Lena Nordheim

How credible is lay health information on the World Wide Web?

-

A study of information about the effects of interventions

Masteroppgave

Avdeling for journalistikk, bibliotek og informasjonsfag

ABSTRACT

Background: The World Wide Web is a source to a wealth of health information that was previously unavailable to lay people. Formal studies have raised concerns about the credibility of online materials on the Web; most of them have been performed on English-language websites.

Objectives: To investigate the credibility of Norwegian-language lay information about the effects of interventions on the World Wide Web.

Methods: I performed a cross-sectional study of Norwegian lay health webpages on two specific topics: 1) Effect of low glycemic diets for weight reduction. 2) Effects of taking extra calcium supplementation during pregnancy to prevent pre-eclampsia and other related outcomes. Webpages were identified via four search engines and SMIL. Information on the included webpages was compared to the results of Cochrane reviews. I used a set of commonly used technical criteria to evaluate the transparency of webpages/sites. Furthermore, I compared the Webpages compliance with the criteria to their accuracy scores to see if there was an association between accuracy (information credibility) and the technical criteria (surface credibility).

Results: 56 webpages were included, 49 from commercial providers. Overall, only above 40% of the webpages were evaluated as accurate. The low glycemic diets web pages were more inaccurate than the calcium supplementation webpages. There were no apparent differences between accurate webpages and inaccurate webpages in complying with the technical criteria.

Conclusion: Lay health information on the Web seems to be inaccurate, unclear and sometimes inconsistent. Technical criteria will not sufficiently distinguish accurate materials from inaccurate materials. Further research is needed on other, more clinical topics to ensure the inclusion of public agency websites.

Masteroppgave ved Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag

Oslo 2009

TAKK

Selv om denne masteroppgaven har engelsk språkform tar jeg meg den uforbeholdne rett til å fremsette alle takksigelser på morsmålet.

Takk til mine gode kolleger Ragnhild Øye Bjarkøy, Marjolein Iversen, Liv Irene Ruud, Birgitte Graverholt, og Nina Rydland Olsen for medvurderinger av nettsidene i dette prosjektet. En spesiell takk til Nina for gjennomlesning av utkast og uvurderlige innspill på innhold, språk og flyt.

Takk til Monica Nortvedt, innspirende og engasjert leder, som har latt meg bruke arbeidstid på å ferdigstille oppgaven. Uten hennes støtte ville denne oppgaven antakeligvis blitt et livsprosjekt.

Takk til venner og familie for all tålmodighet og forståelse for et tilnærmet 100% fravær fra sosiale aktiviteter det siste året. Takk også til katten min Dennis som, til tross for vanskjøtsel over en lengre periode, har vært like henrykt og kjælen hver gang jeg har kommet på tanken om å ta turen hjemom.

Sist, men ikke minst, tusen takk til veileder Nils Pharo for gode råd og klok veiledning i alle ledd av oppgaven. Takk også spesiell for støtte og oppmuntring i perioden hvor det gikk som tråest.

Bergen, 15. Februar 2009

Lena Nordheim

CONTENTS

1. INTRODUCTION 7

1.2 OBJECTIVES 9

1.3 OUTLINE OF THESIS 10

1.3 SPECIFICATIONS AND REMARKS TO THE TEXT 10

1.3.1 definitions 12

2. BACKGROUND 15

2.1 EVIDENCE-INFORMED PATIENT CHOICE 15

2.1.1 Evidence-based health care 15

2.1.2 Patient centred care 17

2.2 THE ROLE OF SYSTEMATIC REVIEWS IN EVIDENCE-BASED HEALTH CARE 19

2.3 CRITISISMS OF EVIDENCE-BASED HEALTH CARE 22

2.4 EVIDENCE-INFORMED PATIENT CHOICE AND INFORMATION ACCESS 23

3. THEORETICAL FRAMEWORK 26

3.1 DEFINING THE CREDIBILITY CONSTRUCT 26

3.2 WHAT MAKES WEBSITES CREDIBLE? 28

3.3 CREDIBILITY FRAMEWORK FOR HEALTH CARE WEBSITES 30

4. PREVIOUS RESEARCH 35

4.1 IDENTIFYING RELEVANT LITERATURE 35

4.2 EVALUATIONS OF THE CREDIBILITY OF ONLINE HEALTH INFORMATION 36

4.2.1 The criterion standard appraoch 36

4.2.2 The evaluation instrument approach 39

4.3 SUMMARY AND RESEARCH QUESTIONS 45

5. METHODS 47

5.1 STUDY DESIGN 47

5.2 THE CRITERION STANDARD 48

5.2.1 Selecting the systematic reviews 48

5.2.2 Comparing webpages to the criterion standard 49

5.3 IDENTIFYING WEBPAGES FOR EVALUATION 54

5.3.1 Search strategy 54

5.3.2 Sampling 56

5.4 EVALUATION CRITERIA 58
5.4.1 Defining the technical evaluation criteria 58
5.4.2 Web operator criteria 60
5.4.3 Content criteria 61
5.4.4 Design criteria 63
5.5 METHODOLGICAL CONSIDERATIONS AND LIMITATIONS 64
5.6 ETHICAL CONSIDERATIONS 66

6. FINDINGS 67

6.1 RESULTS FROM SEARCH ENGINES AND SMIL SEARCHES 67
6.2 ACCURACY OF WEB STATEMENTS (INFORMATION CREDIBILITY) 69
6.2.1 Coding of the criterion standard 69
6.2.2 Coding of the webpages 71
6.2.3 Comparison of Web statements to the cochrane reviews 78
6.3 COMPARING ACCURACY WITH THE TECHNICAL CRITERIA 83
6.3.1 web operator: structural criteria 83
6.3.2 Web operator: development criteria 86
6.3.3 Web content: authorship criteria 88
6.3.4 Web content: currency criteria 91
6.3.5 Web content: attribution criteria 92
6.3.6 Web design criteria 96
6.4 THE HEALTH PORTAL "SMIL" VERSUS SEARCH ENGINES 97
6.5 PUBLIC AGENCY SITES VERSUS COMMERCIAL SITES 97

7. DISCUSSION 99

7.1 ACCURACY OF WEB PAGES (INFORMATION CREDIBILITY) 100
7.2 COMPARING ACCURACY AND COMPLIANCE WITH TECHNICAL CRITERIA 103
7.2.1 Web content: authorship 105
7.2.2 Web content: currency 106
7.2.3 Web content: attribution 107
7.3 CONTENT DEVELOPMENT PROCEDURES 108

8. CONCLUSION 110

8.1 IMPLICATIONS FOR PRACTICE 111
8.2 IMPLICATIONS FOR RESEARCH 111

REFERENCES 113

APPENDICES 124

Appendix 1: medline search strategy 124

Appendix 2: Cochrane Library search strategy for pregnancy reviews 125

Appendix 3: Coding schemes for the Cochrane reviews 126

Appendix 4: coding schemes for Web statements 129

Appendix 5: Search strategies in search engines and SMIL 131

appendix 6: web operator criteria 132

Appendix 7: web content criteria 133

Appendix 8: website design criteria 135

appendix 9: included websites and distribution of webpages 136

appendix 10: references to cited webpages 137

1. INTRODUCTION

In recent years, lay people have become more active in making decisions about their health care. Those with a medical condition or illness want to receive information and participate in planning management of their condition (Benbassat, Pilpel, & Tidhar, 1998). The growing recognition of lay peoples' wish to be informed is reflected in patient rights' legislation established during the last decade. In Norway, the Act relating to Patients' Rights establishes the patient's right to health care (Patients' Rights Act, 1999). It also establishes the patient's right to participation and information. Accordingly, it is the patient's democratic right to access information about diagnosis, treatment and other aspects of health care.

In the health services the principles of *evidence-based health care* have grown in importance and acceptance during the last decades (Eddy, 2005). A main rationale behind the evidence-based model in health care was a growing recognition that there was a gap between clinical research and what was actually happening in clinical practice. The term «evidence» reflects a focus on applying findings from scientific research in clinical decisions about individual patients as well as in public health decisions and policy. The model also emphasises patient, or consumer involvement in decisions about their health (Guyatt, Cook, & Haynes, 2004). This has given rise to the concept of evidence-informed patient choice (V. A. Entwistle, Sheldon, Sowden, & Watt, 1998). A fundament of evidence-based patient choice is to provide patients with evidence from research in order for them to make informed decisions about their treatment and care. This implies that patients or lay people in general, must have access to research-based information, which is also commonly referred to as evidence-based information (Glenton, Nilsen, & Carlsen, 2006).

Only a few years ago lay peoples' access to health information was limited. People had to rely solely on information provided by their doctor, other health care providers, or the public library service. In most libraries the information available was home doctor compendia and encyclopaedias (Muir Gray, 2008). The introduction of the World Wide Web (the Web) has given people, both patients and professionals, access to thousands of pages of health

information that was previously inaccessible. Increasingly, people seek information on the Web to help them with health decisions, and thousands of websites now try to deliver health information to them. A study of US citizens concludes that 60% use the Web for health-related information (Fox, 2006). In Europe the usage has been more moderate, although the number of EU-citizens using the Web for health purposes has increased throughout the years. Surveys have estimated that 30-40% of the European population has searched for health information (Andreassen et al., 2007; European Commission, 2005). The EU-surveys also reveal large variations in usage between countries, with Northern-European countries having higher usage rates than the Eastern- and Southern-European countries.

In Norway, two surveys on Web use for health purposes have been undertaken, in 2001 and 2005 respectively (Andreassen, Sandaune, & Gammon, 2002; Andreassen, Wangberg, Wynn, Sørensen, & Hjortdahl, 2006). Physicians and other health professionals are still the main source for health information. However, an increasing number of people also use the Web for information seeking. Fifty-eight percent used the World Wide Web to search health related content in 2005 compared to 31% in 2001. Both the national and the international surveys reveal that there is a larger proportion of health seekers among individuals with a higher education and income.

Most lay health seekers are looking for information about a particular illness or condition for themselves, family members or friends (McMullan, 2006). Health information covers many aspects, including prevalence of disease, diagnostics, causes and risk factors, effect of treatment or prevention, prognosis, and experiences of having a health condition. People who are faced with illness seem to search for information on diagnostics and treatment purposes. The social aspect of Web use is also important. Searches may serve for experience sharing and support from others with the same health condition via online support groups, chat rooms or bulletin boards. People who are not ill may also search the World Wide Web with a well-defined purpose in mind, for example to prevent disease, improve well-being, or to stay healthy (Renahy & Chauvin, 2006).

One of the great advantages of the World Wide Web, in contrast to traditional media, is that it allows frequent updating of information. However, there are also some disadvantages. Access to an ever increasing range of health information on the Web makes it difficult to identify

relevant and credible information. Anyone without any real competence in a health topic can provide a website. Seeking useful and valid information can be difficult because of the speed and lack of censorship and editorial control of what is published (Wyatt, 1997). Moreover, searching and locating information are only starting points. Judging whether the information is applicable and credible may present a greater challenge than just searching for information. In order for lay people to take well-informed decisions regarding their health, he or she needs access to credible information.

Several formal studies have demonstrated problems with the credibility of health information for lay people on the Web, most of them have evaluated English-language websites (Eysenbach, Powell, Kuss, & Sa, 2002). The studies raise concern over inaccurate or «non-evidence-based» information. Previous studies found that World Wide Web users do not necessarily assess quality when searching for health information online (Eysenbach & Köhler, 2002; Meric et al., 2002). Moreover, they often lack the medical knowledge to accurately assess whether the information in webpages is credible and appropriate (Shon, Marshall, & Musen, 2000). Therefore, the credibility of online health information is an area of concern for researchers, clinicians and consumers.

1.2 OBJECTIVES

As previously described, many of those seeking out health information on the Web do so in order to manage a disease or condition. Thus, I will focus on health information related to the effects of treatment and preventive measures. The overall and broad research question for my thesis is:

How credible is Norwegian health information about the effect of interventions for lay people on the World Wide Web?

My specific research questions are:

1. How accurate is lay health information about the effect of interventions on Norwegian websites when compared to the results of systematic reviews?

2. Can the technical criteria be used to evaluate information accuracy?
3. Is lay information identified via the health portal SMIL more accurate compared to information found using general search engines?
4. Is information on public agencies sites more accurate than information on commercial sites?
5. What procedures are in place to ensure that research evidence is included when developing content, and what sources of evidence are used?

1.3 OUTLINE OF THESIS

The concept underlying my thesis is evidence-informed patient choice, which I describe in more detail in Chapter 2. In Chapter 3 I present a theoretical framework for evaluating credibility in health websites, which constitutes the foundation for the methods used, data analysis and interpretation. Chapter 4 gives an overview of relevant research and outlines the rationale for the specific research questions addressed in my thesis. I present the study design and the methods used for data collection in Chapter 5; before I go on to present my findings in Chapter 6. In Chapter 7 I discuss the findings. Finally, in Chapter 8 I summarise and make suggestions for future research on the topic.

