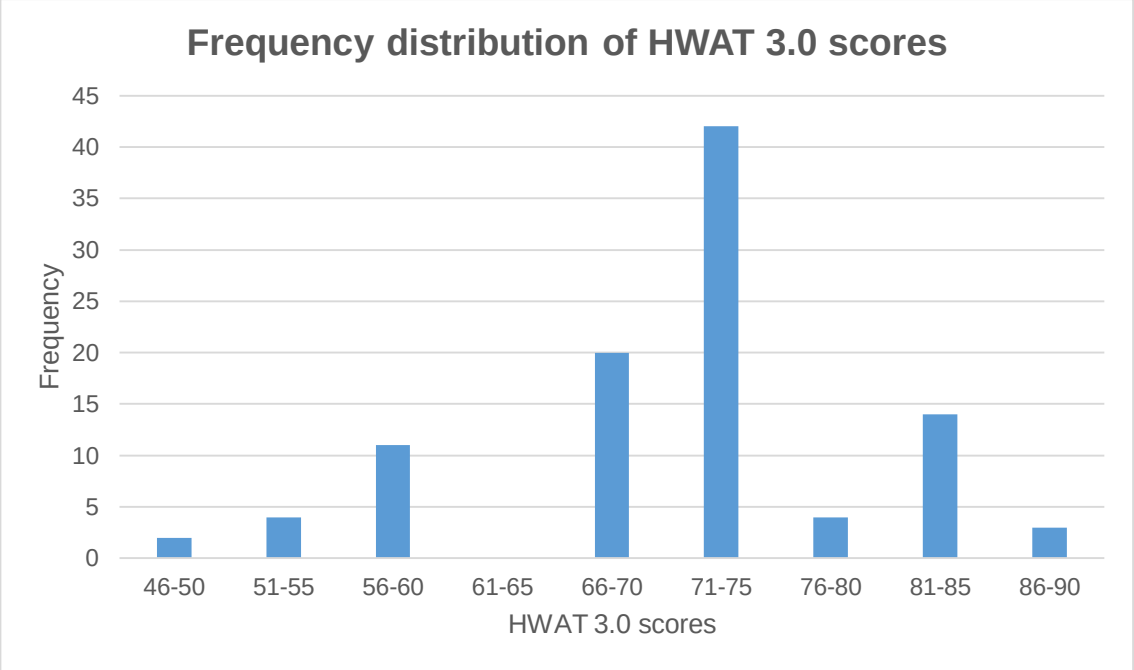


Figure 6: The frequency distribution of HWAT 3.0 scores

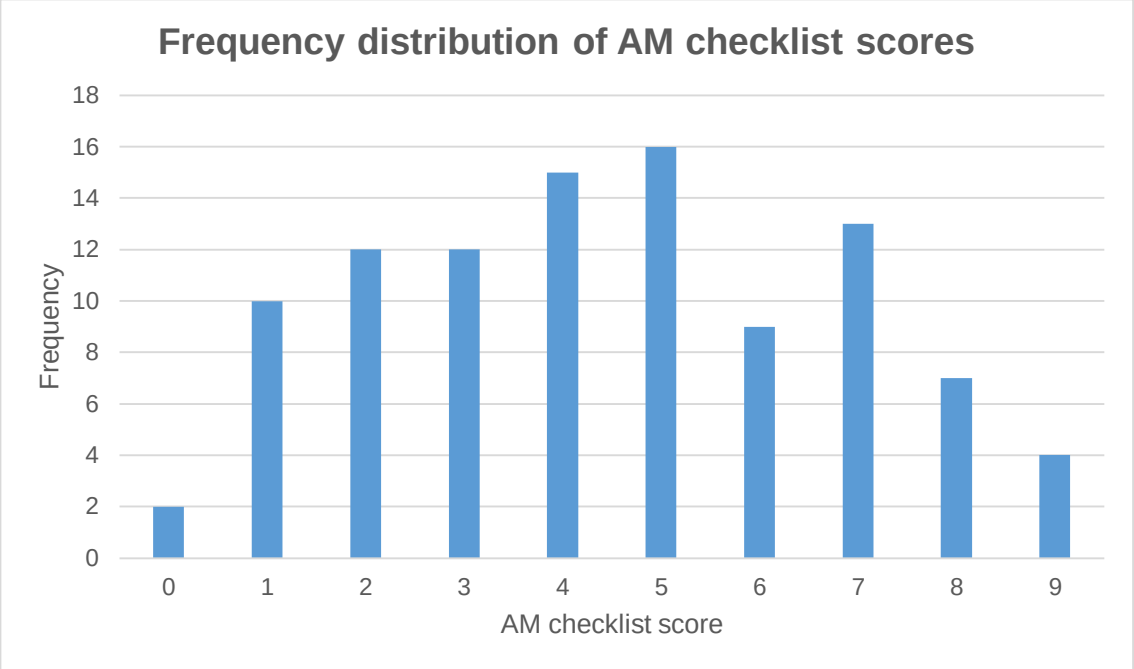


Figure 7: The frequency distribution of the AM checklist scores

3.4.8.2 Comparison of rankings between the three tools and between the tools and the AM checklist

The correlation of the scores between the different tools and the AM checklist was calculated using Pearson's correlation. Pearson's correlation was used because it assesses the strength of the linear correlation between two sets of normally distributed data.

3.4.8.3 Quality according to Google ranking

The correlation of the scores between the three tools and AM checklist with Google ranking was also calculated using Pearson's correlation for the same reasons described in section 3.4.8.2.

3.4.8.4 Inter- examiner and intra- examiner reliability

The strength of correlation between the scores of the primary researcher for intra- examiner reliability and primary researcher and S.H. for inter- examiner reliability was tested using the intraclass correlation coefficient (ICC). This statistical method was used to assess exact

agreement between the scores on the two different occasions for each tool or AM checklist for all 10 websites.

3.4.8.5 Quality according to suffix

Unfortunately, because of the large variation in the number of websites in the suffix groups it was not possible to carry out statistical analysis across all suffix groups as most groups occurred in small numbers. To be able to carry out statistical tests for comparisons between suffix groups, some suffixes were combined to form a '.com' group and an 'other' group. All the scores from the websites with suffixes '.com', '.co.uk', '.info', '.co.nz' and '.au' were combined to give a mean score for the '.com' group. The 'other' group consisted of a combined score for websites with suffixes '.org', '.edu', and '.nhs.uk' to give a mean score which could then be compared. The websites were grouped this way to divide professional organisation websites such as hospital websites from non-professional organisation and commercial websites as much as possible. Significance of the difference between the mean scores of the '.com' and 'other' groups was tested with the t-test for independent samples because there were two categories.

3.4.8.6 Quality according to country of origin

Because of the very small number of websites originating from some countries such as Australia, Spain and New Zealand, statistical tests could only be applied to the websites originating from the USA and the UK. The two tailed t- test was used to assess statistical significance between the mean scores of UK and USA websites using the three tools and the AM checklist.

3.4.8.7 Quality according to type of website

Due to their small size, the data for some groups, e.g. support groups, question and answer and video sites, was combined in order to allow statistical tests to be applied. Firstly, an ANOVA was carried out, which showed significant differences between the mean scores of the three tools, the AM checklist and also within groups. Tukey's HSD test was performed to

identify significant differences between website types for each tool, the AM checklist and Google ranking. ANOVA was then used because there were more than two categories. Tukey's HSD test was used as this test is applied to normally distributed data to determine which means within groups are significantly different from each other.

3.4.8.8 Quality according to British Orthodontic Society (BOS) accreditation and sponsorship

The independent sample t- test was planned to be used to assess significant differences of mean scores of website quality according BOS accreditation and sponsorship.

4. Ethics, sponsorship and data handling

4.1 Ethics

The Research Governance Officer confirmed on 3 June 2013 that ethical approval was not required for this project because this study did not involve participants or privileged data and so could be categorised as an evaluation of service.

4.2 Data handling and confidentiality

The websites are placed on the Internet by their developers and it can be assumed therefore that the information contained is not confidential because it is in the public domain and can be used for analysis. Website names and URLs were not therefore anonymised .

5. Results

5.1 Comparison of rankings between the three tools with the AM checklist

The results of the Pearson's correlation test for the three existing tools and the AM checklist are described in Table 2. A strong correlation was found between the AM checklist and DISCERN ($r= 0.816$, $r^2=0.67$ $p<0.01$). The null hypothesis that there is no correlation between the orthognathic website quality rankings produced by LIDA, DISCERN and HWAT 3.0 and the AM checklist was therefore rejected. This showed that there was high agreement between DISCERN and the AM checklist for which websites both considered as higher quality and thus were ranked higher on the overall score rank list. This result also meant that there was agreement about the websites which were lower quality, had scored lower and thus were further down the overall total score rank for both DISCERN and the AM checklist. These results indicate that DISCERN is effective at correctly identifying high quality orthognathic websites and distinguishing them from lower quality sites by allocating a lower score thereby ranking them lower. No correlations were found between LIDA or HWAT 3.0 with the AM checklist showing that these tools were unable to correctly identify high quality orthognathic websites.

Table 2: Pearson's correlation between the 3 tools and the AM checklist.

	LIDA	**DISCERN**	**HWAT**	**AM checklist**
LIDA	1	0.043	0.394**	0.069
DISCERN		1	-0.228*	0.816**
HWAT			1	-0.150
AM checklist				1

*= $p<0.05$, **= $p<0.01$

5.2 Comparison of rankings between the tools

Table 2 also shows that there was a weak but statistically significant correlation between LIDA and HWAT 3.0 ($r= 0.394$, $r^2= 0.16$, $p<0.01$). The null hypothesis that There is no difference in the orthognathic website quality ranking produced from LIDA, DISCERN and

HWAT 3.0 could therefore not be rejected. However, although this correlation was positive and significant because $p<0.01$, it was a very weak correlation which means that there was only a 16% association between LIDA and HWAT 3.0. There was also a weak and negative correlation between DISCERN and HWAT 3.0 ($r= -0.228$, $r^2= 0.05$, $p= 0.023$) which was significant. This means that a website which ranks higher on DISCERN is likely to have been ranked much lower by the HWAT 3.0 tool and vice versa. Also, if a website is ranked lower by the DISCERN tool it is possible that the same website would be scored higher by the HWAT 3.0 tool, potentially giving misleading information as to which site was of higher quality.

Overall, the results showed that the existing tools used in this study were poorly correlated with each other.

5.3 Quality according to Google ranking

A weak but significant negative correlation was found between the Google ranking and the AM checklist and also with DISCERN (Table 3), which means that websites which were ranked higher using the AM checklist and DISCERN ranked lower in Google, i.e. fell near the top of the Google search results. The null hypothesis that there is no correlation between the quality of orthognathic websites and Google ranking was therefore rejected.

The correlation between the Google ranking and the AM checklist of -0.347 ($r^2=0.12$, $p<0.01$) and between the Google ranking and DISCERN of -0.315 ($r^2=0.1$, $p<0.01$) means that Google only accounts for 12% of the variance in the AM score and 10% in the DISCERN score. Therefore, although there was a negative and significant correlation, it was weak. This would indicate that the ranking in Google alone cannot be used as a quality marker for orthognathic websites.

Table 3: Correlation between the ranking produced by the tools, AM checklist and Google ranking

	LIDA	DISCERN	HWAT	AM checklist
Google rank	0.038	-0.315*	0.035	-0.347*

*= $p<0.05$

5.4 Measurement of reliability of AM checklist in pilot study

The first time the reliability scores were calculated for the AM checklist version 1, the intra-examiner reliability was 0.659 (95% confidence interval 0.083 to 0.904) and inter-examiner reliability was 0.127 (95% confidence interval of -0.389 to 0.650).

After improvements were made to the AM checklist, both the inter- and intra-examiner reliability tests were repeated for AM checklist version 2 and the results showed significant improvement. The intra-examiner reliability improved to 0.972 (95% confidence intervals 0.760 to 0.994) and the inter-examiner reliability improved to 0.780 (95% CI 0.369 to 0.939). Although these confidence intervals were wide suggesting low precision, these results showed an overall large improvement of the AM checklist.

5.5 Intra-examiner reliability of LIDA, DISCERN and HWAT 3.0

The results of the intra- examiner reliability test (Table 4) showed that LIDA had good, DISCERN moderate, and HWAT 3.0 had poor intra-examiner reliability, which in turn meant that, for example, the HWAT 3.0 tool had a high likelihood of producing different sets of results if it was used on two different occasions (even by the same investigator) and therefore, was unlikely to produce consistent results.

Table 4: Intraclass correlation coefficient (ICC) for intra-examiner reliability with 95% confidence intervals

Tool	ICC	95% confidence interval
LIDA	0.82	0.16 to 0.96
DISCERN	0.65	0.08 to 0.90
HWAT 3.0	0.49	-0.13 to 0.84

5.6 Inter- examiner reliability of LIDA, DISCERN and HWAT 3.0

The results of the inter- examiner reliability (Table 5) showed that DISCERN had moderate inter-examiner reliability but LIDA and HWAT 3.0 had moderate/good inter-examiner reliability. There was therefore a moderate to high chance of two examiners producing similar scores using LIDA or HWAT 3.0. The null hypothesis that there is no difference in the inter- and intra-examiner reliability of the three health website quality assessment tools was rejected. However, it should be noted that the confidence intervals for all ICCs and tools were wide which reduced the precision of the data.

Table 5: Intraclass correlation coefficient (ICC) for inter-examiner reliability with 95% confidence intervals

Tool	ICC	95% confidence interval
LIDA	0.72	0.19 to 0.92
DISCERN	0.68	0.10 to 0.91
HWAT	0.79	0.36 to 0.94

5.7 Quality according to suffix

Most websites had the .com or .co.uk suffix (n= 76). Twelve websites had the .org suffix and two websites had the .edu suffix (Figure 8).

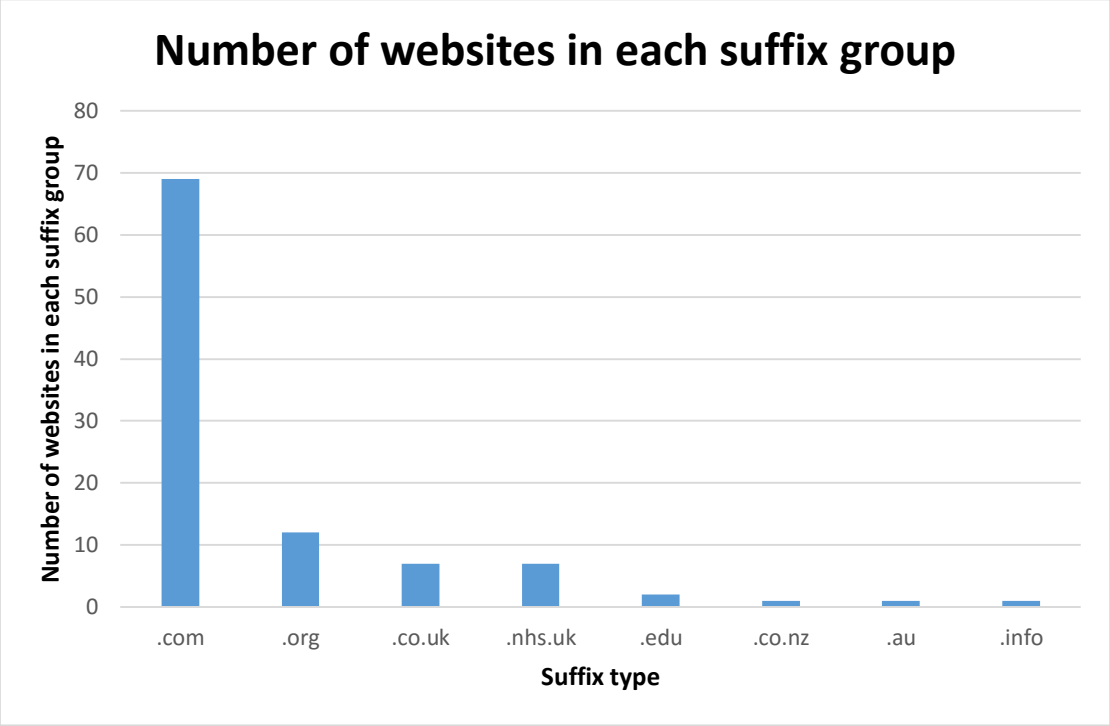


Figure 8: The number of websites in each suffix group.