1.3 SPECIFICATIONS AND REMARKS TO THE TEXT

Although the communicative and supportive aspect of seeking out health information is important, my study will deal with the « informational» facet of cyberhealth rather than its «communication» one. In other words, I will take a closer look at direct information tools and not address online communication tools such as forums and chats. This decision was made due to the scope and constraints of a Master's thesis, and not because I do not acknowledge

the experiential facet of health information. In fact, concerns have been expressed about exchange of non-credible information among lay people using World Wide Web communication tools. However, credibility concerns have also been expressed about expert and authorative sources providing lay health information (Coulter et al., 2006). Thus, my focus will be on information provided or supported by health care providers and others claiming expertise in their respective areas.

Throughout this thesis I use the term «lay people», which I define as anyone without a health or medical background who is searching for and using health-related information. I decided not to use the term «patient» because of its stereotypical connotations as someone who is ill or have a particular health condition. Many people who are looking for health-related information on the World Wide Web will not be ill. Instead, they might be seeking out answers to prevent rather than cure disease, for example on issues related to a healthy lifestyle. However, I had to make one exception in Chapter 2, where I describe the concept of «evidence-informed patient choice». This is a defined concept, although the word «patient» is also used in a broader context.

In Chapter 6 I quote different statements from Norwegian health websites to illustrate and support my assessments. This thesis is written in English and consequently I have translated all statements. Because the meaning and context of statements are assessed in their original language, these aspects may be lost through translation. Accordingly, I decided to accompany all quotes with the original statement in Norwegian.

In the next sub-section I give some definitions of terms and statistical measures that are used throughout this thesis, or that have been otherwise important for the methods and interpretation of findings in my thesis.

1.3.1 DEFINITIONS

Confidence interval (CI)

A confidence interval (CI) quantifies the uncertainty in measurement. A CI calculated for a measure of intervention effect shows a range within which the true effect is likely to lie. It usually reported as 95% CI (Davies, 2001). For example, if taking calcium supplementation in pregnancy reduce the risk of pre-eclampsia by 50% with a 95% CI ranging from 31% to 67%, we can be 95% sure that the true risk reduction was between 31% and 67%.

A CI that embraces the value of no difference indicates that the intervention under investigation is not significantly different from the control. If the outcome is dichotomous (i.e. dead or alive, pregnant or not pregnant) the value of no difference is usually 1. If the outcome is continuous (i.e. level of depression measured on a scale), the value is 0.

Intervention

Here the term *intervention* is used about preventive and curative actions taken to improve an outcome. An intervention can be a medication, a surgical procedure, or a (new) service. Interventions can also be educational, for example educating diabetes patients to cope with their disease, or they can have a lifestyle aspect, for example dietary or exercise interventions.

P-value

The p-value indicates the probability (ranging from zero to one) that the results observed in a study could have occurred by chance. A p-value of 0.05 or below is considered to be statistically significant. If there is an observed difference in effect in favour of the intervention group and the p-value is 0.05 or less, then the difference is probably due to the intervention and has not simply occurred by chance (Houle & Stump, 2008).

Randomised controlled trial

In a randomised controlled trial the patients, or participants, are randomly allocated to two or more different interventions. The control intervention could be a different intervention, a placebo, or no intervention. The purpose of randomisation is to ensure that both known and unknown confounding factors are evenly distributed between the intervention groups, so that the groups differ only with respect to the interventions being compared. The test that randomisation has been successful is that different treatment groups have same characteristics at baseline. For instance, there should be the same number of men and women, or older or younger people, or different degrees of disease severity (Straus, Richardson, Glasziou, & Haynes, 2005).

Relative risk (RR)

The relative risk (RR) is the risk of an event in the intervention group divided by the risk of an event in the control group. It is usually expressed as a decimal proportion, sometimes as a percentage (Davies, 2001).

Statistical significance

Here, statistical significance relates to whether an observed difference in effect between those receiving an intervention (intervention group) and those who don't (control group) is due to chance (Houle & Stump, 2008).

Systematic review

A systematic review is a review of the evidence on a clearly formulated question that uses systematic, explicit and reproducible methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review (Centre for Reviews and Dissemination, 2001)

Webpage

A webpage is a «collection of information, consisting of one or more web resources, intended to be rendered simultaneously, and identified by a single URI» (Lavoie & Nielsen, 1999).

Website

A website is a «collection of interlinked webpages, including a host page, residing at the same network location» (Lavoie & Nielsen, 1999).

Web statements

I my thesis Web statements refer to information about the effects of an intervention provided on a webpage. An example of a Web statement could be «Calcium supplementation is recommended if you don't drink milk or eat cheese».

2. BACKGROUND

2.1 EVIDENCE-INFORMED PATIENT CHOICE

Evidence-informed patient choice «involves providing people with research-based information about the effectiveness of healthcare options and promoting their involvement in decisions about their treatment» (V. A. Entwistle et al., 1998, p. 13). Behind the concept lies the idea that the patient's insight to research-based information will raise the quality of his or her health care decision-making. It is also part of a broader assumption that a better public understanding of science in general will enrich society and the individual (Claire Glenton et al., 2006).

Although the concept includes the term «patient», the term is used within a wider definitional frame, and considers all people who interact with health care professionals or services (Elwyn & Edwards, 2001, p. 11). Evidence-informed patient choice is a paradigm at the intersection of two other concepts: evidence-based health care and patient-centred care (Eysenbach & Diepgen, 2001). The two concepts are closely linked because they both emphasise that it is not the authority of the doctor or expert that justifies a particular clinical intervention. In the next sections I will describe these concepts briefly.

2.1.1 EVIDENCE-BASED HEALTH CARE

The concept of evidence-based health care is derived from evidence-based medicine, originally defined by Sackett, Rosenberg, Gray, Haynes, & Richardson (1996, p. 71) as:

> (…) the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research .

Evidence-based medicine as a concept has grown in importance and acceptance in recent years and its' principles has been transferred to other health professional areas such as nursing, physiotherapy, occupational therapy, and dentistry. The term «evidence-based health care» encapsulates all the different health professions as well as health policy makers and managers (Muir Gray, 2001a). The basic steps of an evidence-based approach for health care professionals are usually outlined as (V. Entwistle & O'Donnell, 2001):

6. Formulate a precise question;
7. Search the health care literature for research evidence that addresses the question;
8. Critically appraise the research evidence to assess its validity and relevance to the question, person, and situation concerned;
9. Apply the best available evidence to make decisions about health care;
10. Reflect on performance

The steps should be viewed as an ongoing process rather than a just a sequence of steps, as the last step, reflection on performance, might generate new questions. Evidence-based health care challenges the traditional clinical decision making where the clinician's understanding of the value of diagnostic tests and treatment efficacy are primarily guided by observations from his or her own experience; through knowledge learnt through the rigors of education; followed by continuing education, journals, and interaction with colleagues. The clinician's experience-based knowledge, also merely referred to as «experience», is still an important component in the decision-making process, yet it is complementary to rather than a substitute for research (Glenton, 2006).

The focus on research evidence as part of clinical decision making started in the late 1980s when audit studies of health service delivery revealed that there were many gaps between what was known and what was done (Muir Gray, 2001b). The gaps identified were:

- failure to implement new interventions that did more good than harm, such as aspirin after a heart attack;

- failure to stop interventions that did more harm than good, for example undergoing mastectomy instead of lumpectomy and radiation in early breast cancer

The audit studies clearly indicated that people did not use research finding systematically and promptly.

In Norway, several national strategic plans and government documents emphasise the principles of evidence-based health care to improve the quality of health services, clinical practice, and education of health professionals. The White Paper 39/1999 (St.meld. nr. 39, 1999), published ten years ago underline the use of research evidence as a basis for future decision making in health care and other professional areas. The National strategy for quality improvement in social and health service explicitly states that research evidence is a prerequisite for good quality care and effective interventions (Directorate of Health and Social Affairs, 2005a). One priority area in the strategy targets those engaged in the higher education of health and social care professionals to develop operating principals that ensure the following vision to be achieved:

> All health professionals should be educated to deliver patient-centred care as members of an interdisciplinary team, emphasising evidence-based practice, quality improvement approaches, and informatics (Directorate of Health and Social Affairs, 2006, p. 4).

2.1.2 PATIENT CENTRED CARE

The first definition of evidence-based medicine by Sackett et al. (1996) cited above does not take the third and important component of evidence-based health care into consideration, namely the patient. However, the latest definition of evidence-based health care (or practice) goes far into putting the patient at front when making decisions regarding their health, by stating that:

> Evidence-Based Practice (EBP) requires that decisions about health care are based on the best available, current, valid and relevant evidence. *These decisions should be made by those receiving care*, informed by the tacit and explicit knowledge of those providing care, within the context of available resources (Dawes et al., 2005, p. 4, my italics).

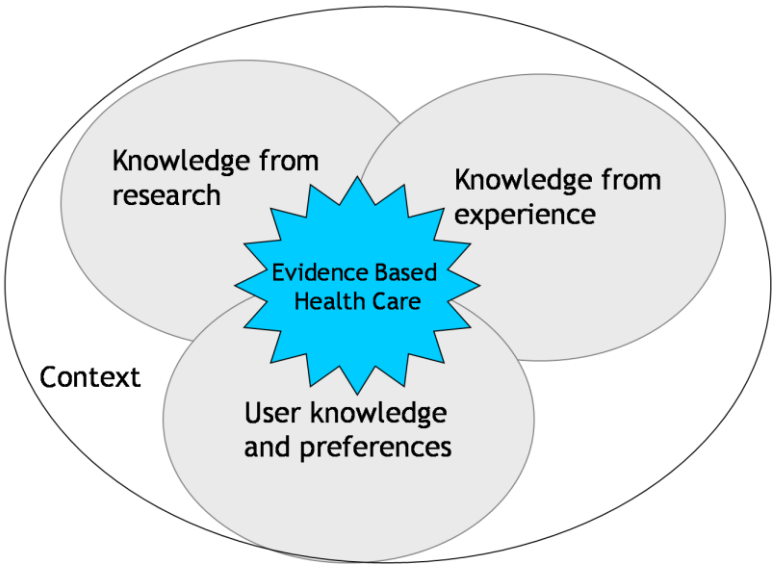
This reflects a general movement in health care, from the paternalistic model to the patient-centred model. The paternalistic model emphasizes doctors' authority and patient compliance with what the doctor decides. In the patient-centred model the patient plays a central role in decisions about health care, stressing the fact that patient preferences must be incorporated into clinical decision making (Benbassat et al., 1998; Glenton, 2006).

The patient-centred care model embraces the concepts of shared decision making and patient choice. In shared decision making, the health care provider and the patient work collaboratively to select treatment and care that includes patient experiences and preferences (Benbassat et al., 1998; Ruland, 2005). This is particularly important when difficult decisions have to be made, such as the choice to undergo prenatal testing. Few screening or treatment options come without potentially harmful consequences for the patient; accordingly patients' values and preferences must be considered in the decision-making process (Rudland, 2005). In Norway, patient or user involvement is declared a high-priority health policy goal (Directorate for Health and Social Affairs, 2005a), which is also reflected it the Patients' Right Act (1999).

Jamtvedt & Nortvedt (2008) suggest that the perhaps most important aspect of evidence-based health care is to use research-based information in collaboration with the patient. Clinical reasoning and experience is used to adapt patient preferences and research evidence to the individual patient situation: What are the experiences from other similar situations? How would the patient's characteristics influence the outcomes?

In summary, the evidence-based health care model integrates different and equally important types of knowledge: knowledge from experience, from patients' and other consumers' ('users') knowledge, and knowledge from research. The context or setting, such as financial and organisational constraints, influences the preferences and values of the patients and the health professional as well as how the research evidence is applied (see Figure 1).

Figure 1: ***Evidence-based health care - the model***[1]



2.2 THE ROLE OF SYSTEMATIC REVIEWS IN EVIDENCE-BASED HEALTH CARE

As described earlier in this chapter, evidence-based health care comprise several steps, including a search for scientific evidence and critical appraisal of the evidence. However, this model has limitations as practicing the steps is time consuming and sometimes impeded by limited access to research databases and full-text journals. Proponents of evidence-based health care have increasingly focused on the need for systems to support evidence-informed decisions, including rapid access to evidence-based information (Haynes, 2007).