As discussed in section 3.4.8.5 due to the small number of websites in some suffix groups, the websites were categorised into two groups; '.com' and 'other'. The mean scores for each tool and the AM checklist were calculated for the '.com' group and the 'other' group and mean difference was calculated (Table 6). This showed that the means of the 'other' group was almost always higher than that for the '.com' group. This was the case for the smaller number of websites in the 'other' group (n=21) compared to the '.com' group (n= 79). However, this was only significant for the DISCERN and HWAT 3.0 tools and not for the AM checklist. Therefore the null hypothesis that there is no correlation between the quality of orthognathic websites and suffix could not be rejected.

Table 6: Mean score from LIDA, DISCEREN, HWAT 3.0, AM checklist and mean rank position from Google for suffixes of '.com' encompassing .com, .co.uk, .aus, .co.nz, .info and 'other' encompassing .nhs.uk, edu, and .org.

Method of assessment	Suffix	Mean	Mean difference	P value	95% Confidence intervals	
					Lower limit	Upper limit
LIDA	.com other	101.19 102.42	-1.24	0.64	-2.38	2.22
DISCERN	.com other	28.54 34.86	-6.31*	0.01	-11.17	-1.45
HWAT 3.0	.com other	70.83 76.57	-5.75*	0.01	-10.10	-1.39
AM checklist	.com other	4.27 5.24	-0.97	0.09	-2.10	0.16

*=p<0.05

5.8 Quality according to country of origin

The majority of the websites originated in the USA (n=52). The UK was the second most common country of origin (n=23) (Table 7). Video diary websites and focus groups had contributors from multiple countries and only two websites had unclear countries of origin.

As mentioned in section 3.4.8.6, as a very small numbers of websites originated from countries other than the UK and USA, statistical tests were only applied to these two countries. The mean scores of each tool and the AM checklist for UK and USA based websites was calculated. The two tailed t-test was used to assess statistical significance of the difference in mean scores between UK and US websites.

This test showed that websites from the UK scored higher compared to the USA using the AM checklist and DISCERN and this was statistically significant ($p<0.05$) therefore the null hypothesis that there is no correlation between the quality of orthognathic websites and country of origin was rejected.

. This means that UK websites were rated as higher quality when assessing orthognathic websites. Websites from UK could therefore be used as a marker of high quality compared with websites from the USA.

Table 7: Mean and standard deviation of scores from the 3 tools and AM checklist according to country of origin. The last row shows the mean difference between the scores of US and UK websites.

Country	Statistical analysis	LIDA	DISCERN	HWAT 3.0	AM checklist
Australia/ New Zeeland (n= 3)	Mean Standard Deviation	106.67 9.86	30.00 13.74	76.00 7.54	4.67 2.08
Europe (N= 4)	Mean Standard Deviation	97.00 16.71	25.5 9.81	61.50 9.94	4.00 3.56
South/ Central America (n= 2)	Mean Standard Deviation	101.50 4.95	24.5 4.95	72.00 4.24	3.50 2.12
Canada (n= 2)	Mean Standard Deviation	102.60 13.39	26.8 7.46	73.80 10.73	5.20 1.64
Asia (n= 5)	Mean Standard Deviation	99.20 7.69	24.60 8.26	72.00 9.24	3.40 3.13
Multiple/ Unknown (n= 6)	Mean Standard Deviation	97.17 15.70	36.17 10.48	58.50 11.02	5.33 2.42
UK (n= 23)	Mean Standard Deviation	102.39 7.63	35.65 11.96	71.74 10.31	5.39 2.25
US (n=52)	Mean Standard Deviation	101.67 11.53	27.92 8.85	74.13 6.80	4.06 2.23
US v UK	Mean difference	0.72	7.73*	-2.40	1.33*

n= total number of website originating from each country

*= $p<0.05$)

5.9 Quality according to type of website

Several website types were identified (Table 8). The most common type was private practice websites offering orthognathic treatment (n=54) followed by hospitals (n=20) and blogs (n=10). Other website types included professional organisations, support groups, a question and answer (Q&A) page and patient information websites written by clinicians or unknown authors.

Table 8: Mean and standard deviations of scores for website types for LIDA, DISCERN, HWAT 3.0, the AM checklist and Google ranking.

Type of website	Analysis	LIDA	DISCERN	HWAT	AM checklist	Google rank
Blog (n= 10)	Mean Std dev	106.50 10.93	37.80 7.51	66.6 7.18	6.2 1.40	101.5 68.02
Hospital (n= 20)	Mean Std dev	100.95 6.30	33.05 13.76	77.10 9.54	4.90 2.71	76.90 50.12
Professional organisation (n= 6)	Mean Std dev	102.40 7.30	28.80 7.73	76.80 7.82	5.00 1.87	45.25 53.30
Private practice (n= 54)	Mean Std dev	101.06 12.54	25.87 7.71	72.44 7.54	3.67 2.13	124.06 47.87
Other patient information (n= 6)	Mean Std dev	100.00 7.61	35.00 8.14	65.00 10.33	5.83 1.94	66.90 54.56
Support group, Q &A, video sites (n= 5)	Mean Std dev	98.40 12.50	39.40 8.05	61.80 12.66	5.80 2.28	77.20 77.18

As outlined in section 3.4.8.7 several types of websites were combined in order to allow statistical analysis to be carried out. Table 9 shows the results of the ANOVA and Tukey's HSD test outlining significant differences in means of different types of websites.

Table 9: Difference in the mean scores according to website type for LIDA, DISCERN, HWA 3.0, A.M checklist and Google ranking.

		Mean difference			
Type of website	**Comparison against (type of website)**	**LIDA**	**DISCERN**	**HWAT 3.0**	**AM checklist**
Blog	**Hospital**	5.55	4.75	-10.50*	1.30
	Professional organization	4.10	9.00	-10.20	1.20
	Private practice	5.44	11.93*	-5.84	2.53*
	Patient information	6.50	2.80	1.60	0.37
	Support groups, videos, Q&A	8.10	-1.60	4.80	0.40
Hospital	**Blog**	-5.55	-4.75	10.50*	-1.30
	Professional organization	-1.45	4.25	0.30	-1.00
	Private practice	-0.11	7.18*	4.66	1.23
	Patient information	0.95	-1.95	12.10*	-0.93
	Support groups, videos, Q&A	2.55	-6.35	15.30*	-0.90
Professional organisation	**Blog**	-4.10	-9.00	10.20	-1.20
	Hospital	1.45	-4.25	-0.30	0.10
	Private practice	1.34	2.93	4.36	1.33
	Patient information	2.40	-6.20	11.80	0.83
	Support groups, videos	4.00	-10.60	15.00	0.80
Private practice	**Blog**	-5.44	-11.93*	5.84	-2.53*
	Hospital	0.11	-7.18*	-4.66	-1.23
	Professional organization	-1.34	-2.93	-4.36	-1.33
	Patient information	1.06	-9.13	7.44	-2.17
	Support groups, videos, Q&A	2.66	-13.54*	10.64	-2.13
Patient information	**Blog**	-6.50	-2.8	-1.60	-0.37
	Hospital	-0.95	1.95	-12.10	0.93
	Professional organization	-2.40	6.20	-11.80	0.83
	Private practice	-1.06	9.13	-7.44	2.17
	Support groups, videos, Q&A	-1.60	-4.40	3.20	0.33
Support groups, videos, Q&A sites	**Blog**	-8.10	1.60	-4.80	-0.40
	Hospital	-2.55	6.35	-15.30	0.90
	Professional organization	-4.00	10.60	-15	0.80
	Private practice	-2.66	13.54*	-10.64	2.13
	Patient information	-1.60	4.40	-3.20	-0.33

*=p<0.05

The groups containing patient written websites (blogs and support groups, videos and Q&A groups) had the highest mean scores across the AM checklist, LIDA and DISCERN compared to other types of websites. The AM checklist showed that blogs, support group websites, video websites and Q&A websites were comparable to hospital, professional organisation and patient information websites. There was no statistically significant difference between the mean scores of these groups ($p > 0.05$). Furthermore, blogs were the only group which had a statistically significantly higher mean score than private practice websites ($p<0.05$), which was generally the lowest scoring group according to the AM checklist, DISCERN and LIDA. The null hypothesis that there is no correlation between the quality of orthognathic websites and website types was therefore rejected.

Other statistically significant differences found were as follows; DISCERN showed hospitals, and support groups, videos and Q&A websites to have significantly higher scores than private practice websites ($p<0.05$). HWAT 3.0 showed that hospitals had significantly better scores than blogs, patient information and support groups, videos and Q&A websites ($p<0.05$). Google ranked professional organisation and hospital websites further towards the top of the search results than private practice websites.

5.10 Quality according to BOS accreditation

Only one website had BOS accreditation, so this analysis could not be carried out.

5.11 Quality according to Sponsorship

None of the websites in this study had any indications of sponsorship, so this analysis could not be carried out.

5.12 Summary of main results

In summary the analysis found the following key findings:

1. There was a strong and statistically significant correlation between the AM checklist and the DISCERN tool. No correlation was found between the AM checklist and LIDA or HWAT 3.0.

2. No correlations were found between LIDA, DISCERN and HWAT 3.0.
3. No correlation was found between the AM checklist and the Google ranking.
4. The AM checklist and LIDA tool have acceptable reliability.
5. Website suffix cannot be used as a marker of quality.
6. UK orthognathic websites had statically significantly higher quality than USA websites.
7. Blogs, support groups, video websites and question and answer websites have comparable quality to hospital and professional organisation websites.

6. Discussion

This study objectively assessed the ability of HWQATs to identify websites with the best content and discriminate these from poorer quality websites. Potential indicators of high quality websites were also explored.

6.1 Discussion of main results

6.1.1 Comparison of rankings between the three tools with the AM checklist

6.1.1.1 Comparison of LIDA and HWAT 3.0 with the AM checklist ranking

No correlations were found between the AM checklist and LIDA or HWAT 3.0 (section 5.12 finding 1). This lack of correlation is most likely because of the questions asked in LIDA and HWAT 3.0. These questions mainly focused on general attributes and the functionality of the websites such as clarity, link functioning and the consistency of the layout throughout the websites. Examples of such questions in LIDA were:

2.1.3 Is the layout of the main block of information clear and readable?

2.1.6 Is the colour scheme appropriate and engaging?

2.3.3 Does the design minimise the cognitive overhead of using the site?

2.3.1 Does the site provide an effective search facility?

2.3.2 Does the site provide effective browsing facilities?

Only the following two questions were loosely reflective of the website content.

2.1.2 Is the level of detail appropriate to their level of knowledge?

3.1.1 Does the site respond to recent events?

Similarly the HWAT 3.0 tool also mainly assessed general parameters with questions such as:

1. Originating person/ organisation identification
2. Clearly stating the subject of the website

3. Is the author or institution providing the information stated?
4. Do the interwebsite links function?

Only two questions were reflective of website content.

1. Basic information e.g. definition, consequence, prevention and treatment
2. Is the source of information credible?

The AM checklist was specifically designed to ascertain whether scores from existing tools correlated with websites of high quality content, and its questions were geared to assess the topic- specific content. As LIDA and HWAT 3.0 contained very few content-specific questions, it is not surprising that a correlation was not found in the ranking of websites from the AM checklist and LIDA or HWAT 3.0.

There are no studies, as far as the authors of this study are aware, which have assessed the correlation of LIDA and HWAT 3.0 with a standard (such as the AM checklist). Therefore, these findings from our study cannot be compared to previous studies.

6.1.1.2 Comparison of DISCERN and AM checklist rankings

The results of this study show that DISCERN had a strong and statistically significant correlation with the AM checklist ($r= 0.82$, $p<0.01$) (section 5.12 finding 1). This means that DISCERN was the only tool which was able to correctly identify the higher quality websites and discriminate between these and the poorer quality ones. This strong correlation is likely to be because of the questions used in the DISCERN tool. Of 16 questions, the first eight were about general quality parameters such as whether the information on the website was unbiased, whether a source for further information was present and if it was clear when the information was published. However, this tool also asked the following more specific content-related quality questions:

1. Does it describe how each treatment works?
2. Does it describe the benefits of each treatment?

3. Does it describe the risks of each treatment?
4. Does it describe what would happen if no treatment is used?
5. Does it describe how the treatment choices affect overall quality of life?
6. Is it clear that there may be more than one possible treatment choice?
7. Does it provide support for shared decision making?

These questions were similar to checklist questions on the AM checklist. For example:

1. Description of procedure including an explanation of the surgery and the fixation of bones using plates and screws
2. Risks
3. Benefits
4. Alternative treatment

The AM checklist did not have questions which were included in DISCERN, for example, about the impact of treatment on the overall quality of life or about shared decision making. Similarly, DISCERN did not include many of the questions in the AM checklist such as whether general anaesthesia is mentioned, questions about the length of treatment and the recovery period. Despite these differences the correlation was still strong. This could be because websites which had the information required by the AM checklist to give a higher score were more comprehensive and also included material related to the questions asked in DISCERN about website content. Furthermore these websites might have been more comprehensive and therefore likely to contain the general information asked by DISCERN.

The strong correlation between DISCERN and the AM checklist was much higher than the correlations found by Hsu and Bath in 2008 (Table 10). These investigators assessed the correlation between three tools (DISCERN, IQ [information quality] tool and the HoN code) and their newly developed breast cancer (BC) tool when applied to breast cancer information on the Internet. Analysis of 40 breast cancer websites yielded a moderate correlation of 0.43 (using the Kendall's Tau B test) between the DISCERN and the BC tools. The difference in

the correlations between DISCERN and the BC tool and DISCERN and the AM checklist described here could be because of the difference in the way that the AM checklist was developed. The BC tool was developed from a search of the available literature, based on the information needs of breast cancer patients. In contrast, the AM checklist was compiled based on the information which clinicians consider essential for patients to know about orthognathic treatment. Additionally, DISCERN was developed predominantly by clinicians with a limited amount of patient input, therefore the questions in DISCERN and the AM checklist would be different to the questions in the BC tool.

Another study, by Khazaal et al. (2012), that assessed the correlation between a developed standard proforma to assess websites about mental health-related issues with DISCERN reported that DISCERN was a good tool to identify high quality websites. In this study, a large sample of websites was assessed with their proforma and DISCERN and the correlation between the two was found to be 0.77 using a Receiver Operating Characteristic curve. Their proforma did not appear to have any patient input, which could explain the similar positive correlation found in our study. However, the proforma was based on guidance from the literature rather than specific input from clinicians, as was the case in our study. Overall, it appears that DISCERN has good ability to identify higher quality websites and discriminate between poorer quality websites.

Table 10: Summary of previous studies assessing the correlation between DISCERN and a standard (such as the AM checklist used in this study)

Study	Tools assessed	Standard used	Correlation
This study	DISCERN	AM checklist	0.82*
Hsu and Bath, 2008	DISCERN	BC tool	0.43*
Khazaal et al., 2012	DISCERN	Authors developed proforma	0.77*
Surman and Bath, 2013	DISCERN	Stroke tool	0.62*

*= $p<0.01$

6.1.2 Comparison of rankings between the tools

The different questions in each tool are also likely to be the reason behind the lack of correlation between the three tools (section 5.12 finding 2). Because each tool assessed each website with its own set of criteria, it is not surprising that rankings were not similar or correlated.

The lack of correlation between the tools in our study agrees with the findings of Prusti et al. (2012), who assessed a smaller sample of 40 anti- depressant information websites (Table 11). They also found a weak Pearson's correlation coefficient of 0.49 ($p<0.05$) between the DISCERN and DARTS tools and also concluded that the DISCERN tool was more reflective of the contents of the websites but the DARTS tool better assessed the aesthetic aspects of the websites.

The findings of other studies disagreed with our findings. Surman and Bath (2013) found a strong correlation between DISCERN and the HoN code and a moderate correlation between DISCERN and the stroke tool (content- specific tool similar to the AM checklist) when these tools were applied to 51 websites that assessed speech and language difficulty after a stroke. Furthermore, Breckons et al. (2008) assessed the correlation between 12 HWQATs (including LIDA and DISCERN) when applied to breast cancer websites. They applied all of the tools to a small sample of 12 websites and tested for correlation using the non-parametric Spearman rank correlation test. They found a correlation between 10 out of 12 tools although the correlation between LIDA and DISCERN of 0.7 was not statistically significant.

Table 11: Summary of previous studies assessing the correlation between existing tools.

Tools assessed for correlation		Correlation			
		This study	**Breckons et al, 2008**	**Surman and Bath, 2013**	**Prusti et al., 2012**
LIDA	**DISCERN**	0.82**	0.7		
LIDA	**HWAT 3.0**	0.39**			
DISCERN	**HWAT 3.0**	0.23*			
DISCERN	**HON**			0.704**	
DISCERN	**DARTS**				0.49*

*= p<0.05, **= p<0.01

Although previous studies show contradicting information about the correlation between the tools it should be noted that most of these studies had smaller sample sizes than our study, particularly Breckons et al. (2008) which sampled only 12 websites. The use of a larger sample size to detect a true correlation between the tools is extremely important because if a small sample is used, the potential for bias is increased because the sample could contain websites of clear and diverse qualities which are easily ranked high by most tools. However, the use of a larger sample like our study is more likely to ensure that websites with a wide variation in quality are analysed and therefore the ranks are more representative of their associated tools for comparison. Additionally, our large sample indicated that the data was generally normally distributed, which allowed the use of parametric tests which might be more appropriate.

6.1.3 Quality according to Google ranking

The weak but statistically significant negative correlation between the Google ranking and the AM checklist suggests that Google can, to a very limited extent, identify websites of higher quality compared to those of lower quality by ranking a website higher in its search results (section 5.12 finding 3). However, because the correlations were weak (r= -0.315 between Google and the AM checklist and r= -0.347 between Google and DISCERN) the

Google ranking alone cannot be used as a marker of quality to discriminate higher quality from lower quality websites.

The reason for the weak correlation between the Google ranking and website quality according to the AM checklist is likely to be because of the difference in the way quality is determined by Google and the AM checklist. Google uses a sophisticated method to search for relevant websites described in section 2.7.1. This method places great emphasis on the number of times a particular website is cited by other websites because the developers believe that the number of citations and links from other websites indicates website quality.

Our study shows that websites that ranked higher on the Google search and thus were likely to have been cited more than others did not necessarily have the highest quality content as measured by the AM checklist. This means that websites which are sometimes ranked low in the Google search results were of higher quality with useful information. However, they were unlikely to be seen by patients because they were low enough that they appeared on the 10^{th} or 11^{th} page of the search results as patients are unlikely to look beyond the first few pages of search engine results (Eysenbach and Kohler, 2002). For example the website www.nycoms.com was ranked 152 on the Google search carried out in our study but had a high score of 7 on the AM checklist, indicating that this website had high quality content. The method by which Google ranks websites makes it very difficult for such a website to be placed towards the top of the search results. This is because websites attempting to link to other relevant websites are likely to use the top results from a search engine. Websites such as www.nycoms.com are likely to remain unnoticed unless higher quality websites are actively sought by looking beyond the first few results from the search engine, which we did in our study.

The weak correlation between the Google ranking, the AM checklist and other existing HWQATs was echoed by a recent study by Kirthi and Modi (2012), which found no correlation between search engine ranking and LIDA scores for 64 coronary angioplasty

websites. However, statistical analysis was not carried out in this study. The authors came to this conclusion by simply descriptively comparing the rankings of the top 10 scoring websites from LIDA with the highest ranking position from any of three search engines used in their study, including Google.

A study by Perez-Lopez (2004) also agreed with our findings. This study assessed 94 menopause websites for content quality and accuracy using general and specific quality parameters described in an earlier study by Sandvik (1999) which included seven parameters. This study concluded that the popularity of a website in a search engine does not necessarily indicate a higher quality website with more accurate medical information. However again in this study, only a descriptive comparison was made for the top 10 highest quality websites according to Sandvik's parameters (Sandvik, 1999). The Google ranking appeared to have a poor correlation with the top 10 high scoring websites because some of these websites were ranked further down the Google ranking.

However, findings from Best et al. (2014) who assessed the correlation between the Google ranking and the scores from four existing tools when applied to information about head and neck cancer, disagree with the findings of our study. In that study, which analysed 40 websites, significant correlations were found between the Google ranking and all four existing tools used in the study; SMOG (Simple Measure of Gobbledygook) readability tool, AMA (American Medical Association) guidelines, LIDA and DISCERN. However, the Best et al. (2014) conclusions were based on regression coefficients which indicate the presence of a correlation but does not give an indication of the strength of the correlation. Furthermore, the sample size was much smaller than the sample size in our study. Moreover, the Google ranking was compared with existing, unvalidated tools and unlike our study that used the AM checklist, this study did not have an objective method to determine website content quality and correlate it with the Google ranking.

Overall, it appears that the Google ranking is largely based on a website's popularity and the number of times that particular website is cited. Although some evidence (Best et al., 2014)

indicates that the Google ranking was a marker of quality for certain medical websites our study found that this was not true for orthognathic websites. Therefore users should be cautious about relying on Google ranking to find high quality orthognathic websites.

6.1.4 Reliability

The results of the reliability tests showed a wide variation. LIDA had the best overall reliability and HWAT 3.0 the worst. Overall, inter-examiner reliability was lower than intra-examiner reliability for DISCERN and HWAT. The reliability of each tool is individually discussed:

LIDA

This study found both the intra-examiner and inter-examiner reliability (ICC 0.82 and 0.72 respectively) for the LIDA tool to be moderate/high. However, having used this tool, the authors feel that this evaluation might not be an accurate appraisal. With 29 questions, LIDA had the largest number of questions of all three existing tools. However, the answers to the following 11 questions were almost always the same:

1.5 Browser Test

1.6 Registration

2.1.4 Is the navigation clear and well structured?

2.1.5 Can you always tell your current location in the site?

2.2.1 Is the same page layout used throughout the site?

2.2.2 Do navigational links have a consistent function?

2.2.3 Is the site structure (categories or organisation of pages) applied consistently?

2.3.4 Does the site support the normal browser navigational tools?

2.3.5 Can you use the site without third party plug-ins?

2.4.1 Can the user make an effective judgement of whether the site applies to them?

2.4.3 Can the user personalise their experience of using the site?

These 11 questions, which are a large proportion of the 41 questions, mainly assessed if the website was operating and accessible. This was assessed by evaluating factors such as whether the website worked in common browsers, whether registration was required to access the webpage and whether the navigation tools in the browser ('Go back' or 'Go Forward' functions) could be used. The scores for most of these questions was the same for 91 websites because as part of this study's exclusion criteria, any websites that required registration or any non-functioning website was not included in any analysis, so some of these questions were not pertinent to the study and the scoring. We conclude therefore that there is potential for shortening this tool.

The good level of inter- examiner and intra- examiner reliability found in this study was in agreement and exceeds those found by the developers of LIDA (LIDA, 2004) who found the inter-examiner reliability to be 0.611 (using Spearman's rank correlation). This may be because of the additional guidance notes used both for intra- and inter-examiner reliability in this study (discussed in section 3.4.5.2). However, Downing et al. (2011) found higher inter-examiner reliability for LIDA of 0.86 (using the ICC) compared to 0.72 found in our study. Overall, the findings of our study and previous studies seem to suggest that LIDA has good inter-examiner reliability.

DISCERN

In this study, DISCERN was found to have moderate inter- and intra-examiner reliability (ICC 0.64 and 0.68 respectively) with wide confidence intervals indicating low precision of the ICC estimates. A major limitation of this tool was the Likert scale scoring system of 1- 5 which made it difficult to use for some questions. For example, questions asking about risks and benefits were difficult to score on a 1–5 scale because orthognathic treatment has several risks and benefits. Therefore it was difficult to score a website which, for example, mentioned some but not all of the risks of orthognathic treatment, so scoring for such a question could be prone to variation due to individual examiner judgement and bias depending on how strictly they wished to score such a question. Additionally it was possible

that some questions which had a ‘yes’ or ‘no’ answer were answered with a 1 for ‘no’ and 5 for ‘yes’; for example, question 5 was ‘Is it clear when the information used or reported in the publication was produced?’. However, this was an observation made by the author as a user of the tool, which highlights the risk of subjective scoring. Another example of subjectivity is the difficulty of answering question 16, ‘Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices’. This is a very subjective question which the researcher found difficult to score from 1 to 5.

The inter-examiner reliability reported in this study was much lower than that found in the study by Ademiluyi et al, 2003. Ademiluyi et al. (2003) tested the inter-examiner reliability and validity of three tools (DISCERN, IQ tool and QS [quality scale]) by applying the tools to smoking cessation websites. They found a higher level of inter-examiner reliability of ICC 0.823 than our study. This was calculated using the total scores from 22 websites which made up 25% of their sample. They also calculated individual question inter-examiner reliability which varied widely between questions. Similarly, Downing et al. (2012) found DISCERN to have an inter-examiner reliability of 0.82 (using the ICC) when they applied this tool to splenectomy information websites.

Interestingly, the developers of DISCERN did not report an overall agreement using the total scores which was how reliability was calculated in our study. Instead, the authors assessed the inter-examiner reliability for each question using the weighted kappa agreement. This ranged from 0.31 and was highest for question 16 at 0.53. The authors considered a kappa agreement equal or greater than 0.4 as acceptable but it could be argued that even a kappa agreement of 0.53, is a level of agreement inadequate for an HWQAT. Rao et al. (2012) also tested individual question reliability between three assessors when DISCERN was applied to dengue-related information on the Internet and using kappa agreement, found a reliability range from 0.21 to 1.0.

Overall, the results from this and previous studies show the reliability of DISCERN to be highly varied which requires improvement.

HWAT 3.0

In this study, the reliability of HWAT 3.0 was found to be the lowest compared to the other two tools. The intra-examiner reliability of ICC 0.49 with a very wide confidence interval meant that there was a 49% agreement between the scores for one examiner on two occasions. This intra-examiner reliability is of concern, but interestingly, the inter-examiner reliability of HWAT 3.0 of ICC 0.79 was much higher than the intra-examiner reliability. Again, the confidence interval was wide. The reason for this difference could be that the primary researcher, who assessed the intra-examiner reliability, found this tool difficult to use for orthognathic surgery. The low intra-examiner reliability score could therefore be reflective of the tool still being open to individual subjectivity.

The intra-examiner reliability score for HWAT 3.0 in this study was similar to that reported by the tool developers (Lewiecki et al., 2006), who reported this to be 88% between physician osteoporosis experts, 79% for osteoporosis nurse educators and physicians, and 71% for osteoporosis nurse educators. These figures represent the average agreement of all questions, i.e. the percentage agreement was calculated for each of the 13 questions individually, followed by calculation of the mean agreement across all questions (using the individual question-based reliability scores). Again, similar to DISCERN, the developers of HWAT 3.0 did not calculate the agreement level between the total scores produced from several websites, how the user should judge the website quality rather than from answers to individual questions. Also, small discrepancies between the agreement of individual questions would not have a significant impact on the reliability level when it is calculated using individual questions. However, such small discrepancies could have a much larger impact on the total score for different websites. For example, for the website ‘http://awimpsguidetoorthognathicsurgery.blogspot.com’, 100% agreement between 10 out of 13 HWAT 3.0 questions would indicate a reliability of 77%. However, the total scores for the website on two occasions were 72 and 48. This is a very large difference and reduced the ICC because this statistical analysis compared the total scores. Therefore, if an

agreement of 77% was used, it would convey a misleading impression of much better agreement.

Overall, agreement between individual questions might be useful to identify questions with lower reliability in order to develop further a tool. However, because the total score from a tool is used to indicate website quality to allow comparison to other websites, ultimately the total score must have a very high reliability for the tool to be deemed appropriate for general use.

6.1.5 Quality according to suffix

A statistical comparison could only be made between websites with the '.com' suffix and the 'other' groups (as discussed in section 3.4.8.5). This did not give an accurate representation of each individual suffix group such as '.org', '.edu' etc. However, according to the DISCERN and HWAT tools, the 'other' group had overall higher quality compared to the '.com' group. There was no significant difference found between the 'com' and 'other' group using the AM checklist, so the findings of our study suggests that the suffix type does not successfully distinguish websites of higher quality from those of lower quality and as such the website suffix cannot be used as a marker of website quality (section 5.12 finding 5).

The findings of our study agree with the findings of Dash et al. (2012), who found no statistically significant difference between malnutrition websites using general and content HWQATs utilised in previous studies. However, the sample size of their study was small (29) and had as little as two websites in the '.gov' group, so their results need to be interpreted with caution.

However, studies by Ansani et al. (2005) and Joshi et al. (2011) disagree with the findings of our study. In a study that assessed the quality of arthritis information websites, Ansani et al. (2005) found websites with the '.gov' suffix to have significantly higher content quality and '.com' websites to have the lowest quality compared to other suffix groups. Similarly, Joshi et al. (2011) found that for nutrition information websites, '.org' websites had the highest

content scores compared to other suffix groups. However, an ANOVA indicated significant difference only between the '.com' and the '.edu' groups although the number of websites in some suffix groups was extremely small. Both these studies assessed website quality using newly developed criteria obtained from the literature. Our study used existing tools and the AM checklist to ascertain website quality which is a more objective method of assessing website quality and therefore this disagreement cannot be deemed conclusive until further investigation has been carried out.

6.1.6 Quality according to country of origin

Statistical analysis could only be carried out between websites originating from the UK and USA which both had samples large enough to allow for it, as discussed in section 3.4.8.6. The results using the AM checklist and the DISCERN tool showed that UK websites were of a significantly better quality than US orthognathic websites (section 5.12 finding 6). This in turn means that websites from the UK can be used as a marker of higher quality orthognathic websites compared with USA websites. However, it should be noted that the UK settings of Google was used and therefore this may have an impact on the results. This potential limitation is discussed further in section 6.2.2.

The reason that UK websites had better quality than US websites is most likely due to the fact that a much larger proportion of UK websites discussed risk, the use of general anaesthetic, the overall length of treatment and the recovery period according to the AM checklist. Other parameters in the AM checklist had similar responses from US and UK websites.

This finding agrees with the conclusions reported by Thompson and Graydon (2009), who investigated the quality of websites with methotrexate information. In this study, quality assessment was based on descriptive statistics. The authors analysed the eight highest ranked websites using their newly developed MWAT tool (a modification of the HWAT 3.0 tool). Their analysis found that three of the top eight websites were from the UK compared to

one website from the USA. However, the sample size in this study was small (28 websites) and because statistical analysis was not carried out on all results, likely because of the small sample size, direct comparison is difficult and the findings of this study should be interpreted with caution.