Along with the introduction of the evidence-based health care in the early 1990s, a new type of publication saw the light of day: the systematic review. A systematic review is a review of the evidence on a clearly formulated question where the authors have used a systematic, explicit and reproducible method to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review (Centre for Reviews and Dissemination, 2009). The production of systematic reviews is

[1] Translated from www.kunnskapsbasertpraksis.no

regarded as a means to make research more easily accessible to health care professionals (Muir Gray, 2001b).

Why systematic reviews? In the past, busy health care professionals turned to textbooks for expert knowledge. However, a landmark study published in 1992 showed that textbooks responded slowly to new knowledge. This study revealed that it took many years for beneficial treatments to enter the standard texts, while treatments that had been proved harmful or at best ineffective continued to be recommended by prestigious authors long after their use had been discredited by research (Antman, Lau, Kupelnick, Mosteller, & Chalmers, 1992).

Health professionals who recognised that textbooks could quickly become out of date used other methods of keeping abreast of new developments; they relied on editorials and review articles. However, researchers who have investigated the scientific rationale behind editorials and traditional reviews have demonstrated that they were in fact biased and unreliable. Two main reasons were that the authors did not describe the methods used to find and evaluate the research papers cited in the review, and citation was selective in that experts only cited those studies that supported their views (Muir Gray, 2001b).

Systematic reviews, on the other hand, are supposed to be objective and unbiased because the authors explicitly state the purpose for the review; they have clearly defined criteria for including and excluding studies and for appraising them; and they explicitly state their methods for combining data (Centre for Reviews and Dissemination, 2009). Another important aspect of a systematic review is the literature search. The review author should perform comprehensive searches and ensure that they identify and include both published and unpublished studies. Finding unpublished studies is important because of two types of biases: submission bias and publication bias. The former relates to researchers being more motivated to submit positive results for publication. The latter relates to journal editors being more likely to publish studies with positive results (Muir Gray, 2001b).

Although systematic reviews can be done on any type of health care questions, the majority of reviews published so far address questions about the effect of different interventions. A main reason for the substantial amount of systematic reviews of effectiveness is the contribution of the Cochrane Collaboration, a non-for-profit organisation based in the United Kingdom. The

main purpose of the Cochrane Collaboration is to develop systematic reviews of health care interventions (Cochrane Collaboration). The Cochrane review is a sub-type of a systematic review. It follows the stringent demands of systematic reviews as described above and review authors must follow the procedures laid down in the Cochrane Collaboration Handbook (Higgins & Green, 2008). A unique feature of the Cochrane review is the updating process. All reviews are published electronically in successive issues of the Cochrane Database of Systematic Reviews, which is part of a larger database named The Cochrane Library. When new clinical studies are completed and reported existing reviews can be brought up-to-date. Authors of Cochrane reviews are responsible for updating their reviews (Higgins & Green).

Cochrane reviews can be lengthy documents; even health professionals frequently find them challenging to access and read (Rosenbaum, Glenton, & Cracknell, 2008). However, the collaboration has expanded its audience to lay users by creating synopses of the reviews (White, 2002). In addition all reviews now have a "plain language summary" written specifically for lay users to highlight information and research findings on the effectiveness of a treatment in a review.

The need for summarised research evidence to inform decision-making, both at an individual, local, and national level, has resulted in the establishment of the Norwegian Knowledge Centre for the Health Services. The mission of the Centre is to summarise and disseminate research evidence about the effect and quality of interventions within the health service as a whole (Norwegian Knowledge Centre for the Health Services, 2008). The Centre also hosts the Norwegian branch of the Nordic Cochrane Centre[1]. Researchers at the Centre have been actively involved in developing and promoting plain language summaries of Cochrane reviews (Santesso et al., 2007).

[1] http://www.cochrane.dk/

2.3 CRITISISMS OF EVIDENCE-BASED HEALTH CARE

The concept of evidence-based health care has been embraced by many, but has also been subject to a lot of criticism and debate (Elwyn & Edwards, 2001). Critics of evidence-based health care worry that it dictates a single "right" way to practice, which does not take the individual patient into account, and that it devalues the clinical experience and expertise of the health care worker (Tonelli, 2001).

Another common critique is related to the so-called "hierarchy of evidence", where systematic reviews of randomised controlled trials are ranked at the top and expert opinion and clinical expertise at the bottom (Mantzoukas, 2008). Randomised controlled trials are recognised as the study design best suited to avoid bias in questions about the effects of interventions (Straus et al., 2005). Some argue that while randomised controlled trials might be well suited when evaluating the effects of medications, it might be inappropriate, or even unethical, to use in other circumstances, such as in open heart surgery (Ashcroft, Hope, & Parker, 2001).

Cochrane reviews focus on questions about effect of interventions, and consequently randomised controlled trials are the preferred study design. A problem in Cochrane reviews and systematic reviews about treatment effects in general, is the lack of information about adverse effects (Glenton, Underland, Kho, Pennick, & Oxman, 2006). This reflects the limitations of randomised controlled trials to measure and report adverse effects. Because such trials are expensive and resource-demanding, the follow-up period is often too short to detect any side effects. However, methods for including nonrandomised studies to incorporate evidence of adverse effects are now being addressed by the Cochrane Collaboration (C. Glenton et al.).

As the two are closely related, criticism of evidence-based health care and the use of systematic reviews often go hand in hand. Miles and colleagues put it this way:

> "The EBM protagonists are dazzled scientists who set out to dazzle, rejoicing like acrobatic children vaulting through the statistical stratosphere, casting down meta-analyses and systematic reviews to clinicians below" (Miles, Bentley, Polychronis, & Grey, 1997, p. 84).

Nevertheless, the principles of evidence-based health care have gained worldwide acceptance. The British Medical Journal has nominated evidence-based medicine as one of the fifteen most important medical milestones since the journal was first published in 1840. In an article commenting the medical milestone, Dickersin and colleagues ask: «How could something so intuitively obvious to lay people not be similarly viewed by clinicians?» (Dickersin, Straus, & Bero, 2007). Their question relates to the fact that lay peoples' response to the evidence-based concept was usually one of surprise. They believed doctors and other health professionals had always been using evidence from research to inform their decisions (Muir Gray, 2001b).

2.4 EVIDENCE-INFORMED PATIENT CHOICE AND INFORMATION ACCESS

Evidence-informed patient choice assumes the presence of two requirements (Eysenbach & Diepgen, 2001):

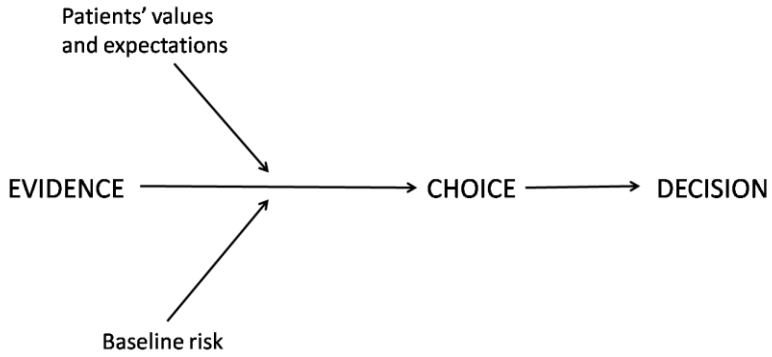
- The patient must have the power and opportunity to choose
- Objective, unbiased information must be available to the patient

The first requirement depends on both the individual patient and the clinician, and also how the health system is organized and structured. Some patients want to participate in the decision-making process, while others prefer to be passive and would avoid any information. Sometimes the clinician fails to empower patients and giving them the option to choose (Eysenbach & Diepgen).

The latter requires the patient having two types of basic information at hand: patient-related information and general information. Patient-related information is specific data and information that concerns the individual patient, such as information about his or her diagnosis, pathology, personal risk factors and more. General information relates to the external research evidence, for example the effect of different treatment options for a given disease (Eysenbach & Diepgen, 2001). For example, a decision on whether or not to undergo

a prenatal test, would involve information about personal health risk factors as well as external evidence about the potential harm and benefits of the test. Accordingly, the use of evidence to arrive to a final decision is influenced by the individual patient's values and preferences, as well as baseline risk factors. Figure 2, which is taken from Muir Gray (2001b) illustrates these relationships:

Figure 2: ***The determinants of an individual decision***



The word «informed» in evidence-informed patient choice indicates that research evidence is there to support, yet not dictate patient decisions. Many patients want experience-based information, that is, personal stories from other people having the same condition or being in the same situation, to supplement research-based information (Claire Glenton et al., 2006). This is not at least important for lay people with a chronic condition, where personal stories about coping strategies, can be just as important as research-based information about the effect of interventions.

The evidence-informed patient choice model emphasises *access* to research based information. In Norway, health professionals have access to a wealth of information resources

through the Norwegian electronic Health Library[1]. Many of the sources are also freely accessible to all Norwegian citizens. For example, the library provides national access to the Cochrane Library. Clinical Evidence is a meta-source that summarises the findings of different systematic reviews, making them more easily available to clinical practice. Clinical Evidence is also freely available nationally via the Norwegian electronic Health Library.

The National electronic Health Library models its English equivalent, the National Library for Health[2], previously known as the National electronic Library for Health (Turner, Fraser, Muir Gray, & Toth, 2002). One of the first moves of the English National Library of Health was to make the Cochrane Library available to the public. Clinicians were not happy about the idea at first, yet in 2009 the picture is changed:

> Initially this was greeted with concern by clinicians but clinicians now know that unless we provide high quality information, members of the public will come in with junk, and that may take much longer to discuss than a quality-assured source of evidence (Muir Gray, 2008, p. 96).

A timely question is of course whether lay people are actually interested in accessing sources that are primarily aimed at health professionals. Yet, when the National Library of Medicine opened free public access to Medline via Pubmed, the number of searches increased from 7 million in 1996 to 120 million in 1997. The new searches were primarily performed by lay people (Sieving, 1999). However, most lay people perform searches and access health related content using a general search engine rather than searching from identified medical websites (Renahy & Chauvin, 2006). Thus, the credibility of lay health information found via general search engines is an important issue to research.

[1] www.helsebiblioteket.no

[2] http://www.library.nhs.uk

3. THEORETICAL FRAMEWORK

3.1 DEFINING THE CREDIBILITY CONSTRUCT

The credibility of health information has been an issue for research since the early ages of the World Wide Web. However, researchers have used various terms for the concept, and until recently no theoretical framework to conduct research on the topic has been developed (O'Grady, 2006). Terms used across studies include “quality”, “trustworthiness” and “credibility”. As definitions and interpretations of these terms may vary, many researchers have called for a common and consistent terminology (Eysenbach et al., 2002; O'Grady, 2006; Provost, Koompalum, Dong, & Martin, 2006).

Terms like «quality» and «credibility» are theoretical *constructs* and cannot be measured directly, as can height and weight (Provost et al., 2006). Such theoretical constructs needs to be operationally defined. Thus, it is necessary to establish rules of correspondence between the construct itself, the observable behaviours of the website, and its content. Indicators such as «authorship» and «date of creation» are examples of the observable behaviour of a website. In an attempt to develop a standard for assessing the *quality* of health related websites, Provost and colleagues give the following definition for the quality construct for health websites:

> Quality of health-related websites: the levels of excellence which characterise the health website or eHealth content based on accepted standards of quality aiming to satisfy health information needs (Provost et al., p. 44).

What constitutes accepted quality standards of health websites or eHealth content? There are different *dimensions* of quality to consider such as accessibility, usability, and the accuracy of the specific information or content (Minervation, 2007). Accessibility concerns the overall access to the website. An example would be whether a website conforms to standards of web accessibility, ascertaining that the site is available to people with disabilities or with low-end

technology (Minervation, 2007). Usability concerns whether the website is designed and structured in such a way that the users can get what they want from it. The most well-known usability standard is probably the guidelines developed by Jacob Nielsen (Nielsen, 2000). Finally, accuracy relates to whether the presented information is up-to-date and in concordance with the best research evidence or with generally accepted medical practice (Eysenbach et al., 2002). Clearly, all these dimensions and their corresponding standards need to be taken into consideration when evaluating the overall quality of a health website. A health website providing accurate information is of no value if the user cannot access it, find it, or read it. On the other side, a health website which meets the appropriate reading levels of its target audience and is easy to navigate, could be misleading and potentially harmful if its content is inaccurate (Chalmers, 2001; Crocco, Villasis-Keever, & Jadad, 2002).

O'Grady (2006) has suggested a theoretical framework for future research on the credibility of health care websites in particular. I will use this framework to guide my study of Norwegian lay health websites. O'Gradys framework is based on the work of Fogg and colleagues at Stanford University; their studies of what strategies people use to evaluate online information in general (Fogg, 2003; Fogg et al., 2001; Stanford, Tauber, Fogg, & Marable, 2002). Although building on the concept of quality, Fogg et al. (2001) uses the term *credibility*. They simply define credibility as believability. Accordingly, credible information is believable information. Furthermore, they take the concept of quality into account by stating that credibility is *perceived quality*; it doesn't reside in an object, a person, or a piece of information.