Our results disagree with the findings of Ogunwale et al. (2009) who in a study of 43 Hip resurfacing websites, using a newly developed quality criteria found that European websites had the highest mean scores compared with websites from other regions/countries. British websites were analysed separately. However, this study only used descriptive statistics, making direct comparison with our study difficult.

6.1.7 Quality according to type of website

The finding that blogs had the highest overall mean score and therefore the highest quality according to their assessment by the AM checklist was unexpected. Blogs had statistically significantly higher scores than private practice websites using the AM checklist and DISCERN (section 5.12 finding 7) despite the fact that the number of blogs (10) was much lower than the number of private practice websites (50). Furthermore, no statistically significant differences were found between the mean scores of blogs, hospitals and professional organisation websites using the AM checklist and DISCERN. This means that blogs had comparable quality with hospital and professional organisation websites.

This finding was surprising because blogs are usually written by lay people and represent the perspective of a single individual which is thought to be prone to bias. However, this study found that blogs provided a very comprehensive account of the orthognathic treatment process because of their diarised structure. This enabled a realistic perspective of the long period for orthognathic treatment and particularly of the recovery period, which was often not covered as well by other types of websites. Furthermore, the AM checklist was developed using information routinely discussed at orthognathic treatment consultation clinics by clinicians present at this clinic. The findings of our study seem to suggest that patients also

value the same information which clinicians consider important (covered by the AM checklist) because they include this information in their blogs.

Similarly to blogs, the 'support groups, videos and question and answer' group that contained information sourced from patients also had generally higher mean scores than the other groups, according to the AM checklist and DISCERN. However, the difference in the mean was statistically significantly higher than only the private practice group on one occasion for the DISCERN tool. Overall, these information sources might prove to be extremely valuable for patients because they describe the treatment process from a patient's perspective, including complications and risks. At orthognathic clinics, the clinicians who describe the treatment to patients often do not have first-hand experience with the effects of the treatment on a patient's day to day life, which was discussed in much more detail on blog websites. Moreover, great emphasis was placed on the implications of, and difficulties with a liquid diet after surgery which appeared to have a much more significant impact on the patient's life than generally thought by clinicians.

There are many studies in the literature which use both existing and newly developed tools to assess website quality, but not many assess blog quality and the quality of other resources written by patients as a separate category. It is therefore difficult to make a direct comparison with other studies. For example, Farrell et al. (2006) compared the quality of urinary incontinence websites using general and specific quality criteria developed by the author. The quality of websites was compared according to broader categories than used in our study; professional websites (which included hospital sites), organisational websites (which included professional organisation websites), and commercial websites. This study used Tukey's HSD test to assess quality differences according to website type and found only one significant difference – that organisational sites had higher general quality than commercial sites. This study generally agreed with the findings of our study in that not many significant differences were found between the majority of the website types according to the different tools and the AM checklist. However, because patient written resources did not

appear to be included in Farrell's study, it is difficult to assess how these would have compared with other urinary incontinence website types.

Similarly, in the study by Best et al. (2014) head and neck cancer websites were categorised according to type into government, commercial, non-profit and university/hospital websites. This study used bivariate correlation and found significant correlations between scores from American medical association guidelines and government websites of 3.117 ($p < 0.05$) and between non-profit groups and SMOG of -1.69 ($p < 0.05$). All other correlations were not significant. Again, this generally agreed with the results of this study where very few significant differences were found for most of website types regardless of whether a tool or the AM checklist was used as the assessment method.

Thompson and Graydon (2009) reported a stronger association between the website type and quality, concluding that non-sponsored sources such as hospitals and national organisations were likely to have better quality than other website types. However, they drew this conclusion from the their top eight rated methotrexate websites (of 28 examined) all from non-sponsored sources and did not carry out further statistical analysis on the results, which is necessary before definitive conclusions can be made. Similarly, Nasser et al. (2012) concluded that commercial warfarin websites were of poor quality compared to other website types. However, this study had a small sample of 11 websites and did not appear to include patient written resources. Furthermore, none of these studies used an objective method (such as the AM checklist used in our study) to compare the results from the tools.

Overall, from the results of our and previous studies, the evidence that website type has an impact on the quality of websites is weak, so website type cannot be used as a quality marker. In our study, blogs had statistically significantly better scores than private practice websites, indicating that they contain the information which clinicians feel essential for patients to know (as assessed using the AM checklist). This finding highlights the potential

benefits of blogs in conveying information to fellow patients undergoing the same medical treatment as seen on orthognathic websites.

6.2 Study limitations

6.2.1 Sample

In this study, the first 100 websites resulting from a single search term in Google were used. This method has limitations. The sample of 100 websites was not based on a sample size calculation and ideally all 339 websites from the search should have been analysed to include as many websites with as diverse qualities as possible. However, this would have been extremely time- consuming beyond the scope of this study. A pragmatic view identified 100 websites to be a good representation of the 339 websites from the Google search.

The Internet constantly changes, so each time a search term is entered into a search engine the new search produces slightly different results. As this was a cross-sectional study only the results from a single search were used. However this method makes our study more generalisable because patients are likely to carry out only a single search and unlikely to carry out multiple searches with the same search term to find different websites (Eysenbach and Kohler, 2002).

The results of the Google search in this study only produced English- language websites and this study also only analysed websites in English (non English- language websites were excluded). This is a limitation because orthognathic websites in other languages could have potentially different quality than websites in English, which may impact on the study results.

The use of ‘orthognathic’ as the search term, rather than multiple search terms which patients might use, was another limitation. For example, patients might search for ‘jaw surgery’ or ‘orthognathic surgery’; Such different search terms would have likely produced a different list of websites and possibly impacted on the results.

6.2.2 Search engine

In our study, the websites investigated were produced from a single search engine. This might not be a true representation of all available orthognathic websites. Google was used as it is the most frequently used search engine and would make the study results generalisable and representative of the websites most likely to be accessed by UK patients. Previous studies in the literature often use results from multiple search engines (Patel and Cobourne, 2011; Downing et al., 2011; Livas et al., 2013). However, the aims of these studies differed from ours. These studies were assessing the quality of websites of specific health topics rather than carrying out an objective assessment of the tools themselves.

Using websites from a single search engine could potentially impact on the results since other websites of different quality may have been missed. However, this is unlikely as a relatively large sample was used, containing websites which may have appeared in other search engines.

Another limitation of this study was that the UK setting of Google was used because the study was carried out in the UK. This may have impacted the results because UK websites were likely to have been ranked towards the top of the search results, as shown in our study. Also this will have an impact on the comparison of quality according to country of origin. However, a UK patient is likely to come across the same UK setting of Google and therefore the results of this study are a good representation of orthognathic websites available for UK patients and possible markers of quality that can be used.

6.2.3 Examiner

In this study, a single examiner carried out all the website assessment ratings. Due to the subjective nature of some questions in the existing tools, examiner bias could therefore have impacted on the results. Furthermore, the websites were only examined on a single occasion and the assessments were not repeated. Close examination of inter- examiner and intra-

examiner reliability provides a measure of how much error is likely to have been introduced by this.

Another source of examiner bias was that the study was not blinded. The websites were saved offline and numerically coded, but almost all websites had the author/ origin displayed on the page. This could potentially impact the results and future studies should try to conceal the website origin. However, often the originator can still be identified from the website text. The best method to reduce examiner bias might have been to employ many examiners with different backgrounds and specialties who might not have been familiar with the quality of orthognathic websites.

6.2.4 Data capture form

To facilitate data analysis, question scores were recorded in a Google docs form that enabled the primary researcher to easily select scores for each question using a drop down menu. It is possible that the wrong number could have been inadvertently selected. It was also possible for a field to be left blank. Both of these scenarios, although unlikely due to the measures outlined in section 3.4.6, would impact the total score for a website. Although this is unlikely to have had an impact on the overall study findings, a separate selection tab for each number instead of a drop down menu might reduce the chance of a selection error using the cursor to make the appropriate selection. Compulsory questions would have removed the likelihood of missed questions.

6.2.5 Websites saved offline

Main page

As this was a cross-sectional study, all the websites should have ideally been assessed at the same time. However, the large number of websites to be assessed with multiple tools and the AM checklist precluded this making it necessary to save the websites offline for assessment later. As outlined in section 3.4.2 in some instances the live version of the

websites had to be used at the time of assessment. The live version of the website could have been different to the version present when the Google search was done initially, particularly if a considerable amount of time had passed since the initial search. This in turn means that the study was not strictly cross-sectional. However, this happened on very few occasions (less than 5) and is unlikely to impact the overall study findings.

To overcome this problem with hindsight it would have been advisable to check the offline version of the websites as soon as they are saved to see if the contents loaded successfully. If not, those few websites could have been examined at the time of the original search and the other saved websites examined later. Alternatively another method for saving websites offline could have been sought.

Linked pages

In this study only the main page of each website from the Google search was saved offline and linked pages had to be assessed using the live version. This could have potentially impacted on the study results because the linked pages could have been updated since the initial search, therefore impacting the cross-sectional study design.

If this study was repeated, a method to save both the main page and all linked pages of each website is viewed as advisable.

6.2.6 Tools and the AM checklist

Investigator guidance notes

As outlined in section 3.4.5.2, brief guidance notes were added for subjective questions in the existing tools before the start of the study. These guidance notes were not present as part of the tools and the tools were not validated. This might potentially impact on the generalisability of the study results. However, these guidance notes were used consistently throughout the study and probably reflect the thought process of the examiner when these tools are used in other studies. The fact that these guidance notes had to be used highlights

the need for additional guidance notes in the future to increase the reliability of these tools if they are to be developed further.

Application to videos and pdf websites

In our study, any website which conveyed information about orthognathic treatment including videos and pdf leaflets were analysed. This is because all these sources of information are available for patient use on orthognathic treatment on the Internet. However some questions in the existing tools concerning, for example, the presence of a search engine or link functioning, were difficult to answer for such websites and little guidance from the tool authors was given about how to best score these questions, so a score of 0 was allocated as decided by the primary researcher.

These questions were not applicable but impacted on the total score for each website. These may have scored lower on various tools, indicating poorer quality, which might not necessarily have been fair. This could potentially have impacted on the study results, but this is unlikely because it happened on only 6 occasions. Also the relatively large sample size in this study means that possible outliers such as these are unlikely to have had a major impact on the principal study results.

Readability

As outlined in section 3.4.5.2, readability was not assessed as part of the HWAT 3.0 tool. It could be argued that because readability assessment was part of the original HWAT 3.0 tool, the results of this study could have been affected by omitting this part of the tool.

Supplemental questions in LIDA

As discussed in section 3.4.5.2, the supplemental questions in the LIDA tool were not used in this study. Of these supplemental questions, two referred to website contents which might have impacted on the overall LIDA scores of the websites used in this study. However, this is unlikely because of the number of other questions which were answered in LIDA.

Total DISCERN score

Some previous studies (mentioned in section 6) only used questions 1- 15 in the DISCERN tool (section 2.8.4.2) when calculating the total DISCERN score for a website and did not include the score for question 16 'Based on the answers to all the above questions, rate the overall quality of the publication as a source of information about treatment choices'. However, as far as we are aware the developers of DISCERN did not endorse this, so the total score of questions 1 to 16 was used to calculate the total DISCERN score for each website in our study. As question 16 is potentially subjective, it could have impacted on the results of our study. However, even when this question was omitted in the total score, the strong correlation between the AM checklist and DISCERN still remained ($r= 0.8$, $p<0.01$) so it was unlikely to have had an impact on the overall study findings.

The AM checklist

The AM checklist was developed by the author of the study so this might have caused examiner bias. This may have led to more careful use of this checklist compared to the other tools which could potentially affect the results.

Another limitation of the AM checklist was that all the checklist items were scored 1 or 0. It could be argued that certain domains were more important than others. For example, risks are more important than epidemiological information, but both were scored 1 or 0. This could have potentially had a negative impact on the results. This is because if one website discussed epidemiological information but not risks and another discussed risks but not epidemiological information, they both would have scored 1 from these two questions giving the impression that they have the same quality. However, this is unlikely to have had an impact on the overall findings of the study because multiple items were included on the AM checklist so the checklist is likely to have reflected the overall quality of each website well.

Finally the risks most commonly discussed at orthognathic joint clinic were included in the AM checklist. However, not all of the risks of orthognathic surgery were included in the AM

checklist .These included chin and tongue paraesthesia, TMJ problems, relapse and re-operation if required. Ideally a website should list as many risks of orthognathic surgery as possible. However, the author proposed a more brief list of the main risks deemed acceptable to discuss on a website and this was approved by Orthodontic and Oral and Maxillofacial Surgeon consultants during the development of the AM checklist, Future studies could include risks which patients deem as most important on a checklist such as the AM checklist.

6.2.7 Patient input

The AM checklist was developed from the input of several clinicians in order to assess whether orthognathic websites included information which they considered essential for patients to know. However, the main users of orthognathic websites are patients. A limitation of our study was that patients had no input into the information they considered essential. The information gathered from the quality assessment of blogs in our study emphasised factors such as the liquid diet that patients might consider very important, which was not included in the AM checklist, so a website that scored high on the AM checklist might not necessarily be regarded as high quality by a patient.

6.3 Implications for practice

The findings from our study have important implications not just for authors of future studies but also for patients. It is likely that there will continue to be a high number of studies which will assess the quality of websites using existing tools, particularly LIDA and DISCERN. The authors of such studies need to use these HWQATs with caution. Before the start of the study authors should objectively assess the tool to be used using a method such as the AM checklist in order to check whether the tool is able to identify the best quality websites for the health topic they are going to investigate.

The findings from this study can also be used in educating orthognathic patients in how to find high quality orthognathic treatment websites. Patients need to be aware that they should

not simply rely on Google ranking to find high quality orthognathic treatment websites. They should also not use website suffix as a marker of quality. Instead they should look beyond the first few pages of the Google results and use UK websites. Patients should also be encouraged to use Blogs as an additional source of information in their preparation of orthognathic treatment.

6.4 Implications for future research

Further objective assessments of HWQATs

In our study to ascertain whether three existing tools correctly identify high quality websites, we developed and used a topic- specific checklist for orthognathic treatment websites. This showed that DISCERN was successful at identifying good orthognathic websites although it had moderate reliability. However, further studies are necessary to ascertain whether this is also the case when the results of DISCERN are compared to topic- specific checklists developed for other health topics.

Another aspect of HWQATs which should be investigated further is the use of categorisation. As mentioned in section 3.4.5.2, website categorisation was not used in this study. Instead, the actual numerical score produced by application of each of the tools to websites was used and analysed. Future studies could investigate the value of categorisation by comparing the categories in which an existing tool places websites against categories produced from a standard, such as the AM checklist.

Other studies to objectively assess the HWAT 3.0 tool further could include the HWAT 3.0 readability assessment to ascertain its ability to distinguish high quality health websites from those of poorer quality. To do this, the standard which HWAT 3.0 is checked against must also include a readability assessment to allow for a fair comparison.

Further development of the AM checklist

To reduce the risk of examiner bias, multiple examiners could use available tools and the AM checklist in future studies. Also in future studies the AM checklist questions could be weighted to place more emphasis on essential information such as risks.

Further studies of search engine ranking as a marker of quality

In this study a single search term was carried out for finding orthognathic treatment websites using one search engine using the UK setting. Future studies should analyse websites produced from different search terms using all available search engines to ascertain which is the best search engine at finding health websites with the best content quality. The results should be compared to topic- specific standards such as the AM checklist. Different country settings should also be used in the search engines and websites from different languages other than English analysed. These measures would mean a wider range of orthognathic websites are assessed.

The method utilised by Google may work well with generic searches, but further attention also needs to be paid to the content of medical websites and future studies could investigate how search engines can implement a more specific content quality assessment method as part of their search for medically related websites.

Studies with patient input

Further research, perhaps through a qualitative design, should propose various quality criteria to clinicians and patients to ascertain which they find most important. This is important because, although tools developed by clinicians might focus on general parameters or website content, patients may consider other features of the website more important. It is expected that such criteria will differ between patients and clinicians. Therefore to make tools more valid weighting of questions can be adopted from the HWAT 3.0 tool to differentiate between factors of higher and lesser importance to patients and clinicians hence better reflect website quality as they have considered all factors important to patients and clinicians and placed different emphasis on certain questions.

Studies of tool reliability

As discussed in section 6.1.4, reliability of tools was tested differently in previous studies and this made comparisons difficult. Future studies could include both question-based reliability tests as well as testing the reliability between the total scores from each website which was how reliability of tools was tested in this study. This would help to identify questions which produce the largest disagreement of scores between examiners and these questions could be further developed to help ultimately improve the overall reliability of tools. Furthermore, in order to improve DISCERN's reliability, the scoring system should be changed, perhaps to include less than five options on a Likert scale. This is particularly important for close- ended questions. Thought also must be given to how questions such as risks and benefits can be accurately scored where there may be several risks and benefits for each treatment. Also, guidance notes for each question could help to improve the reliability of this tool. Ultimately, these suggested changes must be made and the tool retested until the reliability is improved.

Studies of Blogs

The findings of this study showed that Blogs contained some extremely informative information which patients may find very useful. Further studies should explore the role of blogs as a patient information resource. Further studies could also assess patient interaction with Blogs compared with other sources of information such as patient information leaflets in their preparation for orthognathic treatment.

7. Conclusion

In this study, although a strong correlation was found between the DISCERN tool and AM checklist, DISCERN had only moderate inter-examiner and intra-examiner reliability. Therefore generic health website quality assessment tools in their current state should not be generalised and applied to health websites. Further development of such tools is required to increase their reliability and their ability to identify high content quality website.

UK websites can be used as a marker of higher quality orthognathic websites but Google ranking and the suffix of websites are not effective quality indicators.

Blogs have quality comparable to hospital and professional organisation websites and should be further investigated to be considered as a source of useful information for patients in the future.

8. References

ADEMILUYI, G., REES, C. and SHEARD, C., 2003. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. *Patient education and counseling,* **50**(2), pp. 151-155.

ALDAIRY, T., LAVERICK, S. and MCINTYRE, G., 2012. Orthognathic surgery: is patient information on the Internet valid? *The European Journal of Orthodontics,* **34**(4), pp. 466-469.

AMBRE, J., GUARD, R., PERVEILER, F., RENNER, J. and RIPPEN, H., 1997-last update, Criteria for assessing the quality of health information on the Internet. Available:http://www.aeemt.com/contenidos_socios/Informatica/Guias_y_recomendaciones/Criteria_Quality_Health_Inform_19971014.pdf.

ANDREASSEN, H., BUJNOWSKA-FEDAK, M., CHRONAKI, C., DUMITRU, R., PUDULE, I., SANTANA, S., VOSS, H. and WYNN, R., 2007. European citizens' use of E-health services: A study of seven countries. *BMC Public Health,* **7**, pp. 53-60.

ANSANI, N., VOGT, M., HENDERSON, B., MCKAVENEY, T., WEBER, R., SMITH, R., BURDA, M., KWOH, C., OSIAL, T. and STARZ, T., 2005. Quality of arthritis information on the Internet. *American Journal of Health-System Pharmacy,* **62**(11), pp. 1184-1189.

BERLAND, G., ELLIOTT, M., MORALES, L., ALGAZY, J., BRODER, M., KANOUSE, D., PUYOL, J., LARA, M., WATKINS, K., YANG, H., MCGLYNN, E., MUÑOZ, J. and KRAVITZ, R., 2001. Health information on the Internet: Accessibility, quality, and readability in English and Spanish. *Journal of the American Medical Association,* **285**(20), pp. 2612-2621.

BEST, J., MUZAFFAR, J. and MITCHELL-INNES, A., 2015. Quality of information available via the Internet for patients with head and neck cancer: are we improving? *European Archives of Oto-Rhino-Laryngology,* **272**(11), pp. 3499-3505.

BOHACEK, L., GOMEZ, M. and FISH, J., 2003. An evaluation of Internet sites for burn scar management. *Journal of Burn Care and Rehabilitation,* **24**(4), pp. 246-251.

BRECKONS, M., JONES, R., MORRIS, J. and RICHARDSON, J., 2008. What do evaluation instruments tell us about the quality of Complementary Medicine information on the Internet? *Journal of Medical Internet Research,* **10**(1), pp. 4-10.

BRIN, S. and PAGE, L., 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks (Elsevier),* **56**(18), pp. 3825-3833.

CARDELLE, A. and RODRIGUEZ, E., 2005. The Quality of Spanish health information websites : An emerging disparity. *Journal of Prevention & Intervention in the Community,* **29**(1), pp. 85-102.

CARDEN, C., JEFFORD, M. and ROSENTHAL, M., 2007. Information about cancer clinical trials: An analysis of Internet resources. *European Journal of Cancer,* **43**(10), pp. 1574-1580.

CERMINARA, C., SANTARONE, M., CASARELLI, L., CURATOLO, P. and EL MALHANY, N., 2014. Use of the DISCERN tool for evaluating web searches in childhood epilepsy. *Epilepsy and Behavior,***41**(1), pp. 119-121.

CHANG, M., HAN, D., MOON, I., KIM, S., KIM, D., LEE, C., MIN, Y. and RHEE, C., 2010. Assessment of allergic rhinitis websites in Korea. *Clinical and Experimental Otorhinolaryngology,* **3**(1), pp. 32-36.

CHARNOCK, D. and SHEPPERD, S., 2004. Learning to DISCERN online: applying an appraisal tool to health websites in a workshop setting. *Health education research,* **19**(4), pp. 440-446.

CHARNOCK, D., SHEPPERD, S., NEEDHAM, G. and GANN, R., 1999. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology & Community Health,* **53**(2), pp. 105-111.

CHESTNUTT, I. and REYNOLDS, K., 2006. Perceptions of how the Internet has impacted on dentistry. *British Dental Journal,* **200**(3), pp. 161-165.

CLINE, R. and HAYNES, K., 2001. Consumer health information seeking on the Internet: The state of the art. *Health education research,* **16**(6), pp. 671-692.

COLELLA, G., CANNAVALE, R., VICIDOMINI, A. and LANZA, A., 2007. Neurosensory disturbance of the inferior alveolar nerve after bilateral sagittal split osteotomy: A systematic review. *Journal of Oral and Maxillofacial Surgery,* **65**(9), pp. 1707-1715.

COULTER, A., 1998. Evidence based patient information. *British Medical Journal,* **317**(7153), pp. 225-226.

COULTER, A., 1997. Partnerships with patients: the pros and cons of shared clinical decision making. *Journal of Health Services Research and Plicy,* **2**(2), pp. 112-121.

CUNNINGHAM, S., HUNT, N. and FEINMANN, C., 1996. Perceptions of outcome following orthognathic surgery. *British Journal of Oral and Maxillofacial Surgery,* **34**(3), pp. 210-213.

DASH, N., PIPAL, K., KAZA, S., PANIGRAHI, P. and JOSHI, A., 2012. Assessment of the malnutrition related information on the Internet. *Technology and Health Care,* **20**(2), pp. 117-125.

DELAMOTHE, T., 2000. Quality of websites: kitemarking the west wind. *BMJ,* **321**(7265), pp. 843-844.

DIAZ, J., GRIFFITH, R., NG, J., REINERT, S., FRIEDMANN, P. and MOULTON, A., 2002. Patients' use of the Internet for medical information. *Journal of General Internal Medicine,* **17**(3), pp. 180-185.

DIAZ, J., SCIAMANNA, C., EVANGELOU, E., STAMP, M. and FERGUSON, T., 2005. Brief report: What types of Internet guidance do patients want from their physicians? *Journal of General Internal Medicine,* **20**(8), pp. 683-685.

DOUPI, P. and VAN DER LEI, J., 1999. Rx medication information for the public and the WWW: quality issues. *Informatics for Health and Social Care,* **24**(3), pp. 171-179.

DOWNING, M., OMAR, A., SABRI, E. and MCCARTHY, A., 2011. Information on the Internet for asplenic patients: a systematic review. *Canadian journal of surgery,* **54**(4), pp. 232-236.

EDEJER, T., 2000. Disseminating health information in developing countries: the role of the Internet. *British Medical Journal,* **321**(7264), pp. 797-800.

EKMAN, A., HALL, P. and LITTON, J., 2005. Can we trust cancer information on the Internet? A comparison of interactive cancer risk sites. *Cancer Causes and Control,* **16**(6), pp. 765-772.

EYSENBACH, G. and 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), pp. 573-577.

EYSENBACH, G., POWELL, J., KUSS, O. and SA, E., 2002. Empirical studies assessing the quality of health information for consumers on the world wide web. A systematic review. *Journal of American Medical Association,* **287**(20), pp. 2691-2700.

FAHY, E., HARDIKAR, R., FOX, A. and MACKAY, S., 2014. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. *Australasian Medical Journal,* **7**(1), pp. 24-28.

FALLOWFIELD, L., HALL, A., MAGUIRE, G. and BAUM, M., 1990. Psychological outcomes of different treatment policies in women with early breast cancer outside a clinical trial. *BMJ: British Medical Journal / British Medical Association,* **301**(6752), pp. 575-580.

FARRELL, K., ROBINSON, L., BAYDOCK, S., FARELL, S., IRVING, L. and O'CONNELL, C., 2006. A survey of Canadian websites providing information about female urinary incontinence. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstétrique et gynécologie du Canada : JOGC,* **28**(8), pp. 700-712.

FOX, S. and DUGGAN, M., 2013. *Health Online 2013.* http://pewInternet.org/Reports/2013/Health-online.aspx.

GAGLIARDI, A. and JADAD, A., 2002. Examination of instruments used to rate quality of health information on the Internet: chronicle of a voyage with an unclear destination. *BMJ,* **324**(7337), pp. 569-573.

GASTON, C. and MITCHELL, G., 2005. Information giving and decision-making in patients with advanced cancer: A systematic review. *Social science & medicine,* **61**(10), pp. 2252-2264.

GOLDSMITH, J., 2000. How will the Internet change our health system? *Health affairs,* **19**(1), pp. 148-156.

GREENFIELD, S., KAPLAN, S. and WARE, J., 1985. Expanding patient involvement in care: Effects on patient outcomes. *Annals of Internal Medicine,* **102**(4), pp. 520-528.

HAINER, M., TSAI, N., KOMURA, S. and CHIU, C., 2000. Fatal hepatorenal failure associated with Hydrazine Sulfate. *Annals of Internal Medicine,* **133**(11), pp. 877-880.

HSU, W. and BATH, P.A., 2008. Development of a patient-oriented tool for evaluating the quality of breast cancer information on the Internet. *Studies in health technology and informatics,* **136**, pp. 297-302.

HUANG, J., DISCEPOLA, F., AL-FOZAN, H. and TULANDI, T., 2005. Quality of fertility clinic websites. *Fertility and sterility,* **83**(3), pp. 538- 544.

HUNT, O., JOHNSTON, C., HEPPER, P. and BURDEN, D., 2001. The psychosocial impact of orthognathic surgery: A systematic review. *American journal of orthodontics and dentofacial orthopedics,***120**(5), pp. 490-497.

IRWIN, J., THYVALIKAKATH, T., SPALLEK, H., WALI, T., KERR, A. and SCHLEYER, T., 2011. English and Spanish oral cancer information on the Internet: a pilot surface quality and content evaluation of oral cancer web sites. *Journal of Public Health Dentistry,* **71**(2), pp. 106-116.

JADAD, A. and GAGLIARDI, A., 1998. Rating health information on the Internet- Navigating to knowledge or to babel? *JAMA,* **279**(8), pp. 611-614.

JAIN, T. and BARBIERI, R., 2005. Website quality assessment: Mistaking apples for oranges. *Fertility and sterility,* **83**(3), pp. 545-547.

JENKINS, K. and BAKER, A., 2003. Consent and anaesthetic risk. *Anaesthesia,* **58**(10), pp. 962-984.

JOSHI, M., BHANGOO, R. and KUMAR, K., 2011. Quality of nutrition related information on the Internet for osteoporosis patients: A critical review. *Technology and Health Care : Official Journal of the European Society for Engineering and Medicine,* **19**(6), pp. 391-400.

KAPLAN, S., 1989. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. *Medical care,* **27**(3), pp. S110-S127.

KEE, F., 1996. Patients' prerogatives and perceptions of benefit. *British Medical Journal,* **312**(7036), pp. 958-960.

KHAZAAL, Y., CHATTON, A., ZULLINO, D. and KHAN, R., 2012. HON Label and DISCERN as content quality indicators of health-related websites. *Psychiatric quarterly,* **83**(1), pp. 15-27.

KIRTHI, V. and MODI, B., 2012. Coronary angioplasty and the Internet: What can patients searching online expect to find? *Journal of interventional cardiology,* **25**(5), pp. 476-481.

KRUSE, R., KOOPMAN, R., CANFIELD, S., MEHR, D., WAKEFIELD, B., KEPLINGER, L. and WAKEFIELD, D., 2012. Internet use by primary care patients: Where is the digital divide? *Family medicine,***44**(5), pp. 342-347.

KUMAR, V., SUBRAMANI, S., VEERAPAN, S. and KHAN, S., 2014. Evaluation of online health information on clubfoot using the DISCERN tool. *Journal of Pediatric Orthopaedics: Part B,* **23**(2), pp. 135-138.

LEE, P., 1999. Why literacy matters- Links between reading ability and health. *Archives of Ophthalmology,* **117**(1), pp. 100-103.

LEIRA-FEIJOO, Y., 2014 (in press). *Available web-based dental implants information for patients. How good is it?* .

LEWIECKI, E., RUDOLPH, L., KIEBZAK, G., CHAVEZ, J. and THORPE, B., 2006. Assessment of osteoporosis-website quality. *Osteoporosis International,* **17**(5), pp. 741-752.

LIVAS, C., DELLI, K. and REN, Y., 2013. Quality evaluation of the available Internet information regarding pain during orthodontic treatment. *Angle Orthodontist,* **83**(3), pp. 500-506.

LÓPEZ-JORNET, P. and CAMACHO-ALONSO, F., 2010. The quality of patient-orientated Internet information on oral lichen planus: a pilot study. *Journal of evaluation in clinical practice; Journal of evaluation in clinical practice,* **16**(5), pp. 883-886.

LUTHER, F., MORRIS, D.O. and HART, C., 2003. Orthodontic preparation for orthognathic surgery: how long does it take and why? A retrospective study. *British Journal of Oral and Maxillofacial Surgery,* **41**(6), pp. 401-406.

MARSHALL, L. and WILLIAMS, D., 2006. Health information: does quality count for the consumer?: How consumers evaluate the quality of health information materials across a variety of media.*Journal of Librarianship and Information Science,* **38**(3), pp. 141-156.

MCLELLAN, F., 1998. `Like hunger, like thirst': Patients, journals, and the Internet. *The Lancet (North American Edition),* **352**(9134), pp. SII39- SII43.

MCLEOD, S., 1998. The quality of medical information on the Internet. a new public health concern. *Archives of Ophthalmology,* **116**(12), pp. 1663-1665.

MCMULLAN, M., 2006. Patients using the Internet to obtain health information: How this affects the patient–health professional relationship. *Patient education and counseling,* **63**(1), pp. 24-28.