Credibility perceptions result from evaluating multiple dimensions simultaneously (Fogg et al., 2001). These dimensions can be categorised into two key components: trustworthiness and expertise. The trustworthiness component captures the perceived goodness or morality of the source. This component is defined by terms like «well-intentioned», «truthful» and «unbiased». The experience component refers to the perceived knowledge and skill of the source. It is defined by terms such as «knowledgeable», «experienced» and «competent». People combine assessments of both trustworthiness and expertise to arrive at final credibility perception. A highly credible website will therefore be perceived to have high levels of both trustworthiness and expertise.

3.2 WHAT MAKES WEBSITES CREDIBLE?

Fogg and colleagues investigated specific website elements that could affect peoples' credibility perceptions (Fogg, 2003, p. 150). 1n 1999 and 2002, they conducted a web-based survey of more than 3000 web users from the United States and Finland. The survey questionnaire consisted of about 50 pre-defined items describing a website element. Study participants were asked to rate each element on a scale to show how that element would influence their perception of the credibility of the website's information.

Findings from the surveys suggest that within the trustworthiness dimension, listing an organisation's physical address and a contact phone number have a high impact on peoples' credibility impressions. Fogg (p. 156) denotes this as the principle of «real-world-feel». If people believe that there are real people behind the website, they are more likely to show confidence in the information provided. Another element that increases credibility perceptions is when a site discloses citations and references in the articles provided, and when it links to outside materials and sources. On the opposite, peoples' credibility perceptions decrease when a website does not distinguish advertisements from content properly, and when advertisements pop up in a new window (p. 158).

The principle of «real-world-feel» is also present within the expertise dimension. Quick responses to users' or customers' questions have high impact on credibility. Other important attributes related to expertise are listing authors' credentials for each article and regular updating of content (Fogg, p. 160). Moreover, peoples' credibility perceptions are also positively affected when a website looks professionally designed. Conversely, typographical errors, dead links, and outdated content influence credibility and peoples' impressions about a site's expertise in a negative way.

Based on his and fellow researchers' previous work, Fogg (2003, p. 163) separates between four types of credibility: presumed, surface, reputed, and earned. *Presumed credibility* has to do with initial assumptions about a source before it is even examined. These assumptions help people evaluate – rightly or wrongly – the credibility of websites. In a web environment, lay

people might assume that health websites provided by non-for-profit organisations are more credible than commercial information providers. Providing links to competitors' sites has also been found to increase peoples' belief in a website's content. Then the source is perceived to be honest and unbiased (p. 164).

Reputed credibility is earned by external reputation from other respectable sources that people find credible. For example, a website can gain credibility by being linked to by a well-known and respected organization, or by being recommended by a relative, friend, or colleague (Fogg 2003, p. 165-7). Another way of gaining reputed credibility is to use seals of approval. Seals of approval are third-party endorsements from companies to convey credibility. In the health area, the Health on the Net Foundation (HON) produces the oldest and perhaps best known seal of approval, which is called the HON-code (Health on the Net Foundation, 1997). Websites that display the HON-code logo (see Figure 3) are supposed to meet the credibility standards set by the foundation.

Figure 3: ***The HON-code***

Surface credibility involves investigating appearance features to determine credibility (Fogg 2003, p. 167). Surface credibility relates to the components of the website itself. One component is the website interface, such as links, structure, and visual design. The information, that is, the written text and accompanying illustrations, constitutes the other component of a website. These elements, and subsequently, the impressions of these elements by the lay individual, all relates to surface credibility. As stated earlier, people add credibility to a site that looks professionally designed. A site that makes it hard to distinguish advertisements from content may have the opposite effect. These two elements both relate to surface credibility.

Earned credibility is gained when the individual experiences the website as useful and reliable over time (Fogg, 2003, p. 170). Once established, earned credibility is probably the type of credibility that is most likely to change attitudes and behaviours. Elements that seem to establish an ongoing relationship between a user and a website is ease-of-use, personalisation, and responsiveness. If a site is arranged in a sense that makes sense to users it will gain credibility. Personalisation relates to the site's ability to recognise you and offer individualised services, such as selecting news stories according to your preferences. Moreover, responsiveness relates to the website provider's ability to quickly respond to users' questions.

The four credibility types are not mutually exclusive. For example, presumed credibility can also affect surface credibility. Surface credibility is partly based on judging the site's features, which in turn can be based on underlying assumptions on credibility (Fogg, 2003, p. 137). For instance, if a lay person assumes that commercial information providers are less credible than non-for-profit organizations (presumed credibility), then advertisements on a website might serve as an indication that its content is less credible (surface credibility).

O'Grady (2006) builds on the four credibility types when modelling a theoretical framework for health care websites. In the next section I describe the framework and discuss its' implications for evaluating lay health websites.

3.3 CREDIBILITY FRAMEWORK FOR HEALTH CARE WEBSITES

To maintain consistency, O'Grady (2006) also uses the term «credibility» when transferring Fogg's taxonomy to health care websites. Still, she acknowledges that terms like «quality» and «trustworthy» could be used in similar contexts. Furthermore, she introduces two facets that can be associated with the credibility types: (1) the *individual* facet, which relates to those using the website; and (2) the *web-based* facet, which relates to the website itself. The web-based facet includes the creators of a website, in other words the site's developers and designers, as well as the information on the site.

Presumed and earned credibility are associated with the individual facet only. Automatic belief in an information source (presumed) and re-use of a source based on positive experience over time (earned) is based entirely within the consumer. Surface and reputed credibility are associated with the website at first, and thereafter with the individual when it is viewed or used. Moreover, O'Grady (2006) emphasises that those who want to investigate web credibility in health must bear in mind that web-based credibility is a separate issue from individual-based credibility. By making this distinction, she denotes an important issue not explicitly discussed by Fogg (2003). Yet, it is based on findings from a study which builds on his work, and where Fogg is one of the co-authors (Stanford et al., 2002). In this study the researchers compared how experts and lay people evaluated the same health and finance sites. They found that experts were concerned with the content of a site, such as providing references and author's credentials for each article, and whether the site was provided by a well-respected organisation. Although listing author's credentials and references on a site added credibility for lay people as well, they relied on design features and visual appeal for much of their credibility appraisal. However, as pointed out by both Stanford et al. (2002) and O'Grady (2006), a professional-looking site is not a primary marker of credibility. This indicates that the credibility of a health care website also relates to other elements than those perceived by the lay individual.

A main purpose of the work undertaken by Fogg et al. (2001) was to gain more knowledge on how to design for credibility. This probably explains the individual, or subjective, approach to credibility. To be able to design for credibility, Web designers need to know how web users evaluate credibility. In fact, Fogg separates between two credibility goals; those of the web surfers and those of the Web designers:

> Web surfers need to evaluate what information is credible. This issue of «information quality» has been embraced by librarians and teachers. To better assess the credibility of an online source, librarians and others advocate that web surfers examine who the author is, how timely the content is, and how the content compares with similar content from trusted sources, such as experts in the field. (...) The other aspect of credibility relates primarily to website designers. The main issue for designers is how to create websites that convey appropriate levels of credibility (Fogg, 2003, p. 149-50).

Fogg indicates that there are elements of credibility related to the information on a website that users may not know about unless instructed by someone else, for example a librarian. O'Grady (2006) and Stanford et al. (2002) also emphasize the educational aspect: lay people need to be educated about web credibility. For the two web-based credibility types (surface and reputed) imply that, in order to be perceived credible, a website must provide credible information and in turn the individual must learn what constitutes relevant credibility criteria. O'Grady exemplifies this by stating that if providing citations and references is an indicator of surface credibility for information, then a lay person must know that this is an important credibility criterion. If the HON-code indicates reputed credibility of a website, then a lay person must acknowledge the HON-code as a marker for credibility.

Within the web credibility framework, Fogg introduces three categories which describe a website: the operator of the site, the content of the site, and the site's design (Fogg, 2003 p. 173). These categories are not considered by O'Grady, but are useful in distinguishing between the many issues relating to Web credibility. The *website operator* is the organisation or person offering the site. The *website content* is by Fogg defined to be what the site provides its' users in terms of information and functionality. Information comprises text, images and sounds that have meaning for users. Functionality relates to the work the site can do for the user, such as translating from English to Norwegian, or calculating the body mass index. The *website design* concerns how the site is put together. It relates to how the information is structured on each page and throughout the site (information design); to technical issues such as search facilities (technical design); to aesthetical features such as colours (aesthetic design); and to the interaction between the user and the site (interaction design). Please see Table 1 for an overview of the theoretical framework and some examples of evaluation criteria within each credibility type and category.

Table 1: ***Theoretical framework for evaluating Web credibility***

	Presumed credibility *Individual-based*	**Reputed credibility** *Web- and individual-based*	**Surface credibility** *Web- and individual-based*	**Earned credibility** *Individual-based*
website operator	The operator is a nonprofit organisation	The person writing the web article is a recognised expert	The site displays their content policy	The operator always sends quick answers to site users' questions
website content	The site has ads from reputable companies	The content has been endorsed by Health on the Net Foundation	Citations and references are provided The author's credentials are listed	Pages are tailored to individual users The site's content has always been accurate and unbiased
website design	The site has a search feature on the top of page	The search feature is powered by Google or another respected search engine	The site looks professionally designed	The site is easy to navigate

Accuracy is another dimension of credibility and is related to the specific information on a website. In fact, when lay people make credibility judgements they are concerned about the accuracy of the information provided (Stanford et al., 2002). However, their evaluations are often based on elements outside the actual information, such as visual appeal, because they do not always have the expertise to evaluate accuracy:

> Many researchers have conducted studies on consumers to understand various aspects of website credibility. However, some consumers may not be well equipped to make informed decisions about the accuracy of information in technical fields such as health or finance. (...) Given the various results of previous studies of consumer credibility assessments, we are left wondering whether consumers' credibility evaluations of websites are correct (p. 3).

Fogg (2003, p. 139) defines two types of errors in peoples' credibility evaluations: gullibility and incredulity People commit the gullibility error when they perceive a website to provide accurate information, even though it is not. On the opposite, people commit the incredulity error when they reject information from a website, even if it is accurate. Once more he mentions the librarians' role in educating users to avoid these errors. Librarians are most

concerned with the gullibility error. They teach information seekers to look for other credibility cues to determine information accuracy, such as authors' credentials, regular updating and the authority of the organisation. These cues, or credibility criteria, can all be said to belong to the reputed or surface credibility types. Moreover, the criteria are mostly concerned with issues related to the operator or content categories.

Another way of evaluating accuracy is to compare the information to some type of standard, by Fogg denoted as "trusted sources" (Fogg, 2003, p. 150). He suggests experts in the field as one possible source that can be trusted. Within the evidence-based practice model, systematic reviews of the appropriate kind of research evidence would be preferred to experts as the best source of comparison (Badgett & Lawrence, 1998; Coulter et al., 2006; Wyatt, 1997). Accordingly, if the information complies with the set standard then it will be accurate. Unless you are an expert, or always abreast of the summarised research findings on a particular topic, accuracy is not something that can be easily evaluated or spotted by the individual. Consequently, accuracy relates to the web-based facet of credibility.

In summary, the theoretical framework consists of four types of credibility: presumed, reputed, surface, and earned. Each of these types can be associated with two facets: those using the website (individual-based) or the website itself (web-based). In health care websites there are various criteria for each of the four credibility types. O'Grady (2006) points out that it is important to distinguish between what criteria lay individuals use when evaluating credibility and what field experts or opinion leaders know to be credible. Consequently, researchers must be clear about what facet of Web credibility they are investigating. Either way, the criteria could be classified into whether they are related to the operator, the content, or the design of a website.

The focus of my study is the web-based facet, rather than the individual-based of credibility. I will be focusing on the accuracy of the information about the effects of interventions across various Norwegian lay websites. I define will also refer to accuracy as *information credibility* to separate it from the other credibility types. In the next chapter I review some of the literature related to health information credibility on the World Wide Web.

4. PREVIOUS RESEARCH

4.1 IDENTIFYING RELEVANT LITERATURE

I performed an initial scope search in Medline (Ovid interface) using the free-text terms «quality», «credibility» and «internet» and «world wide web». This search identified a large systematic review of 79 empirical studies that evaluated the credibility of lay health information on the World Wide Web (Eysenbach et al., 2002). The review authors searched multiple databases, Web search engines, and contacted experts in the area for identification of any unpublished studies. Due to the extensive work done by the authors, I decided to limit the final database searches to studies and reviews published after 2002. The search strategies were broad, but limited to articles in English and the Scandinavian languages. Please find the search strategy for Medline in Appendix 1. The Medline strategy was modified to the databases Embase, Cinahl, PsycInfo, LISTA and Svemed+. I performed the searches in December 2007.

The number of unique hits exceeded 5000. After screening titles and abstracts, 410 titles remained that were somehow related the broad topic of lay health information on the World Wide Web. Some of the titles referred to studies of lay peoples' evaluations and use of health information on the Web. Since my focus was the web-based facet of Web credibility, I did not include them in my literature review. However, some of them were used to inform the methods of my study and will be referred to in Chapter 5.

I made no attempt to summarise all the literature published about information credibility in the recent years. Rather, I chose a pragmatic approach by using systematic reviews primarily. I included primary studies where no reviews were available. I have also added a few relevant studies published after literature search was conducted. The various reviews and studies used the terms «quality», «trustworthiness» and «credibility» in similar contexts. In order to be accurate to the cited papers, I use the terms as they appear in the publication when that particular work is discussed herein.

4.2 EVALUATIONS OF THE CREDIBILITY OF ONLINE HEALTH INFORMATION

4.2.1 THE CRITERION STANDARD APPRAOCH

Currently there is no internationally accepted quality standard of health information websites (Provost et al., 2006). However, numerous efforts have been undertaken to measure the credibility of lay health information on the World Wide Web. Although various methods have been used there are two main approaches apparent in the published literature: (1) to compare the information on the website against some criterion standard; and (2) to rate various aspects of the website using an evaluation instrument. The first approach could be a part of the latter, but not necessarily.

The criterion standard approach is used to assess the accuracy, or information credibility, of health content. As stated in Section 3.1, accuracy could be defined as the degree of concordance of the provided information with the best research evidence or with generally accepted medical practice (Eysenbach et al., 2002). The criterion standard refers to the type of evidence used to measure this «degree of concordance». Proponents of the evidence-based practice model argue that the information on a health care website is accurate if it complies with the conclusions from up-to-date systematic reviews where the authors have summarised the appropriate kind of research evidence (Badgett & Lawrence, 1998; Wyatt, 1997). For information about treatments, systematic reviews based on well-conducted randomised controlled trials are regarded to provide the best evidence (Straus et al., 2005, *see section 2.3*). For other core aspects of health care, such as risk factors, diagnosis, and patients' experiences, other types of studies should be systematically evaluated and summarised.

If «best research evidence» is defined as a systematic review, or a clinical guideline based on systematic reviews, then the review or the guideline would constitute the criterion standard. When using the criterion standard approach, the statements on different websites on a particular topic could be compared to the conclusions of a systematic review on the same topic. For example, if a systematic review concludes that there is no clear evidence that echinacea (an herb) cures common cold, then websites with information on echniacea and common cold could be assessed by whether they comply with the conclusions of the review.

Researchers have used different types of criterion standards when evaluating the accuracy of lay online health information. In their review, Eysenbach et al. (2002) found that the type of standard used influenced the overall conclusions of a study. The studies using systematic reviews, or guidelines based on systematic reviews, reported a higher proportion of inaccurate websites than did those using other types of evidence, such as single studies, textbooks, and expert consensus. Furthermore, studies not using a criterion standard, or where the researcher used his or her personal opinion to assess accuracy, came to more positive conclusions with regard to the quality of information across websites. Nevertheless, a majority of the included studies (70%) found quality to be a problem on the World Wide Web. Powell et al. (2005) undertook a follow-up review of the literature on lay health information on the Web. They identified more than 160 studies which described evaluations of the quality of information found on different health topics. Several of the studies were conducted after the publication of the systematic review by Eysenbach et al. (2002). In fact, online information has been studied so frequently that this field has been given a name, "information epidemiology" or «infodemiology» (Eysenbach, 2002).

The more recent studies also use a variety of methods for evaluating the accuracy of lay information on the web. Some authors have used a criterion standard; others have not. In addition, the types of evidence used as a criterion standard varies. In a study of thyroid cancer, the authors compared online information to expert consensus and practice guidelines (Air et al., 2007). They found that the completeness of information was a bigger problem than its accuracy. Completeness relates to whether all treatment options for a particular condition are described, including their benefits and possible adverse effects. The authors do not specify the development process behind the guidelines used as a criterion standard, for example if they were expert based or based on systematically summarised research evidence. This might have influenced the accuracy scores.

A study of Canadian websites on female urinary continence used clinical guidelines as a criterion standard and found accuracy to be generally low. The guideline recommendations were based on systematically reviewed research evidence (Farrell et al., 2006). Coulter et al. (2006) came to similar conclusions when they compared online and printed patient information on four different health topics to the results of systematic reviews. The topics included arthritis, chronic obstructive pulmonary disease (COPD), healthy eating and

measles, mumps and rubella (MMR). They selected materials developed by commercial, public, and non-for-profit organisations in the United Kingdom. There were small differences in accuracy scores between the four topics; accordingly the authors conclude that it is not possible to tell whether information on one topic is better or worse than the information available on another. This contradicts the conclusions of the review by Eysenbach et al. (2002), where studies of diet and nutrition websites revealed high proportions of inaccurate materials compared to studies of online cancer information.

Although several studies of the accuracy of online health information have been published internationally, there are few such evaluations of Scandinavian language websites. My literature search only identified only two studies where Web materials were evaluated using a criterion standard approach. One of them was a study of Swedish breast cancer websites (Nilsson-Ihrfelt et al., 2004). The majority (70%) of the included websites were targeted at patients. Their findings revealed that information about risk factors and alternative therapies was frequently inaccurate and doubtful, while the information about conventional therapies such as chemotherapy and radiation treatment was highly accurate. However, the authors are quite unspecific about the type of criterion standard used, they refer to «scientific knowledge and current medical practice» (p. 379).

Jørgensen & Gøtzsche (2004) undertook a study on websites about mammography screening, including Norwegian language websites. They compared screening information provided by public and non-for-profit organisations to the findings of a Cochrane review. The Cochrane review question whether screening does more harm than good due to overdiagnosis and overtreatment. However, the information identified via the Web was often biased in favour of mammography. The researchers also included websites from the other Scandinavian countries and English-language sites, and they did not perform a separate analysis on the Norwegian-language Web materials. Because only 27 sites were included in total, it is difficult to generalise the results to a Norwegian population of websites.

One could argue that with no domain knowledge it is difficult or even impossible to assess the accuracy of the information on a website. Moreover, assessing accuracy using a criterion standard is time consuming and may not be feasible in cases where websites provide information on a broader range of topics (Breckons, Jones, Morris, & Richardson, 2008). In

order to meet these challenges evaluation instruments have been developed to help lay people make credibility judgments and thereby predict accuracy. In the next section I describe some of the research on evaluation instruments and their usage to evaluate health information credibility.

4.2.2 THE EVALUATION INSTRUMENT APPROACH

An evaluation instrument is composed of one or more criteria. Each criterion consists of one or more elements, which are items of information that must be assessed in order to evaluate compliance with the criterion (Bernstam, Sagaram, Walji, Johnson, & Meric-Bernstam, 2005). For example the instrument commonly known as the «EU quality criteria for health related websites» (Commission of the European Communities, 2002) consists of six criteria. One of them are «Updating of information», which is defined as:

> Clear and regular updating of the site, with date of up-date clearly displayed for each page and/or item as relevant. Regular checking of relevance of information (p. 6).

To assess whether a given website complies with the EU-criteria, a user has to determine whether the site complies with «updating of information». This requires answering at least the following questions: (1) «does the site clearly display date of up-date for each page and/or item as relevant?» and (2) «is the relevance of information regularly checked?» Therefore, to evaluate the single criterion «updating of information», a user would have to assess two elements.

As pointed out in Chapter 3, the credibility or quality of online information in general can be evaluated along several dimensions, and the evaluation criteria will differ accordingly. Eysenbach et al. (2002, p. 2694-5) identified several criteria used across studies, and divided them into five categories:

5. *Technical criteria*, which are criteria related to the transparency of a website. Examples include authorship criteria (e.g. disclosure of authorship); currency criteria (e.g. provision

of an creation date), and attribution criteria (e.g. the sources used to compile the information).

6. *Design criteria*, which are subjective site design features and layout criteria that relate to the visual appeal and navigation of a website.
7. *Readability criteria*, which are criteria related to the complexity of the text, such as length of words and sentences. The use, or non-use, of medical jargon is a facet of readability.
8. *Comprehensiveness*, which refers to the information provided being complete (e.g. state all treatment options for a given condition); or balanced (cover both positive and harmful effects of a treatment).
9. *Accuracy*, which by the authors is defined as agreement with the best research evidence or with generally accepted medical practice. The comparison of web materials with a criterion standard belongs to this category.

The technical criteria have been the most commonly used among the five categories, although quite a few studies have also been concerned with website design and navigation. Both technical and design criteria relates to surface credibility because they involve evaluating appearance features such as whether the author's credentials are listed or not; or whether the same page layout is used throughout the site.

During the years, numerous evaluation instruments that comprise technical and design criteria have been developed. The first systematic review of evaluation instruments was published as early as 1998 (Jadad & Gagliardi, 1998). The authors identified 47 instruments. They found that only one-third of the instruments offered a description of the criteria used, and very few provided instructions for their use. Many of the reviewed instruments were concerned about «quality» in general. Yet, many of the criteria used were related to design. This allowed websites to receive a high quality rating based solely on visual appeal and ease-of-use. This was criticised by the review authors, and have also been criticised by others (Badgett et al., 1998). When the authors updated their review four years later, they identified 98 instruments

(Gagliardi & Jadad, 2002). Only six of the instruments previously available still existed, and they found little change with regard to the limitations identified in the first review.

Instruments should be easy to use for the intended audience. In a subsequent systematic Bernstam et al. (Bernstam, Shelton, Walji, & Meric-Bernstam, 2005) defined ten elements to be the most that a motivated lay person is likely to be able to practically assess. The authors identified more than 250 different instruments. One-fifth of the instruments comprised ten of fewer criteria, and one-tenth had ten or fewer elements. This indicates that many instruments are comprehensive and unlikely to be utilised by lay people. In addition, many instruments seem to have a limited life span. In a recently published review on evaluation instruments, only three of the instruments identified by Bernstam, Shelton, et al. (2005) still existed (Breckons et al., 2008).

The purpose of an evaluation instrument is to allow those looking for health information to filter out inaccurate and potentially misleading content. The criteria need to be unambiguous so that different people using the same criteria will agree on their application and use similar ratings for the same material (Bernstam, Sagaram et al., 2005). In other words, the criteria need to be *valid* and *reliable*. Validity means that an instrument measures what it sets out to measure. An instrument developed to evaluate the accuracy of health information on a website is valid if the websites that meet with its inherent criteria actually contain accurate content. The instrument criteria are reliable if they can be consistently repeated to produce the same results when employed by different people (Twycross & Shields, 2004a, 2004b).

A common critique in all the reviews cited above has been that none of the instrument developers have evaluated them properly with regard to the reliability and validity of the measurements. Quite a few evaluation instruments contain criteria related to the usability of a website, including readability. Usability elements are considered to be rather subjective and context-dependent. Although end-users often emphasise such elements when evaluating the credibility of a website (Eysenbach & Köhler, 2002; Fogg, 2003), researchers agree that to evaluate them in a valid and consistent way is even harder than evaluating the information itself (Bernstam, Sagaram et al., 2005; Breckons et al., 2008; Gagliardi & Jadad, 2002; Jadad & Gagliardi, 1998).

Even technical criteria are open to interpretation and should be properly validated. In their review, Eysenbach et al. (2002) found that studies varied considerably in how they defined the technical criteria. They concluded that further operationalisation was needed to: (1) ensure consistency when evaluating online health information; and (2) to make meaningful comparisons across studies. A few years later, Bernstam, Sagaram et al. (2005) attempted to operationalise most of the technical criteria defined in the review. They measured their reliability by estimating the level of agreement between two raters, when using the criteria across a sample of forty-two alternative medicine websites. This is referred to as interobserver reliability (Bernstam, Sagaram et al.). Although the operational definitions improved interobserver agreement for some technical criteria, agreement remained low for other seemingly objective criteria, such as «disclosure of sponsorship. Furthermore, the authors found that agreement was lower for criteria that required review of the entire website, rather than a single webpage.

Evaluation instruments are usually generic, so that they can be applied to websites providing a wide range of health information covering a variety of different conditions (Harland & Bath, 2007). As previously described, the generic, technical criteria are the most frequently used across studies that have evaluated online health information. Such criteria are often used by librarians to educate end-users to find credible health information (Childs, 2004; Fogg, 2003, p. 149). Librarians also use them to select and recommend sources to lay people, one example is the Scandinavian health portal SMIL[1] (Scandinavian Medical Information for Laypersons). SMIL is developed by librarians from Nordic university and hospital libraries. They use the EU-criteria to guide the selection and inclusion of sources into the portal (Munthe & Nordnes, 2004).