MEREDITH, P., EMBERTON, M. and WOOD, C., 1995. New directions in information for patients. *British medical journal,* **311**(6996), pp. 4-5.

MURRAY, E., LO, B., POLLACK, L., DONELAN, K., CATANIA, J., LEE, K., ZAPERT, K. and TURNER, R., 2003. The Impact of Health Information on the Internet on Health Care and the Physician-Patient Relationship: National U.S. Survey among 1.050 U.S. Physicians. *Journal of Medical Internet Research,* **5**(3), pp. 38- 53.

NÄRHI, U., POHJANOKSA-MÄNTYLÄ, M., KARJALAINEN, A., SAARI, J., WAHLROOS, H., AIRAKSINEN, M. and BELL, S., 2008. The DARTS tool for assessing online medicines information. *Pharmacy World & Science,* **30**(6), pp. 898-906.

NASSER, S., MULLAN, J. and BAJOREK, B., 2012. Assessing the quality, suitability and readability of Internet-based health information about warfarin for patients. *Australasian Medical Journal,***5**(3), pp. 194-203.

NEUMARK, Y., FLUM, L., LOPEZ-QUINTERO, C. and SHTARKSHALL, R., 2012. Quality of online health information about oral contraceptives from Hebrew-language websites. *Israel Journal of Health Policy Research,* **1**(38), pp. 1-10.

NÍ RÍORDÁIN, R. and MCCREARY, C., 2009. Dental patients' use of the Internet. *British dental journal,* **207**(12), pp. 583-586.

NWOSU, C. and COX, B., 2000. The impact of the Internet on the doctor-patient relationship. *Health informatics journal,* **6**(3), pp. 156-161.

OFFICE FOR NATIONAL STATISTICS, 2015-last update, Internet Access – Households and Individuals 2015. Available: http://www.ons.gov.uk/ons/dcp171778_412758.pdf.

OFFICE FOR NATIONAL STATISTICS, 2009-last update, Internet Access - Households and Individuals, 2009. Available: http://www.ons.gov.uk/ons/rel/rdit2/Internet-access---households-and-individuals/2009/index.html.

OGUNWALE, B., CLARKE, J., YOUNG, D., MOHAMMED, A., PATIL, S. and MEEK, R., 2009. Direct to Consumer Advertising via the Internet, a Study of Hip Resurfacing. *Scottish Medical Journal,***54**(1), pp. 10-13.

OXMAN, A., GUYATT, G., COOK, D., JAESCHKE, R., HEDDLE, N. and KELLER, J., 1993. An index of scientific quality for health reports in the lay press. *Journal of clinical epidemiology,* **46**(9), pp. 987-1001.

PAREKH, J. and GILL, D., 2014. The quality of orthodontic practice websites. *British Dental Journal,* **216**(10), pp. E21.

PARIKH, A.R., KOK, K., REDFERN, B., CLARKE, A., WITHEY, S. and BUTLER, P.E.M., 2006. A portal to validated websites on cosmetic surgery: The design of an archetype. *Annals of Plastic Surgery,***57**(3), pp. 350-352.

PARK, M., JO, J. and PARK, J., 2012. Quality and content of Internet-based information on temporomandibular disorders. *Journal of Orofacial Pain,* **26**(4), pp. 296-306.

PATEL, U. and COBOURNE, M., 2011. Orthodontic extractions and the Internet: Quality of online information available to the public. *American Journal of Orthodontics and Dentofacial Orthopedics,***139**(2), pp. e103- e109.

PÉREZ-LÓPEZ, F., 2004. An evaluation of the contents and quality of menopause information on the World Wide Web. *Maturitas,* **49**(4), pp. 276-282.

POHJANOKSA-MÄNTYLÄ, M., SAARI, J., NÄRHI, U., KARJALAINEN, A., PYLKKÄNEN, K., AIRAKSINEN, M. and BELL, J., 2009. How and why do people with depression access and utilize online drug information: A qualitative study. *Journal of Affective Disorders,* **114**(1), pp. 333-339.

POWELL, J. and CLARKE, A., 2002. The WWW of the World Wide Web: Who, What, and Why? *Journal of Medical Internet Research,* **4**(1), pp. e4.

PROFFIT, W. and MIGUEL, J., 1995. The duration and sequencing of surgical orthodontic treatment. *The international Journal of Adult Orthodontics and Orthognathic Surgery,* **10**, pp. 35- 42.

PRUSTI, M., LEHTINEVA, S., POHJANOKSA-MÄNTYLÄ, M. and BELL, J., 2012. The quality of online antidepressant drug information: An evaluation of English and Finnish language Web sites.*Research in Social and Administrative Pharmacy,* **8**(3), pp. 263-268.

RAO, N., MOHAPATRA, M., MISHRA, S. and JOSHI, A., 2012. Evaluation of Dengue-Related Health Information on the Internet. *Perspectives in Health Information Management / AHIMA, American Health Information Management Association,* **5**(1), pp. 1-8.

RUSS, H., GIVEON, S., CATARIVAS, M. and YAPHE, J., 2011. The Effect of the Internet on the Patient-Doctor Relationship from the Patient's Perspective: A Survey from Primary Care. *The Israel Medical Association journal: IMAJ,* **13**(4), pp. 220-224.

SANDVIK, H., 1999. Health information and interaction on the Internet: a survey of female urinary incontinence. *British Medical Journal,* **319**(7201), pp. 29-32.

SARVER, D. and JOHNSTON, M., 1993. Orthognathic surgery and aesthetics: planning treatment to achieve functional and aesthetic goals. *British Journal of Orthodontics,* **20**(2), pp. 93-100.

SATTERLUND, M., MCCAUL, K. and SANDGREN, A., 2003. Information Gathering Over Time by Breast Cancer Patients. *Journal of Medical Internet Research,* **5**(3), pp. e15.

SCHMIDT, K. and ERNST, E., 2004. Assessing websites on complementary and alternative medicine for cancer. *Annals of Oncology,* **15**(5), pp. 733- 742.

SILBERG, W., LUNDBERG, G. and MUSACCHIO, R., 1997. Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet Caveant Lector et Viewor—Let the Reader and Viewer Beware. *Generations,* **227**(15), pp. 1244-1245.

SOUSA, C. and TURRINI, R.N., 2012. Complications in orthognathic surgery: A comprehensive review. *Journal of Oral and Maxillofacial Surgery Medicine and Pathology,* **24**(2), pp. 67- 74.

SURMAN, R. and BATH, P., 2013. An assessment of the quality of information on stroke and speech and language difficulty web sites. *Journal of Information Science,* **39**(1), pp. 113-125.

THOMPSON, A. and GRAYDON, S., 2009. Patient-oriented methotrexate information sites on the Internet: a review of completeness, accuracy, format, reliability, credibility, and readability. *Journal of Rheumatology,* **36**(1), pp. 41-49.

TONES, K., 2002. Health literacy: new wine in old bottles? *Health education research,* **17**(3), pp. 287-290.

URAC., 2005-last update, URAC Health web site check-up service and accreditation program.
Available: http://www.urac.org/prog_accred_HWS_po.asp?navid=accreditation&pagename=prog_accred_HWS.

V. CRUZ, A. and SANTOS, A., 2006. Blindness after Le Fort I osteotomy: A possible complication associated with pterygomaxillary separation. *Journal of Cranio-Maxillofacial Surgery,* **34**(4), pp. 210 -216.

WALJI, M., SAGARAM, S., SAGARAM, D., MERIC-BERNSTAM, F., JOHNSON, C., MIRZA, N. and BERNSTAM, E., 2004. Efficacy of quality criteria to identify potentially harmful information: a cross-sectional survey of complementary and alternative medicine web sites. *Journal of Medical Internet Research,* **6**(2), pp. e21.

WILES, R., PAIN, H., BUCKLAND, S. and MCLELLAN, L., 1998. Providing appropriate information to patients and carers following a stroke. *Journal of Advanced Nursing,* **28**(4), pp. 794-801.

WILSON, P., 2002. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the Internet. *British Medical Journal,* **324**(7337), pp. 598- 600.

YAP, K., RAAJ, S. and CHAN, A., 2010. OncoRx-IQ: a tool for quality assessment of online anticancer drug interactions. *International Journal for Quality in Health Care,* **22**(2), pp. 93-106.

ZAHEDI, R., TAHERI, B., SHAHRZADI, L., TAZHIBI, M. and ASHRAI-RIZI, H., 2013. Quality of Persian Addiction Websites: A Survey Based on Silberg, DISCERN and WQET Instruments 2011. *Acta Informatica Medica,* **21**(1), pp. 46-50.

9. Appendices

Appendix 1: LIDA Tool (Shortened version)

Accessibility

HTTP-Equiv Content-Type (in header)
HTML Language Definition
Page Title
Meta Tag Keywords
Document type definition
Image Alt Tags
Specified Image Widths
Table Summaries
Frames
Body Tags - Body Background Colour
Body Tags - Body Topmargin
Body Tags - Body Margin Height
Table Tags - Table Background Colour
Table Tags - Table Column (td) Height
Table Tags - Table Row (tr) Height
Font Tags - Font Color
Font Tags - Font Size
Align (non style sheet)

Usability

Is the site design clear and transparent?
Is the site design consistent from one page to another?
Can users find what they need on the site?
Is the format of information clear and appropriate for the audience?

Reliability

Is it clear who has developed the web site and what their objectives are?
Does the site report a robust quality control procedure?
Is the page content checked by an expert?
Is the page updated regularly?
Does the page cite relevant sources where appropriate?

Appendix 2: LIDA Tool (Full version)

The LIDA Instrument

Minervation validation instrument for health care web sites

Full Version (1.2) containing instructions

Most internet users look for health information online,[1-3] but finding unreliable information can lead to harm.[4] There is no shortage of health information out there. The problem most people have is finding **good quality** information that's **relevant** to them.

These are the challenges to information providers:

1. How can you make sure that the information you are providing is accessible, relevant and high quality?
2. Could your site fall foul of legislation affecting visually impaired Internet users?
3. Does your site's poor usability waste your audience's time by making it hard for them to find what they need?
4. How can you be sure that the information you publish is up-to-date, accurate and reliable?

These difficulties have prompted Minervation to develop a set of free validation tools to help web site developers answer these questions.

How to use this document:

For Level 1 (Accessibility) go to www.minervation.com/validation and type in the URL of the site you wish to assess. This will generate answers to questions 1.1 to 1.4.

For questions 1.5, 1.6, Level 2 (Usability) and Level 3 (Reliability), view the web site as normal and enter your scores in the boxes provided. Score each question on a scale of zero to 3, where:

0 = Never
1 = Sometimes
2 = Mostly
3 = Always

The Minerva*lida*tion tool evaluates the design and content of health web sites.

The tool measures three areas:

1. **Accessibility**
   a. Can your audience access your web site?
   b. Does your site conform to legal accessibility standards?
   c. Are your competitors ahead of you?
   d. Does your site reflect "best practice" in coding and relevant metadata?

2. **Usability**
   a. Can your users find what they need to know?
   b. Can they use your web site effectively?
   c. What does it cost people to use your web site?
   d. Do your site visitors return to use the site again and again?

3. **Reliability**
   a. Does your site keep up to date with the latest research?
   b. Does your site reflect best current knowledge?
   c. Do your users trust you to provide them with unbiased information?
   d. Does your site conform to the highest information quality standards throughout?
   e. Is your site harmful or dangerous?

Why does validation matter?
These three areas are important for a number of reasons: some legal, some political, some financial:

Level 1 Accessibility
- Making sure that web sites are accessible to *all* is now law.[5-7]
- By conforming to accessibility standards, NHS and not-for-profit sites producing health information will be permitted to join the NHS Information Partners Programme[8], and will therefore be searchable via NHS Direct Online[9], leading to increased traffic.
- Research information which is available full-text online has a higher impact than information which has restricted access.[10]

Level 2 Usability
- If people cannot use your web site effectively, they'll go elsewhere.[11-13]
- Your web site may be costing your users time which they cannot afford.[14-16]
- Most health web sites present information in a way that is hard for users to understand.[17-19]
- If your site suffers from poor usability, your users may not come back.[13;20;20-22]

Level 3 Reliability
- Users will not trust your web site if it does not have a clear quality control policy.[23]
- Web health information often contains inaccuracies[22] and is usually incomplete.[24;25]
- In some cases web sites have actually been proven to be harmful or dangerous.[26-30] Can you be sure that your site is safe?
- Even "evidence-based" guidelines have been shown to be subject to bias.[31;32]

Aren't there other evaluation tools we can use?
- Yes, there are hundreds, but almost none have been tested for their reliability.[33]
- Those that have been tested are mostly unreliable.[34]
- The few that are reliable do not adequately address the issues of accessibility and usability.[35]
- Information which is validated according to well-known quality schemes still tends to be unusable.[36]

Level 1 Accessibility

- *Does the web site meet W3C and Bobby standards?*
- *Can users access the information in the web site?*
- *Is the web site "future proof"?*

1.1 Page Setup
Characteristics which identify a web page so that web browsers can interpret it correctly.

1.1.1 Document Type Definition

1.1.2 HTTP-Equiv Content-Type (in header)

1.1.3 HTML Language Definition

1.1.4 Page Title

1.1.5 Meta Tag Keywords

1.2 Access Restrictions
These factors can restrict users' access to the site, especially those with disabilities.

1.2.1 Image Alt Tags

1.2.2 Specified Image Widths

1.2.3 Table Summaries

1.2.4 Frames
Web sites must not use frames because they confuse disabled users' screen readers and cause usability problems for other users.[37]

1.3 Outdated Code
HTML elements which will not be used in future versions; should be done using style sheets to eliminate inefficient and inconsistent design practices.

1.3.1 Body Tags

1.3.2 Table Tags

1.3.3 Font Tags

1.3.4 Alignment

1.4 Dublin Core Title Tags
Metadata which will ensure compatibility with NHS directives.

1.5 Browser Test
The web site should work in all commonly used browsers and on Macintosh

For a review of current web browsers, see:
http://www.minervation.com/news_archive.asp?t=10&nid=160&d=200503

1.6 Registration
Is the information available full text without registration, login or subscription?[10]

3 = No login or registration essential for certain features (e.g. eCommunity)
1 = Free registration
0 = Paid registration

Level 2 Usability

- *Can users find the information they need?*
- *Poor usability increases costs (for both you and your users)*
- *Good usability increases usage, stickability and revenues.*

2.1 Clarity

Total:

Clear design increases usability by promoting accessibility, signposting content and encouraging exploration.[37-40;40;41]

2.1.1 Is there a clear statement of who this web site is for?

- *Did it take you long to find this information (No=2, Yes=1, Couldn't=0)?*
- *Is this information on the home page (Yes = 3)?*

2.1.2 Is the level of detail appropriate to their level of knowledge?

When assessing this question, try to think of a typical user from the group specified in 2.1.1.

- *Does the site lead the user into the right level of detail in the right sequence?*
- *Is there a lot of jargon that they would not understand?*
- *Is the language of the right complexity?*
- *Does the site make good use of graphics to explain complex information?*

2.1.3 Is the layout of the main block of information clear and readable?

Look at the "block of content"

- *Is the font size appropriate?*
- *Scannability: use of subheadings?*
- *Use of bulleted lists and internal links within a long document (good)*
- *Text wrapping*
- *Length of the page (long = bad; may need "go back to the top" links)*

2.1.4 Is the navigation clear and well structured?

Look at the buttons, links and menus

- *Can you tell what is a link or button?*
- *Are they readable?*
- *Is it clear which menu you need to click to find what you need (e.g. mixing up subtopics with publication types would make this hard)?*

2.1.5 Can you always tell your current location in the site?

- *There may be breadcrumbs or changes in the menu system telling you which section you're in, though they can be confusing.*

2.1.6 Is the colour scheme appropriate and engaging?

- *Is it appropriate for the target audience?*
- *Is it tasteful?*
- *Is it readable?*
- *Print out a black and white page to see if there's enough contrast for colour blind people.*
- *Remember to check the colours of mouse-overs and previously-clicked links etc.*

Additional Comments on Clarity:

2.2 Consistency

Total: ☐

Consistent design helps users to learn how a web site works and where to look for the information they need.[42;43]

2.2.1 Is the same page layout used throughout the site?

☐

Are the menus, text blocks, header, footer etc consistent throughout?