Health portals like SMIL have been proposed by those responsible for them, as a solution to the quality problem on World Wide Web. Health portals provide third-party evaluation of health-related websites; the websites are selected according to defined criteria. In a study of four large English-language government-run health portals, researchers compared information on the effects of treatments for various health conditions to the results of Cochrane reviews (Glenton, Paulsen, & Oxman, 2005). They revealed that the information

[1] http://smil.uio.no/

available was often unclear, incomplete and misleading. The portal developers used generic, technical criteria when selecting sources. Some researchers have criticised the generic criteria because they are structural or process measures that only act as proxies for the accuracy of the information provided (Seidman, D., Steinwachs, & Rubin, 2003a). Seidman et al. (2003a) use the example of diabetes. A structural accuracy measure could be to assess whether the author is a physician with additional training or specialisation in endocrinology or diabetes. Process measures relate to how the information is developed; whether the website providers follow procedures which are likely to result in accurate information. An example of a process measure could be peer review. Peer review means that the content is reviewed by independent experts.

Seidman et al. (2003a) argue that structural or proxy measures do not tell a user whether the actual content is accurate or comprehensive. They proposed a disease specific instrument that took into account both accuracy and comprehensiveness of online diabetes information (Seidman, Steinwachs, & Rubin, 2003b). It was based on a review evidence-based guidelines and expert opinion. When testing the instrument on a sample of English-language diabetes websites, the authors found a wide variability in accuracy and comprehensiveness of information. Although inter-observer agreement among raters was moderate to high, the instrument was not tested by lay people with diabetes. The instrument included more than 40 elements, which might be too comprehensive for actual users to employ. Furthermore, disease-specific instruments needs to be continuously updated to reflect recent findings from research.

The British Library has developed the DISCERN guidelines, an instrument aimed at evaluating treatment information in particular (Shepperd, Charnock, & Cook, 2002). The instrument comprises 15 criteria related to transparency and comprehensiveness of websites. It has been tested for its validity and reliability among health professionals as well as lay people (K. M. Griffiths & Christensen, 2002; Kathleen M. Griffiths & Christensen, 2005). Findings from the studies suggested a strong association between the health professionals' DISCERN ratings and the accuracy of content. The association was moderately high for the lay ratings. Accuracy was compared to evidence-based guidelines.

A few studies have investigated whether the technical criteria can be used for distinguishing between accurate and inaccurate information. England & Nicholl (2004) tested a set of criteria related to authorship, currency and attribution on lay websites about coealic disease. Clincial guidelines were used as a criterion standard. The authors found no significant association between the technical criteria and accuracy. Bernstam et al. (2008) came to similar conclusions in a study of breast cancer websites. The used the technical criteria defined by Eysenbach et al. (2002) thay were validated in a previous study (Bernstam, Sagaram et al., 2005).

As previously described, synthesising research evidence and developing systematic reviews has increasingly become acknowledged as a prerequisite to producing credible health information. Some of the instruments available include criteria that reflect these issues (Minervation, 2007). Their intention is to evaluate if a website operator has taken the necessary steps to find, assess and use scientific evidence as a basis for their content. Coulter et al. (2006) investigated 31 different information providers in the United Kingdom. One of the criteria used was «reviews the clinical research evidence and use systematic reviews where available» (Coulter et al., p. 37). Their findings revealed that almost all providers considered it important to produce evidence-based patient resources. However, two-thirds of them did not provide content authors with guidelines for identifying and appraising available research evidence. It was felt unnecessary, or even inappropriate, to provide authors with guidelines because they were thought to be familiar with the evidence in their respective fields. However, when comparing their content to the results of systematic reviews, the researchers identified several examples of inaccurate information.

I did not find many Norwegian studies where lay health webpages have been evaluated using an instrument. Norem & Moen (2004) evaluated 54 Norwegian hospital cancer websites using a set of 16 general criteria for public web services. Most of the criteria related to how well the websites provided information about the hospital services, such as waiting times, car parking, addresses and maps. One criterion specifically addressed information on treatments available. The item focused on comprehensiveness rather than accuracy, by scoring sites to their listing of different cancer therapies. In general, the hospitals gave very limited information about the treatments available. The university hospitals scored better than other

types of hospitals in this respect. Moreover, more than half the sites gave no details on their date of creation or update.

Eiring & Tvedten (Eiring & Tvedten, 2007) evaluated the quality of Norwegian lay mental health sites by using the LIDA instrument (Minervation, 2007). The LIDA instrument measures different quality dimensions: accessibility, usability and reliability. The authors pre-selected 13 non-commercial Norwegian-language websites and included an additional ten sites by searching Google. The quality scores varied across sites, with the non-commercial sites generally having higher scores than the commercial sites. However, the two websites having the highest reliability score were targeted at health professionals rather than lay people. The LIDA instrument has a separate section on content development (Minervation, 2007, p. 9). It includes criteria about literature searching and critical appraisal of research evidence. As these criteria are supplemental, it is not clear whether they were used in by the authors.

4.3 SUMMARY AND RESEARCH QUESTIONS

Various studies have evaluated the web-based credibility of lay health information. The two most commonly used methods are the criterion standard approach and the evaluation instrument approach. The criterion standard approach is used to evaluate information accuracy solely. The studies that have used a criterion standard vary with regard to their methodological rigour: studies that have used systematic reviews have revealed larger proportions of inaccurate webpages than studies using single studies, textbooks, or personal opinion.

Evaluation instruments contain criteria to assess different dimensions and elements of a website. Many of them comprise generic, technical criteria that act as proxy measures for information accuracy. Such criteria are often used by health portal providers. They are also used by librarians to educate lay people on how to find credible information on the Web. Despite numerous efforts to develop instruments, many of them have not been properly operationalised and validated.

There are few studies that have evaluated the accuracy of infor ation Norwegian-language websites. Only one small study used a criterion standard to determine information accuracy, however in this study the authors also included websites in other languages and no separate analysis was performed on the Norwegian websites (Jørgensen & Gøtzsche, 2004).

Based on the principles of evidence-based informed choice, the credibility framework for health web sites, and findings from the previous literature, my specific research questions are:

10. How accurate is lay health information about the effect of interventions on Norwegian websites when compared to the results of systematic reviews? (information credibility)
11. Can the technical criteria be used to evaluate information accuracy? (surface credibility)
12. Is lay information identified via the health portal SMIL more accurate compared to information found using general search engines?
13. Is information on public agencies sites more accurate than information on commercial sites?
14. What procedures are in place to ensure that research evidence is included when developing content, and what sources of evidence are used?

5. METHODS

5.1 STUDY DESIGN

I used a cross-sectional design where information from a sample of webpages within two specific topics was captured at one moment in time. Cross-sectional studies are also referred to as descriptive surveys (R. Powell & Connaway, 2004, p. 87). The purpose of a cross-sectional study is to describe characteristics of a population of interest, to estimate proportions in the populations, and to identify if any associational relationships exist between two or more variables. Such studies are useful in identifying current situations and conditions, in my case the state-of-the-art of health information for lay people on the World Wide Web.

In my study the population of interest was Norwegian webpages with lay health information about the effect of two specific interventions. My purpose was to describe elements of surface credibility by employing a set of commonly used technical criteria. I estimated the proportion of webpages that complied with each criterion. Furthermore, I wanted to check credibility in terms of accuracy by comparing the information to a criterion standard. My criterion standard consisted of two systematic reviews. I estimated the proportion of webpages that complied with the conclusions of the reviews. I also wanted to explore if there were any associations between compliance with the technical credibility criteria and accuracy. In other words, I wanted to see if technical criteria could be used to identify accurate health information. Finally, I wanted to explore if there was an association between accuracy and (1) webpages indexed via a health portal (SMIL); and (2) webpages provided by public agencies.

In the next sections I describe in more detail how I selected the two systematic reviews that comprised the criterion standard, how I searched for and selected webpages, and what technical criteria I used to evaluate surface credibility.

5.2 THE CRITERION STANDARD

5.2.1 SELECTING THE SYSTEMATIC REVIEWS

I decided to use systematic reviews developed by the Cochrane Collaboration because the objective of my study was to investigate health information about the effects of interventions. As described in Section 2.2, Cochrane reviews address intervention effectiveness. Furthermore, they follow stringent development procedures and are continuously updated. However, as updating is sometimes delayed, I only considered reviews published in the last two years. Although there is little research on when to update a systematic review, the Collaboration recommends updating every two years (Moher et al., 2008).

Issues related to a healthy lifestyle are a popular topic across different lay populations when it comes to World Wide Web searching, with 40% of health seekers searching for leading topics such as nutrition or dietics (Renahy & Chauvin, 2006). The sub-topic of healthy eating is subject to a lot of media interest (Monsen, 2007). There is an ongoing debate between different experts on what dietary advice should be given to the general population as such, and to people with overweight and obesity (Dommerud, 2003; Poleszynski & Mysterud, 2008). In the recent years the concept of «glycemic index» has become a term on almost everyone's lips, not at least because of the Greek-Norwegian internist and nutritional expert Fedon Lindberg[1]. Glycemic index is a measure of the effects of carbohydrates on blood glucose levels. Foods containing a low glycemic index or load, such as lentils, release glucose gradually into the bloodstream, and thereby stimulate less insulin release and reduce blood lipids (Jenkins et al., 1981). Low glycemic foods could play a positive role in appetite and body weight regulation, yet researchers disagree on this issue (Raben, 2002). In 2007 a Cochrane review on the effect of low glycemic diets for weight reduction in overweight and obesity was published (Thomas, Elliott, & Baur, 2007). I chose the review as one of two systematic reviews for my criterion standard. I will refer to low glycemic index and low glycemic load diets as «low glycemic diets» from now on.

[1] www.drlindbergs.no

Being a woman, at younger age, and to have frequent contact with health professionals are characteristics found to be associated with more regular use of online health information (Andreassen et al., 2006; Renahy & Chauvin, 2006). Thus, I chose pregnancy as the second broad topic for my study as being pregnant encapsulates all the characteristics above in most cases. The growing number of Norwegian websites and online forums targeted at pregnant women also indicates that seeking out information online is a popular activity in pregnancy.

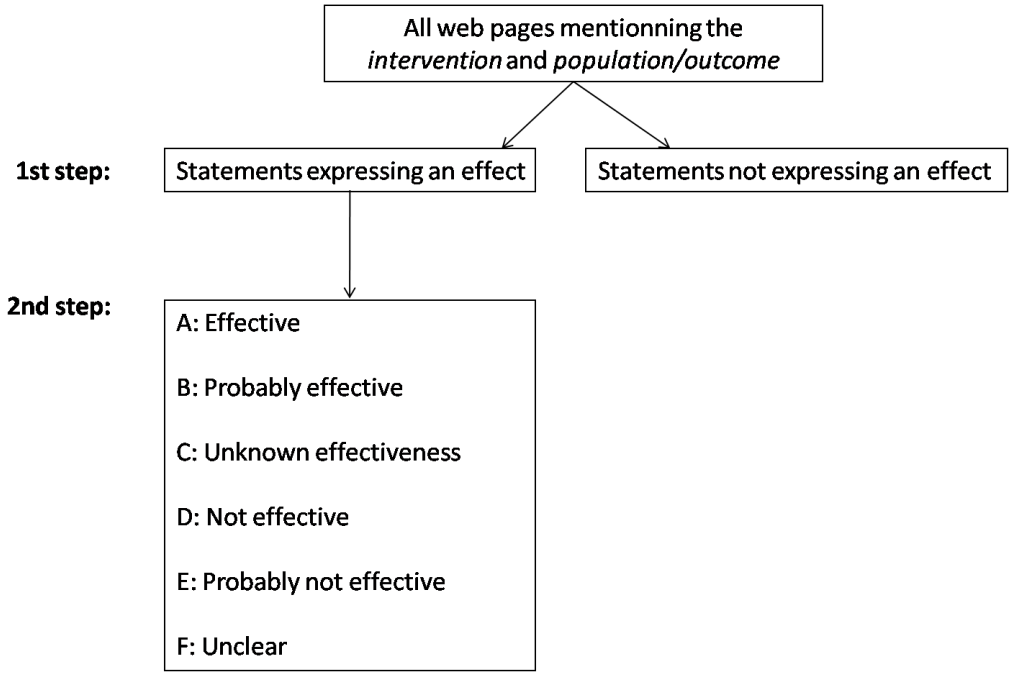
To identify a relevant review, I performed a broad search on pregnancy in the Cochrane Library using different search strategies (see Appendix 2 for more details). A pregnant woman volunteered to examine the list of potentially relevant review titles (n= 95), which resulted in a list of 16 relevant reviews. After discussion, we chose a review on calcium supplementation during pregnancy for preventing high blood pressure (hypertension), pre-eclampsia, and other hypertensive problems related to high blood pressure in pregnancy (Hofmeyr, Atallah, & Duley, 2006). Pregnant women are generally interested in preventive measures to take care of their own health and the child's. If calcium supplementation is a means for preventing high blood pressure and pre-eclampsia it would be interesting to see how this is reflected in the advice on calcium intake found on Norwegian lay websites.

5.2.2 COMPARING WEBPAGES TO THE CRITERION STANDARD

To compare the information on the included webpages to the criterion standard, I used a coding system published and tested in a previous study (Glenton et al., 2005). In the study the authors compared online lay information about the effects of eight different interventions to the results of Cochrane Reviews. The study authors chose conditions to cover a variation in age, sex, mental/physical health, and chronic/acute illness. I decided to use their coding system because it was directly relevant to my own study purpose. Furthermore, their findings could constitute a frame of reference to my own results and assessments. From now on I refer to information on a webpage about the effects of the included interventions (low glycemic diets or calcium supplementation) as *Web statements*.

As shown in Figure 4 the coding system comprised a two-step process. The first step was to decide whether a web statement actually expressed an effect or not. The second step was to evaluate the degree of effect for those statements expressing an effect, which the authors defined as «effective», «probably effective», «unknown effectiveness», «probably no effect», and «no effect». Moreover, a statement was categorised as «unclear» if it expressed an effect, but it was not possible to classify the degree of the effect.

Figure 4: ***Coding the Web statements***



Glenton et al. (2005) asked a panel of 16 people to use the coding system to categorise a selection of Web statements about treatment effects. Based on their input the authors agreed on a final coding list. Statements were categorised as not expressing an effect where an intervention, for example a drug, was mentioned as being used for a particular condition, but without any specification of effect. For example, statements such as «the treatment can be used» or «treatment may include» were classified as not expressing an effect.

For statements categorised as expressing an effect, Glenton et al. (2005) provide an extensive list of examples based on their final coding. There was frequently little agreement among the panel participators in Glentons' study. Statements including the English verbal auxiliary «may» were reported to be particularly confusing. As a result, such statements were coded as «unclear». I give an overview of the graded effect categories along with a selection of their generic examples in Table 2. None of the statements were found to indicate a treatment being «not effective» or «probably not effective», and consequently they are not included in the table.

Table 2: Some examples of generic effect statements[a]

Code	Examples
Statements coded as «effective»	Has proven effective/helpful/beneficial Works/(usually) works well X was more effective than Y Many studies have shown the treatment to be useful/can ease the symptoms These treatment have been beneficial to many people/have helped many patients Is recommended/should be recommended Helps some people Of some value in certain patients
Statements coded as «probably effective»	There have been reports of some benefits Appears to be helpful Results/evidence suggests that the treatment is beneficial/has beneficial effects
Statements coded as «unknown effect»	Have not been proven Results are variable Some studies suggest this is effective..other studies have found that it doesn't help Effectiveness for many people is unknown The evidence is uncertain
Statements coded as unclear	Not all people taking these drugs benefit from them The treatment is not always successful Does the treatment work? Not always The treatment is of not more benefit than the other treatments May/might help/relieve pain/improve symptoms/relieve symptoms/be a good option Can help/improve symptoms/relieve symptoms Is said to improve symptoms

a) As defined by Glenton et al. (2005)

Due to the reported difficulties in consistently coding the Web statements in the study by Glenton et al. (2005), I considered it important to involve other people in the coding process.

In addition I coded the statements twice, with a few days interval, to see whether my own assessments varied substantially.

Five colleagues (four nurses and one physiotherapist) volunteered to categorise the Web statements using the coding system. Four and three raters independently coded the Web statements on low glycemic diets and calcium supplementation respectively. Where agreement was low, I involved another two raters. I kept the codes in their original language (English), yet I provided a Norwegian explanation to prevent possible confusions on how to apply the codes. Based on the input from my colleagues I did the final coding of statements. Please find the coding schemes for each of the topics in Appendix 3. The results from the coding are presented in Section 6.2.2.

To allow comparison between the Web statements and the criterion standard, I evaluated the results of the Cochrane reviews by using the same coding system. A brief description of the reviews and their defined outcomes is provided in Box 1 below.

Box 1: ***Description of the Cochrane reviews (criterion standard)***

Cochrane review: Low glycaemic index of low glycaemic load diets for overweight and obesity (Thomas et al., 2007)
The authors included randomised controlled trials where a low glycemic diet was compared to other diets. The study participants were males and females of any age, including children, who were classified as overweight or obese using validated and specific criteria. People with diabetes mellitus were excluded. The primary outcomes were:

- change in body mass (kg)
- body mass index (BMI)
- fat distribution
- any occurring adverse effects

The authors distinguished between short term efficacy (less than six months), intermediate efficacy (six months to less than 12 months) and long-term efficacy (12 months and over) respectively.

Cochrane review: Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems (Hofmeyr et al., 2006)
The authors included randomised controlled trials where supplementation of at least one gram (1 000 milligrams) of calcium per day, starting from 34 week of gestation at the latest, was compared to placebo tablets. The primary outcomes for the women were:

- high blood pressure (women)
- pre-eclampsia (women)
- preterm birth (child)
- admission to an intensive care unit (child)
- stillbirth or neonatal death (child)

Outcomes were assessed for all pregnant women, and for specified subgroups. Subgroups were defined as: (1) women with an adequate dietary calcium intake (≥ 900 milligrams of calcium a day); (2) women with a low dietary calcium intake (<900 milligrams of calcium a day); and (3) women at high risk. High risk was defined to be teenage pregnancies, pre-eclampsia in previous pregnancies, and pre-existing high blood pressure before gestation.

The review authors provided effect estimates for the different outcomes and used statistical measures such as confidence intervals and p-values to determine their statistical significance (see Definitions in Section 1.3.1). Nevertheless, the *clinical significance* of the results, which relates to practical and applied value of an intervention effect, is still open to interpretation (Houle & Stump, 2008). I therefore asked my colleagues (n=5) to independently rate the results for the different review outcomes.

My colleagues did not read the full-text reviews; rather they were given an overview of the effect estimates and significance measures for the primary outcomes in the reviews and asked to use the coding system to rate them accordingly. The Cochrane reviews provided effect estimates for all defined outcomes, accordingly the effect code «unclear» was not applicable. When coding the review outcomes, we considered the overall results as well as results for specific subgroups defined in the reviews. The coding schemes used for the rating can be found in Appendix 4. The results from the coding are presented in Section 6.2.1.

A web statement was considered accurate if its coding matched the coding of the review results. If their coding differed from that of the reviews, they were considered inaccurate (accuracy was «not met»). Moreover, I scored webpage accuracy dichotomously. A webpage that contained at least one inaccurate statement was scored as «not met». For example, if a webpage about calcium supplements included statements about women with low dietary calcium intake as well as adequate dietary intake (see Box 1), both statements had to comply with the results of the Cochrane review to be evaluated as «met» (webpage is accurate).

A webpage considering only one of the aspects or sub-groups defined in the reviews, for example low glycemic diets in overweight children, was classified as «met» (webpage is accurate) if the statement was consistent with the review result for that particular finding.

5.3 IDENTIFYING WEBPAGES FOR EVALUATION

5.3.1 SEARCH STRATEGY

Studies of how lay people search and navigate for health information have revealed that searches are mostly done from general search engines than from identified medical websites (Morahan-Martin, 2004; Renahy & Chauvin, 2006). Thus, I decided to use four popular search engines to identify webpages: Google[1], Kvasir[2], Yahoo[3], and MSN[1]. The Norwegian

[1] www.google.no

[2] www.kvasir.no

[3] http://no.yahoo.com

search engines Sesam and ABC Startsiden were not included as their web search is being carried out by Yahoo. I did a test search across search engines to see how they overlapped by using two search phrases («glykemisk indeks overvekt» and «lavglykemisk fedme») and then comparing the first search results on each page. In many cases the overlap was a 100% between Google and Kvasir, however sometimes they provided unique URLs. This is probably due to that Kvasir's search is being carried out by Google, yet their default search options are different, with «search the web» and «search Norwegian sites» being the default in Google and Kvasir respectively. Accordingly, I decided to include both engines.

I performed the search for low glycemic diets over two subsequent days early September 2008. The search for calcium supplementation was performed over two subsequent days early October 2008. I used a variety search terms that lay people are likely to use and these were adapted to local spellings and expressions. I used terms related to both the *condition* (i.e. overweight) and the *intervention* (i.e. low glycemic index diets). To allow comparison with information indexed in the health portal SMIL, I performed the same searches in the portal's internal search engine. In addition, I used the list of predefined health topics to find relevant information. Please see Appendix 5 for an overview of the search strategies.

Lay peoples' search skills are limited; they tend to use short phrases and do not employ any advanced search strategies, and they seldom go beyond the first page of a search (Eysenbach & Köhler, 2002; Morahan-Martin, 2004; Renahy & Chauvin, 2006). Accordingly, I did not use any advanced search techniques and only the URLs found on the first result page in each search engine were recorded when performing the different searches. All engines were searched using the default search options. I recorded both sponsored (advertisements) and unsponsored results. In SMIL, all hits identified through searches and the predefined topics were recorded. I removed duplicate hits across search engines and dead links.

I followed internal links within a retrieved webpage if their labels indicated that relevant content could be found elsewhere on the site. For example, a webpage on glycemic index provided by the lay health information source *Lommelegen* was retrieved by many search engines and strategies. On the page, related articles are linked to in a separate box (see Figure 5). I did not follow any external links from retrieved webpages.

[1]http://no.msn.com/

Figure 5: ***Screenshot from Lommelegen.no***

5.3.2 SAMPLING

I defined a set of eligibility criteria to select webpages for further evaluation. Only Norwegian webpages/sites providing general health information intended for a lay audience were considered for inclusion. Providers could be governmental (public), commercial, or non-profit organizations. Webpages were included if they:

a) Referred to both the intervention and the condition, for instance calcium supplementation and pregnancy; or if they

b) Referred to both the intervention and the outcomes of interest, for instance low glycemic diets and weight reduction

In the case of calcium supplements in pregnancy, I included webpages giving advice on supplementation even if they did not mention the specific outcomes in the Cochrane review, i.e pre-eclampsia and high blood pressure. This compares to the general advice of taking folic acid before and during the first two months of pregnancy, where the specific outcome of preventing neural tube defects frequently is not mentioned. Moreover, I included webpages about low glycemic diets even if they did not mention the population of overweight or obese people specifically. Accordingly, webpages about such diets were included if they mentioned it as part of a weigh reduction strategy in general.

My study deals with the «informational» facet of cyberhealth rather than its «communication» one. Accordingly, I excluded information and advice on lay discussion forums and blogs. Analysing such forums would be like analysing conversations, which whilst potentially valuable in terms of finding out what lay people may believe, was beyond the scope of this study. However, I did include information provided by health professional on online question-and answers services. Webpages were excluded if there was no information on who owned the host website (disclosure of ownership).

I excluded webpages aimed at health professionals, lay peoples private pages, general news magazines or news services webpages, commercial product pages (i.e. selling medical remedies or books), websites requiring subscription and/or registration, and webpages that consisted only of a list of external links. Other webpages clearly not relevant, such as phone directories, minutes of meetings, and students' assignments were also excluded.

Webpages containing information aimed at a patient group excluded in the Cochrane reviews were also excluded from the analysis. In the review on low glycemic diets, studies of obese people with diabetes were excluded (Thomas et al., 2007). In the review on calcium supplementation, women already diagnosed with pregnancy-induced hypertension were excluded (Hofmeyr et al., 2006). Consequently, I omitted webpages targeted at these sub-groups from my evaluation.

Included webpages were classified by provider type (public agencies, commercial, non-profit organization) and retrieval method (SMIL or search engines) to allow comparisons on these attributes to the webpages scores on accuracy.

5.4 EVALUATION CRITERIA

5.4.1 DEFINING THE TECHNICAL EVALUATION CRITERIA

As described in Chapter 4, Eysenbach et al. (2002) identified a set of technical criteria that have been commonly used across studies. I decided to use their list of criteria (n=24) as a starting point. To avoid bias due to subjective assessment of the criteria, I only included those having an acceptable level of inter-observer reliability in the validation study by Bernstam, Sagaram, et al. (2005), which I refer to in Chapter 4. The authors determined inter-observer agreement by calculating percentage agreement and kappa coefficient scores (kappa scores) between two rates.