- *Sometimes it's a good thing to have a different layout, for example when moving from a text-heavy explanation page into a multiple choice question, or if it's a gateway site that links to other resources.*
- *Ask yourself, would this inconsistency be confusing to the user? Does it make sense to use a different layout for this page? Can the user still "retrace their steps" if they need to?*

2.2.2 Do navigational links have a consistent function?

☐

Think about what happens when you click the link, e.g.

- *Do external links always open in a new window?*
- *Does the home page or logo link always take you to the home page?*
- *Does the search or feedback button always work in the same way?*

Again, inconsistency may be appropriate depending on whether it would make sense to the user. If it doesn't make sense to you, it certainly won't make sense to everyday users.

2.2.3 Is the site structure (categories or organisation of pages) applied consistently?

☐

Think about whether the subsections used in different areas of the site are consistent.

- *If they are, users will find it easier to predict where to find what they need on the site.*
- *The site map should help to assess this question.*

Additional Comments on Consistency:

2.3 Functionality

Total: ☐

Web sites must provide users with the right tools to find what they need without overburdening them with unnecessary functions[40;44].

2.3.1 Does the site provide an effective search facility?

☐

Browse to a section and think of a typical term that might require that information and a synonymous term people might search for which isn't on that page (e.g. Fluoxetine and Prozac). Do a search for each.

- *Did you find the page in question?*
- *Does it work with synonyms?*
- *Is the ranking of results sensible?*
- *Does it display sufficient information on the hits for you to choose the right one?*
- *Can you refine your search?*
- *Is the complexity of the search engine appropriate for the site?*

2.3.2 Does the site provide effective browsing facilities?

☐

As above, find a page and think of a typical query that a user of this site might have which requires that page. Go to the site home page.

- *Can you find your page by browsing?*
- *Would it be obvious what to click on to get that page?*
- *How many clicks did it take (target ≤ 3)?*

2.3.3 Does the design minimise the cognitive overhead of using the site?

Cognitive overhead means "the additional effort and concentration necessary to maintain several tasks or trails at one time".[45] *So, it's a general term to describe whether a web site requires its users to learn, do, remember or read lot of unnecessary information before they get what they want.*

- *If you very quickly get accustomed to a site and how it works, it probably has a low (i.e. good) cognitive overhead.*
- *The sorts of things that increase cognitive overhead are: having to go to lots of different areas to get the information you need; not being able to tell where to go to get what you want; or not getting what you expected when you click on a link; unusual design or layout that is inconsistent with user expectations, especially in search engine and results pages*[46].

2.3.4 Does the site support the normal browser navigational tools?

A usable web site shouldn't change what you'd expect to be able to do with your web browser:

- *e.g. mouse-over a link to get the target, page address displayed in the address bar, title in the window title, browser toolbar buttons present and consistent (back, forward, home, etc)*

2.3.5 Can you use the site without third party plug-ins?

Typical scores:

- *No plug-ins or PDF equivalent of text that's available elsewhere on the site = 3*
- *Appropriate use of freely available plug-in (such as PDF) **and** it adds value = 2*
- *As above but it could have been done in another way without a plug-in = 1*
- *Gratuitous = 0*

Additional Comments on Functionality:

2.4 Engagability

Total:

Web sites which provide users with a satisfying experience are more effective and more popular[47;48].

2.4.1 Can the user make an effective judgment of whether the site applies to them?

- *Could they make this judgment within a few seconds of visiting the site?*
- *Can the user quickly find the subsection of the site that has been produced specifically for them?*

2.4.2 Is the web site interactive?

Newsletters, eCommunities, chat, enquiry and feedback forms, animations or illustrations:

- *Think about how the site compares with others in the same topic.*
- *For newsletters – look for the ability to specify topics of interest, rather than general updates.*
- *For eCommunities – look for active bulletin boards with lots of users.*
- *For feedback mechanisms – look for forms rather than simple email addresses; is it clear who you are sending feedback to?*

2.4.3 Can the user personalise their experience of using the site?

2.4.4 Does the web site integrate non-textual media?

This includes drawings, diagrams, graphs, photographs as well as audio, video and animation.

- *Do they look professional?*
- *Are they appropriate?*

Additional Comments on Engagability:

Level 3 Reliability

Does the site provide comprehensive, relevant and unbiased information? If not, it is unreliable and may be harmful.[4]
In a systematic review of studies of the quality of health information on the web, 70% found that quality is a problem[17].

3.1 Currency — Total: ☐

If a site is not updated regularly, new evidence may emerge which conflicts with it and which renders the site redundant[49].

3.1.1 Does the site respond to recent events? ☐

Look for coverage of recent events, news items, etc.

3.1.2 Can users submit comments on specific content? ☐

Look for 'in page' comments (these often appear towards the bottom of the page), rather than simple feedback functionality which does not affect the actual site content.

3.1.3 Is site content updated at an appropriate interval? ☐

Is the clinical content updated frequently enough to be up to date? Look for a statement in site policy, the date on each page.
- *Can't tell = 0;*
- *For treatment, an ideal target would be 6 monthly updates; for diagnosis and background information it can be longer.*

Additional Comments on Currency:

3.2 Conflicts of interest — Total: ☐

Surveys show that disclosure of sponsorship is a key issue for users of health web sites.[17]

3.2.1 Is it clear who runs the site? ☐

3.2.2 Is it clear who pays for the site? ☐

Current practice is not good in this area. If you have to look for organisational reports, etc, which disclose funding sources, then that scores 1.

3.2.3 Is there a declaration of the objectives of the people who run the site? ☐

Are these consistent with the objective of providing you with unbiased and accurate information?

Additional Comments on Conflicts of Interest:

3.3 Content production	Total:
Where information is not gathered using a rigorous methodology, the findings are likely to be biased[31;50-52].	☐
3.3.1 Does the site report a clear content production method? *Look for a statement that tells you how information on the site was produced and its quality checked. This might be in an About Us, About this Site or Editorial Policy section.*	☐
3.3.2 Is this a robust method? *Ideally, it should include:* ○ *User-driven identification of user needs and validation of site design* ○ *Comprehensive searching for relevant literature* ○ *Appraisal of the validity of sources using evidence-based guidelines* ○ *Review of the site content by independent experts* ○ *Review of the site by target audience*	☐
3.3.3 Can the information be checked from original sources? *Use your judgment to decide what statements require references. Background information may not need a reference, but clinical definitions of disease usually do; statements of the findings of research certainly do.*	☐

Additional Comments on Content Production:

Questions 3.4 and 3.5 are supplemental questions which require a detailed examination of the web site production process. This may not be possible from looking at the site; you may have to find out more by contacting the host organisation.

3.4 Content production procedure - supplemental	***Total:***
Where the purpose is providing high quality answers to users' questions about health care.	☐
3.4.1 Are the audience needs identified in advance? *Determining needs in advance leads to more robust answers[53]; involving users in this process leads to more effective[20], satisfying (by as much as 40%) and cheaper[16] web solutions.*	☐
3.4.2 Is comprehensive literature searching conducted? *This is necessary to make sure all the relevant documents are found[54], and language[55] and publication[56] biases are eliminated.*	☐
3.4.3 Are retrieved documents critically appraised? *Critical appraisal should be conducted independently using validated appraisal tools.[57]*	☐
3.4.4 Is content authored by subject experts?	☐
3.4.5 Is content reviewed by an independent expert or panel?	☐

Additional Comments on Content Production – Supplemental:

3.5 Output of content - supplemental	***Total:***
Does the site provide accurate and reliable information?	☐
3.5.1 Has literature searching found the right information? *Are there any important data sources missing from the search?*	☐
3.5.2 Does the content check out? *Is the content consistent with current best practice in the topic area?*	☐
3.5.3 Is the content accurate? *Here we're checking for editorial mistakes such as the classification of information (e.g. information about metastatic cancer located in a section header about non-metastatic cancer), use of incorrect references and spelling mistakes.*	☐

Additional Comments on Output of Content – Supplemental:

minervation

Summary Sheet

Calculate totals for each section and record them here

URL: _____

Site Owner: _____

1 Accessibility Total

Total (out of 63): ☐

Enter the totals from Level 1:

1.1-4. Automated test	☐	*out of 57*
1.5. Browser test	☐	*out of 3*
1.6 Full text availability	☐	*out of 3*

Key comments / priorities:

2 Usability Total

Total (out of 54): ☐

Enter the totals from Level 2:

2.1. Clarity	☐	*out of 18*
2.2. Consistency	☐	*out of 9*
2.3. Functionality	☐	*out of 15*
2.4. Engagability	☐	*out of 12*

Key comments / priorities:

3 Reliability Total

Total (out of 27): ☐

Supplementary Total (out of 24): ☐

Enter the totals from Level 3:

3.1. Currency	☐	*out of 9*
3.2. Conflicts of Interest	☐	*out of 9*
3.3. Content Production	☐	*out of 9*
3.4. Content Production - Supplemental	☐	*out of 15*
3.5. Output of Content - Supplemental	☐	*out of 9*

Key comments / priorities:

Appendix 3: The HWAT 3.0 tool

Content

1) Originating person/ organisation identification (weighting 9)
2) Basic information e.g. definition, consequence, prevention and treatment (weighting 6)
3) Clearly stating the subject of the website (weighting 9)
4) Education distinguished from advertising (weighting 9)

Credibility

1) Is the author or institution providing the information stated? (weighting 9)
2) Is the source of information credible? (weighting 9)
3) Is there a seal of approval present? (weighting 6)

Navigability

1) Is the margin cut off if page printed? (weighting 9)
2) Do the interwebsite links function? (weighting 9)
3) Do the intrawebsite links function? (weighting 9)
4) Is there the opportunity to email feedback? (weighting 6)

Currency

1) Is there a revision date or copyright date if it is within 12 months? (weighting 6)

Readability

1) Is the Flesch-Kincaid grade level at 8^{th} grade or below? This is the average reading level of the general population(Lee, 1999) (weighting 4)

Appendix 4: Data for intra- examiner and iner- examiner reliability

Table 12 LIDA, DISCERN, HWAT 3.0 and AM checklist (fist version) scores for intra-examiner and inter- examiner reliability

Website	Assessor	LIDA total	DISCERN total	HWAT 3.0 total	AM Checklist (first version) total
http://www.bos.org.uk/Resources/British%20Orthodontic%20Society/Author%20Content/Documents/PDF/Orthognathic%20Surgery%20May%2009.pdf	Primary researcher (First occasion)	99	40	66	7
	Primary researcher (second occasion)	96	39	60	5
	S.H.	105	49	66	8
http://en.wikipedia.org/wiki/Orthognathic_surgery	Primary researcher (First occasion)	114	46	72	7
	Primary researcher (second occasion)	111	49	90	7
	S.H.	112	39	75	6
http://www.qvh.nhs.uk/assets/patient_information/A%20guide%20for%20pts%20considering%20orthognathic%20surgery%20-Rvw%20March%202013.pdf	Primary researcher (First occasion)	96	63	66	8
	Primary researcher (second occasion)	93	56	60	9
	S.H.	99	60	72	8
http://www.cuh.org.uk/cms/addenbrookes-hospital/services/oral-and-maxillofacial-surgery-and-orthodontics/orthognathic-surgery	Primary researcher (First occasion)	106	34	78	7
	Primary researcher (second occasion)	104	35	84	5
	S.H.	110	32	84	7
http://www.lpch.org/DiseaseHealthInfo/HealthLibrary/craniofacial/maxface.html	Primary researcher (First occasion)	102	22	84	6
	Primary researcher (second occasion)	95	27	84	6
	S.H.	102	22	84	4
http://www.baoms.org.uk/What_is_Oral_and_Maxillofacial_Surgery/Sub_specialist_Areas/Orthognathic_Surgery	Primary researcher (First occasion)	114	22	84	4
	Primary researcher (second occasion)	107	30	84	4
	S.H.	103	45	84	6
http://www.youtube.com/watch?v=xTpllmuKmSE&list=TL6DuumWGJT08Nx84AD-RZTq8cIxYBlo8t	Primary researcher (First occasion)	114	22	72	7
	Primary researcher (second occasion)	109	44	39	9
	S.H.	108	15	54	1
http://www.orthognathicsurgery.info	Primary researcher (First occasion)	96	36	57	8
	Primary researcher (second occasion)	100	45	57	8
	S.H.	101	39	57	7
http://awimpsguidetoorthognathicsurgery.blogspot.com	Primary researcher (First occasion)	98	44	72	7
	Primary researcher (second occasion)	97	42	48	6
	S.H.	95	26	66	3
http://www.aaoms.org/conditions-and-treatments/corrective-jaw-surgery	Primary researcher (First occasion)	112	42	84	7
	Primary researcher (second occasion)	106	28	84	7
	S.H.	113	37	84	6

Table 13 AM checklist (final version) scores after guidance notes added.

Website	Assessor	AM Checklist (final version) score
http://www.caoms.com/orthognathic-surgery.aspx;"3";"error";"0";"211";"77";"6672592";"January 25	Primary researcher (First occasion)	3
	Primary researcher (second occasion)	2
	S.H.	3
http://www.ashfordstpeters.nhs.uk/attachments/157_Orthognathic%20Surgery.pdf;"0";"error";"0";"22396";"7	Primary researcher (First occasion)	5
	Primary researcher (second occasion)	5
	S.H.	7
http://www.oralmaxillofacialpartnership.co.uk/conditions-and-treatments/orthognathic-surgery/;"n/a";"error";"0";"0";"21";"n/a";"June 18	Primary researcher (First occasion)	8
	Primary researcher (second occasion)	7
	S.H.	9
http://underbitemyshorts.wordpress.com/;"0";"327";"10";"13";"71";"3868460";"May 8	Primary researcher (First occasion)	8
	Primary researcher (second occasion)	7
	S.H.	5
http://sites.google.com/site/maxillofacialkochikerala/orthognathic-surgery;"0";"n/a";"0";"6303059";"8	Primary researcher (First occasion)	7
	Primary researcher (second occasion)	7
	S.H.	7
http://www.wales.nhs.uk/sitesplus/866/opendoc/140196/;"n/a";"498	Primary researcher (First occasion)	8
	Primary researcher (second occasion)	7
	S.H.	8
http://www.gotbraces.com/Treatment/Orthognathic-Surgery.aspx;"wait...";"wait...";"wait...";"wait...";"30";"6375871";"July 4	Primary researcher (First occasion)	1
	Primary researcher (second occasion)	1
	S.H.	5
http://www.cosmeticdentistryguide.co.uk/articles/orthognathic-surgery.html;"3";"n/a";"wait...";"wait...";"27	Primary researcher (First occasion)	3
	Primary researcher (second occasion)	2
	S.H.	4
http://oralsurgeonhouston.com/services/orthognathic-reconstructive-surgery;"wait...";"wait...";"wait...";"wait...";"29";"14303204";"April 6	Primary researcher (First occasion)	2
	Primary researcher (second occasion)	2
	S.H.	2
http://www.interfacelondon.com/patients.php?action=jaw/	Primary researcher (First occasion)	0
	Primary researcher (second occasion)	0
	S.H.	2

Appendix 5: SCOPUS search term

(TITLE-ABS-KEY (qualit*) OR TITLE-ABS-KEY (evaluation) AND TITLE-ABS-KEY (Internet) OR TITLE-ABS-KEY (online) OR TITLE-ABS-KEY (world wide web) OR TITLE-ABS-KEY (net) AND TITLE-ABS-KEY (health WEBSITE) OR TITLE-ABS-KEY (medical WEBSITE) OR TITLE-ABS-KEY (dental WEBSITE) OR TITLE-ABS-KEY (patient information) OR TITLE-ABS-KEY (patient education) OR TITLE-ABS-KEY (health education) AND TITLE-ABS-KEY (tool*) OR TITLE-ABS-KEY (evaluation) OR TITLE-ABS-KEY (instrument)) AND (LIMIT-TO (LANGUAGE , "English"))

Appendix 6: Data for the 100 websites analysed in the study

Table 14 LIDA, DISCERN, HWAT 3.0 and AM checklist scores, website type, suffix and country of origin.

	Website	LIDA score	DISCERN score	HWAT 3.0 score	AM checklist score	Google rank	Suffix	Country of origin	Type of website
1	http://www.bos.org.uk/Resources/British%20Orthodontic%20Society/Author%20Content/Documents/PDF/Orthognathic%20Surgery%20May%2009.pdf	99	40	66	7	1	.org	UK	Professional organisation
2	http://en.wikipedia.org/wiki/Orthognathic_surgery	114	46	72	7	2	.org	Multiple	Other Patient information
3	http://www.qvh.nhs.uk/assets/patient_information/A%20guide%20for%20pts%20considering%20orthognathic%20surgery%20-Rvw%20March%202013.pdf	96	63	66	8	3	.nhs.uk	UK	Hospital
4	http://www.cuh.org.uk/cms/addenbrookes-hospital/services/oral-and-maxillofacial-surgery-and-orthodontics/orthognathic-surgery	106	34	78	7	4	.org.uk	UK	Hospital
5	http://www.lpch.org/DiseaseHealthInfo/HealthLibrary/craniofacial/maxface.html	102	22	84	6	5	.org	US	Hospital
6	http://www.baoms.org.uk/What_is_Oral_and_Maxillofacial_Surgery/Sub_specialist_Areas/Orthognathic_Surgery	114	22	84	4	6	.org.uk	UK	Professional organisation
7	http://www.youtube.com/watch?v=xTpllmuKmSE&list=TL6DuumWGJT08Nx84AD-RZTq8cIxYBlo8t	114	36	72	7	8	.com	Multiple	Support group/ video site/ Q& A site
8	http://www.orthognathicsurgery.info	96	44	57	8	10	.info	Multiple	Other Patient information
9	http://www.aaoms.org/conditions-and-treatments/corrective-jaw-surgery	98	33	72	7	14	.org	US	Professional organisation
10	http://awimpsguidetoorthognathicsurgery.blogspot.com	112	42	84	7	19	.com	UK	Blog
11	http://steffies-orthognathic-surgery.blogspot.com	105	42	72	5	20	.com	UK	Blog
12	http://medicine.yale.edu/surgery/plastics/care/jaw/ortho.aspx	103	32	84	1	23	.edu	US	Hospital
13	http://www.guysandstthomas.nhs.uk/resources/patient-information/childrens/jaw-(orthognathic)-	97	57	66	9	24	.nhs.uk	UK	Hospital
14	http://groups.yahoo.com/group/orthognathicsurgerysupport	99	31	48	5	26	.com	Multiple	Support group/ video site/ Q& A site
15	http://www.drbonine.com/orthognathic_surgery.html	107	41	75	6	33	.com	US	Private practice
16	http://www.cmsllc.com/toport.html	102	38	75	6	38	.com	US	Private practice

	Website	LIDA score	DISCERN score	HWAT 3.0 score	AM checklist score	Google rank	Suffix	Country of origin	Type of website
17	http://ortho_club.tripod.com	86	42	48	3	39	.com	Multiple	Support group/ video site/ Q& A site
18	http://www.childrenshospital.org/health-topics/procedures/orthognathic-surgery	112	43	84	8	42	.org	US	Hospital
19	http://www.thefacesurgeon.co.uk/surgical-treatments/orthognathic-surgery	107	28	75	3	43	.co.uk	UK	Private practice
20	http://iagreedtoorthognathicsurgery.blogspot.com	106	30	66	5	45	.com	Canada	Blog
21	http://www.caoms.com/orthognathic-surgery.aspx	96	22	84	4	46	.com	Canada	Professional organisation
22	http://www.drpitts.com/orthognathic	117	37	75	7	47	.com	US	Private practice
23	http://www.massgeneral.org/omfs/services/procedure.aspx	108	21	84	2	50	.org	US	Hospital
24	http://www.palm-panek.com/procedures/oral-maxillofacial-surgery/corrective-jaw-surgery-	107	23	75	4	51	.com	US	Private practice
25	http://www.ashfordstpeters.nhs.uk/attachments/157_Orthognathic%20Surgery.pdf	92	35	90	6	52	.nhs.uk	UK	Hospital
26	http://www.mtlmf.com/en/services-en/orthognathic-surgery	101	19	84	5	53	.com	Canada	Private practice
27	http://www.londonorthognathiccentre.com	105	33	75	6	54	.com	UK	Private practice
28	http://www.visagefacialsurgery.com/orthognathic-surgery.html	58	24	69	2	55	.com	US	Private practice
29	http://surgery.med.umich.edu/plastic/patient/ped_procedures/orthognathic	104	16	78	1	56	.edu	US	Hospital
30	http://www.drposnick.com/orthognathic/man_deficiency.html	74	18	54	2	57	.com	Multiple	Private practice
31	http://www.mayoclinic.org/tests-procedures/jaw-surgery/basics/definition/PRC-20013370	106	34	90	5	60	.org	US	Private practice
32	http://oralfacial.com/orthognathic-surgery.php	113	29	75	5	61	.com	US	Private practice
33	http://www.oralmaxillofacialpartnership.co.uk/conditions-and-treatments/orthognathic-surgery	111	39	69	7	64	.co.uk	UK	Private practice
34	http://meshamcd.wordpress.com	109	32	60	5	66	.com	US	Blog
35	http://agaveclinic.com/en/orthognathic.php	72	40	51	9	67	.com	Europe	Private practice
36	http://patient-info.com/ortho.htm	93	34	75	7	68	.com	US	Other Patient information
37	http://www.ndcs.com.sg/ForPatientsAndVisitors/ConditionsAndTreatments/Glossary/Pages/JawSurgery.aspx	103	21	75	4	69	.com	Asia	Hospital
38	http://underbitemyshorts.wordpress.com	101	31	66	5	70	.com	US	Blog
39	http://stanfordhospital.org/clinicsmedServices/COE/surgicalServices/oralsurgery/procedures/jaw.html	113	33	78	5	72	.org	US	Hospital
40	http://www.nbt.nhs.uk/webfm_send/1161	98	47	60	6	83	.nhs.uk	UK	Hospital

	Website	LIDA score	DISCERN score	HWAT 3.0 score	AM checklist score	Google rank	Suffix	Country of origin	Type of website
41	http://sites.google.com/site/maxillofacialkochikerala/orthognathic-surgery	97	39	84	8	89	.com	Asia	Hospital
42	http://www.wales.nhs.uk/sitesplus/866/opendoc/140196/	95	53	60	9	90	.nhs.uk	UK	Hospital
43	http://www.drjui.com/surgical-instructions/after-jaw-surgery.html	110	27	75	5	91	.com	US	Private practice
44	http://nyp.org/health/orthognathic-surgery.html	96	28	84	2	92	.org	US	Hospital
45	http://www.smilesolutions.com.au/orthodontics/orthognathic-jaw-surgery	100	18	69	3	93	.au	Australi	Private practice
46	http://doublejawsurgery.com/day-9-the-purpose-of-the-splint	123	38	60	8	94	.com	Canada	Blog
47	http://www.jawpain.com/surgical-instructions/jaw-surgery.html	107	18	75	3	95	.com	US	Private practice
48	http://www.maxfaxsho.co.uk/Orthognathic-surgery	95	31	60	6	98	.co.uk	UK	Other Patient information
49	http://institutomaxilofacial.com/en/orthognathic-surgery	107	19	60	2	103	.com	Europe	Private practice
50	http://www.caryoralsurgery.com/oral_surgery_services/orthognathic_surgery.html	94	23	75	4	105	.com	US	Hospital
51	http://www.toothloop.com/index.php/adults/oral-a-maxillofacial-surgery/orthognathic-surgery.html	112	29	75	4	106	.com	UK	Private practice
52	http://montanaoralsurgery.com/orthognathicsurgery.php	97	18	75	3	107	.com	US	Private practice
53	http://www.implantandfacialsurgery.com/south_florida_dental_surgery/oral_maxillofaclal_surgery/orthognathic_jaw_surgery.html	97	30	69	5	110	.com	US	Private practice
54	http://www.novamedical.co.nz/procedures/dentistry/orthognathic--jaw--surgery	102	27	75	4	111	.co.nz	Australi	Other Patient information
55	http://borealisplasticsurgery.com/orthognathic-surgery	108	23	75	1	112	.com	US	Private practice
56	http://www.cincinnatijawsurgery.com/downloads/A_Patient's_Guide_to_Orthognathic_Surgery.pdf	57	39	51	5	115	.com	US	Private practice
57	http://www.la-coms.com/patient-info/surgical-instructions/orthognathic-surgery	92	26	69	3	117	.com	US	Private practice
58	http://www.drwmcdonald.com/Orthognathic.htm	87	25	75	4	118	.com	Canada	Private practice
59	http://www.wssoms.com/jaw-surgery.aspx	105	27	78	3	120	.com	US	Professional organisation
60	http://www.omfsaboutface.co.uk/orth.htm	86	36	69	5	122	.co.uk	UK	Support group/ video site/ Q& A site
61	http://www.maxillosurgeon.com/orthognathic-jaw-surgery-tijuana.html	98	21	69	2	123	.com	South/ C	Private practice
62	http://www.thailandvipservices.com/medical-services/dental-services/dental-implants/orthognathic-surgery	102	23	60	4	125	.com	Asia	Private practice
63	http://www.faceandjawsurgery.com/oral-surgery-procedures-nd/jaw-surgery.html	110	18	69	1	126	.com	US	Private practice
64	http://www.ofscms.com/public/procedures/orthognathic.html	88	16	69	2	128	.com	US	Private practice

	Website	LIDA score	DISCERN score	HWAT 3.0 score	AM checklist score	Google rank	Suffix	Country of origin	Type of website
65	http://www.pineypointoms.com/orthognathic-surgery.html	102	29	66	6	129	.com	US	Private practice
66	http://www.dchft.nhs.uk/patients/wards-depts/orthodontics/patientleaflets/OrthognathicSurgery.pdf	91	35	60	5	131	.nhs.uk	UK	Hospital
67	http://drrichardjoseph.com/jaw-surgery/benefits.php	114	39	69	7	132	.com	US	Private practice
68	http://www.kiourtsisortho.com/treatment/orthognathic-surgery.aspx	105	19	75	1	135	.com	US	Private practice
69	http://www.ips-kr.com/eng/beautryshow.php?beautry_number=150	87	18	66	1	137	.com	Asia	Private practice
70	http://www.lashrubinorthodontics.com/orthognathic-surgery	109	24	75	2	138	.com	US	Private practice
71	http://whitemountainoms.com/procedures/orthognathic.html	100	23	75	4	140	.com	US	Private practice
72	http://www.oralsurgeryandimplants.com/oral-surgery/jaw-surgery.html	103	33	75	7	141	.com	US	Private practice
73	http://www.uhns.nhs.uk/OurServices/ClinicalServices/AZofClinicalServices/Orthodontics/Patientsintreatment/SurgicalOrthodonticorthognathicTreatment.aspx	106	16	84	2	142	.nhs.uk	UK	Hospital
74	http://www.brilliantortho.com/orthognathic-surgery.php	105	24	75	4	143	.com	US	Private practice
75	http://www.gotbraces.com/Treatment/Orthognathic-Surgery.aspx	103	16	75	0	148	.com	US	Private practice
76	http://www.cosmeticdentistryguide.co.uk/articles/orthognathic-surgery.html	100	28	51	3	149	.co.uk	UK	Other Patient information
77	http://jaw.muppethouse.com	110	45	66	6	151	.com	US	Blog
78	http://www.nycoms.com	86	29	75	7	152	.com	US	Private practice
79	http://www.ourorthodontist.com/orthognathic-surgery	108	28	75	2	154	.com	US	Private practice
80	http://www.spielberg-ortho.com/treatment/orthognathic-(jaw)-surgery.aspx	109	26	75	3	156	.com	US	Private practice
81	http://www.kalantoms.com/orthognathic-surgery.html	106	23	75	3	157	.com	US	Private practice
82	http://www.cosmeticvacations.co.uk/cosmetic_dentistry/orthognathic_surgery.php	105	28	75	5	159	.co.uk	South/ C	Private practice
83	http://3fivetwo.com/group/blogpost/Life_changing_Orthodontics_and_Orthognathic_surgery_available_at_3fivetwo	104	20	60	1	160	.com	Europe	Private practice
84	http://www.chla.org/site/c.ipINKTOAJsG/b.7866455/k.C019/About_Jaw_Surgery__Orthognathic_Surgery__Under_Bite__Over_Bite.htm	101	22	84	3	164	.org	US	Hospital
85	http://oralsurgeonhouston.com/services/orthognathic-reconstructive-surgery	102	16	75	1	165	.com	US	Private practice
86	http://www.midlandoms.com/treatment/orthognathic-surgery	104	20	75	3	168	.com	US	Private practice
87	http://www.interfacelondon.com/patients.php?action=jaw	104	20	75	2	170	.com	UK	Private practice
88	http://www.knightortho.com/orthognathic-surgery.php	107	24	75	4	173	.com	US	Private practice
89	http://peacefrog1987.tripod.com	80	29	60	5	175	.com	US	Blog

	Website	LIDA score	DISCERN score	HWAT 3.0 score	AM checklist score	Google rank	Suffix	Country of origin	Type of website
90	http://www.drbdorfman.com/ortho.htm	92	16	75	1	176	.com	US	Private practice
91	http://whitneysjawsurgery.blogspot.com	110	52	66	8	183	.com	US	Blog
92	http://www.drnoor.com/services/oral-surgery/corrective-jaw-or-orthognathic-surgery	105	26	75	4	184	.com	US	Private practice
93	http://www.gopherbraces.com/Treatment/Orthognathic-Surgery.aspx	105	21	84	2	186	.com	US	Hospital
94	http://www.lingualuk.com/orthognathic.html	109	16	75	1	188	.com	UK	Private practice
95	http://ask.metafilter.com/238992/howto-orthognathic-surgerybracesgeneral-facefixing-for-adults	107	52	72	9	191	.com	US	Support group/ video site/ Q& A site
96	http://aimeesfunkyjaw.blogspot.com	109	37	66	8	192	.com	UK	Blog
97	http://www.smilewithconfidence.com.au/Treatment/Orthognathic_Surgery.aspx	118	45	84	7	195	.com	Australi	Private practice
98	http://cirugiafacialbenidorm.com/units/unit-rhinoplasty-and-orthognathic-surgery/?lang=en	105	23	75	4	198	.com	Europe	Private practice
99	http://www.seoultouchup.com/cheekbone-jawbone-surgeries	107	22	75	0	199	.com	Asia	Private practice
100	http://trumanorthodontics.blogspot.co.uk/2012/05/orthognathic-surgery.html	106	37	90	4	202	.co.uk	UK	Private practice