The kappa coefficient is a statistical measure of agreement between two raters who, independently, assign items to one of m categories, i.e. *yes/no/not available* or *met/not met* (Altman, 1991). It is generally thought to be a more robust measure than simple percent agreement calculation since the kappa coefficient takes into account the agreement occurring by chance. The kappa coefficient ranges from 0 to 1. If there is complete agreement between raters, then the kappa coefficient is 1. If there is no agreement between raters the kappa coefficient is 0. In general, a kappa score between 0.60 and less than 0.80 is considered to indicate good agreement between raters. A score of 0.80 up to 1 indicates excellent agreement.

Bernstam, Sagaram, et al. (2005) defined criteria with a kappa score of 0.60 or higher to have an acceptable inter-observer agreement. Accordingly, I only considered these criteria for inclusion in my study. During the validation process the authors agreed on whether a single webpage or the entire website should be reviewed to determine whether a given criterion was met. An overview of the criteria that obtained adequate kappa scores, along with their level of evaluation, is provided in Table 3.

Table 3: ***Overview of technical criteria with a kappa score of 0.60 or above***[a]

#	Technical criterion	Reliability	Kappa score	Evaluation level
1.	Date of creation disclosed	Excellent	1	Page
2.	Date of last update disclosed	Excellent	1	Page
3.	Date of creation or update disclosed[b]	Excellent	1	Page
4.	References provided	Excellent	0.905	Page
5.	Editorial review process	Excellent	0.905	Site
6.	General disclosures (authorship, ownership, sponsorship, date/update)[b]	Excellent	0.810	Page
7.	Sources clear	Excellent	0.810	Page
8.	Disclosure of authorship	Excellent	0.810	Page
9.	Copyright notice[b]	Excellent	0.810	Site
10.	Credentials of physicians disclosed[c]	Excellent	100%	Page
11.	Authors' credentials disclosed	Good	0.779	Page
12.	Authors' affiliation disclosed	Good	0.779	Page
13.	General disclaimers provided	Good	0.714	Site
14.	Disclosure of ownership	Good	0.714	Site
15.	Internal search engine present	Good	0.619	Site
16.	Feedback mechanisms provided	Good	0.619	Site
17.	Fax number provided[b]	Good	0.619	Site
18.	Disclosure of advertising	Good	0.619	Page
19.	Statement of purpose	Good	0.619	Site

a) After Bernstam, Sagaram, et al., 2005

b) Not included in my study

c) It was not possible to calculate the kappa score for this criterion. Thus, the percentage agreement was used.

I found some criteria to be overlapping and thereby redundant. This applied to the criterion «fax number provided» which is covered by the more general criterion «feedback mechanism provided». Moreover, the criterion «general disclosures» was by the authors defined as «disclosure of either authorship, ownership, sponsorship or currency of information» (Bernstam, Sagaram, et al., 2005, p. 679). With the exception of sponsorship, which had poor inter-observer agreement (K=, all the other issues were covered as separate criteria (i.e. «disclosure of authorship»). Hence, I excluded the umbrella criterion «general disclosures».

This also applied to the criterion «date of creation or last update disclosed». I excluded the criterion «copyright notice» as Norwegian copyright laws give all creators of an original work exclusive right to control its distribution (Norwegian Copyright Act, 1961).

To summarise, I employed a total of 15 technical criteria on the included webpages and their host sites. As defined by Bernstam, Sagaram, et al. (2005) the criteria rating options were «yes» (criterion met), «no» (criterion not met), and «not applicable» (e.g. when no author was identified, then author credentials were identified as not applicable).

I used simple, descriptive statistics to analyse the data. For each criterion, I calculated the proportion of accurate webpages rated as «yes» (webpage is accurate and complies with criterion) and «no» (webpage is accurate, but does not comply with criterion). I did the same calculations for the inaccurate webpages. Moreover, I compared the proportions of accurate and inaccurate webpages that complied with each criterion to see if any differences occurred.

I classified the technical criteria by using the three categories defined by Fogg: operator, content, and design (see Chapter 3). In the next sections I describe each category and the respective criteria briefly.

5.4.2 WEB OPERATOR CRITERIA

Web operator criteria relate to the people or organisation that runs the website. I considered the following technical criteria to belong to the web operator category:

- disclosure of ownership
- statement of purpose
- general disclaimers provided
- provision of a feedback mechanism
- editorial review process

The criterion «editorial review process» concerns whether the web operator clearly states how the information is produced and reviewed. Thus, it relates to the content development process. However, its' operational definition merely asks for a *presence* of a claim of how the content was produced (Bernstam, Sagaram, et al. 2005). If following evidence-based guidelines for developing content, a web operator should take the necessary steps to systematically search for and appraise existing research evidence. This also includes the use of systematic reviews where available. Therefore, I decided to add another two criteria from the LIDA instrument (Minervation, 2007) that target literature searching and critical appraisal. As described in Chapter 4, this instrument was used by Eiring & Tvedten (2007) in their study of Norwegian websites. I defined the criteria more precisely because the original instrument does not provide specific operational definitions.

In the final checklist I separated between «structural criteria» and «content development criteria» to distinguish the development criteria from the other technical criteria within the web operator category. In line with recommendations from Bernstam, Sagaram, et al. (2005), all criteria were evaluated at site level. The criteria and their operational definitions and rating options are found in Appendix 6.

5.4.3 CONTENT CRITERIA

Content criteria relate to aspects of the information provided. Many of the included technical criteria could be classified as content criteria. I decided to distinguish between three sub-categories of content criteria: those related to *authorship*, to *currency*, and to *attribution*. The following criteria are classified within the authorship category:

- disclosure of authorship
- authors' affiliation disclosed
- authors' credentials disclosed
- physician credentials' disclosed

The last three criteria could be seen as attributes to the authorship criterion. In the original articles the authors distinguish between disclosure of authors' credentials and physicians' credentials (Bernstam, Sagaram, et al., 2005; Eysenbach et al., 2002). Although not explicitly stated, the distinction is probably made to separate between advice given by authors with medical training and those in allied health, such as nurses, physiotherapists, dieticians, et cetera. Information about authors' attributes is not always found on a webpage level. Although Bernstam, Sagaram, et al. evaluated the attribution criteria at webpage level, I decided to rate them as «yes» (criterion met) if their details could be obtained at site level.

The currency sub-category included the criteria about date disclosures:

- date of creation disclosed
- date of last update disclosed

Attribution measures whether the source of information is identified, with references given, including links to scientific studies. The attribution sub-category consisted of the following criteria:

- sources clear
- references provided

As the attribution criteria' operational definitions merely request the presence of a source or a conventional reference (Bernstam, Sagaram, et al. 2005), I decided to register the *type* of sources or evidence used. The evidence types were informed by those used as criterion standards in the various studies identified by Eysenbach et al. (2002): guideline, systematic review, single study, expert or textbook, and personal opinion.

If the authors explicitly stated that the information was based on national guidelines (e.g. «According to the nutritional recommendations in pregnancy from Helsedirektoratet...»), I defined the evidence type to be «guideline» even if no conventional reference was provided. On the other hand, to be registered as a systematic review or single study, a conventional reference had to be present. Medical textbooks and compendia tend to be written by experts

(Antman et al., 1992); accordingly I merged textbooks and experts into the same category. If the information was authored by a person with some kind of specialisation (i.e. a clinical dietician), yet no other sources were referenced, I defined the evidence type to be «expert». If an author name was present, yet no details on his or her credentials were provided and no other sources were referenced, I defined the evidence type to be «personal opinion». I used the category «other» in cases where the evidence used differed from the options defined above. Where no author, source, or reference was present, the type of evidence was defined as «not available».

With the exception of the authorship attribution criteria, all content criteria were evaluated at webpage level, as defined by Bernstam, Sagaram, et al. (2005). The final set of web content criteria with the operational definitions, rating options, and descriptions of evidence types can be found in Appendix 7.

5.4.4 DESIGN CRITERIA

The design criteria relate to how the site is put together. Two of the included technical criteria were defined as design criteria:

- disclosure of advertising (information design)
- internal search engine present (technical design)

The set of web design criteria with their operational definitions and rating options are found in Appendix 8.

5.5 METHODOLGICAL CONSIDERATIONS AND LIMITATIONS

As explained in the beginning of this chapter a cross-sectional study, or descriptive surveys, captures information about a situation at one moment in time. Online web content changes continuously and a cross-sectional will only gather snapshots of information at a specific point of time, i.e. the time of retrieval of a webpage. A longitudinal survey would have tracked changes in the «behaviours» of webpages and websites because information is gathered at multiple points in time (Powell & Connaway, 2004). For example, a webpages evaluated as inaccurate at one point of time could have been assessed as accurate later due to updates made by the web operator. My study will not reveal such changes; consequently the accuracy status of webpages might be different as of today. Due to the time frame of a Master's thesis, a longitudinal survey was not possible.

A descriptive cross-sectional study can only be used to describe associations of relationships but cannot establish a causal relationship (Powell & Connaway, 2004, p.87). In other words, by using a cross-sectional design I can say something about the association between established technical criteria (surface credibility) and the accuracy of information on the sampled webpages (content credibility). However, this study will not reveal how indicators for accuracy and website and webpage characteristics are related. Such relationships should be explored by an analytical approach, where statistical methods such as multivariate regression are used to test whether technical criteria can predict accuracy. By using a multivariate regression analysis it is possible to measure a number of characteristics of webpages, such as «disclosure of authorship» and «date of update» and use those as a basis for predicting how successful a website is in providing accurate information. However, to perform a regression was beyond the scope of my Master's thesis.

I considered using the EU-criteria because they are employed by those developing the SMIL-portal. However, some of the operational definitions for the criteria were vague and open to interpretation. For example, the criterion «Accessibility» is defined as «attention to guidelines on physical accessibility as well as findability, searchability, readability, usability, etc.» (European Commission, 2005, p. 6). Because I was doing the technical evaluations of the Web pages alone, I wanted to use criteria that had been properly operationalised and validated.

I also considered the DISCERN-instrument. This would have been a relevant instrument to use because it specifically targets information about treatment interventions. As described in Section 3 it has been subject to validation. Nevertheless, my aim was to see if technical criteria commonly promoted by librarians and health professionals could be used to discriminate accurate and inaccurate information on the Web. Accordingly, the instrument did not fit my purpose.

I decided to use previously operationalised and validated criteria to ensure objective assessment of the technical web characteristics, yet anything that involves assessment is open to judgement of the person performing the evaluation. I did not involve a second assessor when evaluating the technical and development criteria for the included webpages. This might have influenced the final evaluations. There are also other important non-technical features of websites, such as usability and readability, which are not evaluated in this study.

One of my purposes was to investigate whether the health information identified via the health portal SMIL[1] proved more credible than information retrieved via the search engines. SMIL is currently undergoing a major revision and the version available at the time of retrieval had not been updated since February 2008. Despite the chance of missing relevant websites, I included SMIL in the evaluation. I considered would still be possible because the portal providers link to *external* content providers rather than creating the actual content themselves. The providers of the SMIL portal use the EU quality criteria for health related websites.

Although I tried to replicate how a typical user would search, i.e. by using several variants of search terms and to use lay terms instead of medical terms, I did not involve actual users when searching for and selecting webpages. This could affect the representativeness of the included webpages as other users might have searched differently and accordingly selected other webpages. Except from me, all members of the panel assessing the Web statements were health professionals. The lack of involvement of actual users of the information evaluated in this study might have influenced the results.

[1] http://smil.uio.no/

Finally, I am a lay person myself without a health or medical background. I tried to compensate for this by choosing topics which do not require specialist knowledge, as would have been the case with a surgical procedure. I also involved health professionals (five colleagues) in coding the Cochrane reviews and the Web statements. Yet there might be important clinical aspects which I have disregarded and that could have influenced the outcomes of the evaluation.

5.6 ETHICAL CONSIDERATIONS

As this was a study of published information and involved no participants no ethics approval was required. (National Committee for Research Ethics in the Social sciences and the Humanities, 2003, September 24).

6. FINDINGS

6.1 RESULTS FROM SEARCH ENGINES AND SMIL SEARCHES

After removing duplicates and dead links, the search for low glycemic diets retrieved a total of 452 unique webpages. Of these, 20 webpages were included for analysis, whereas 432 webpages were excluded. The calcium supplementation searches retrieved 520 unique pages; of which 25 web pages were included and 495 were excluded.

Webpages were excluded because they were aimed at health professionals; they were lay discussion forums and blogs, news magazines articles, commercial product sites, or portal pages; they required subscription or registration; or they were otherwise irrelevant. Webpages were considered irrelevant if the mentioned the intervention (e.g. calcium supplementation), but targeted a different population (e.g. infants or children instead of pregnant women). A list of all the excluded webpages is available from the author on request.

I identified an additional eleven relevant webpages for low glycemic diets by investigating internal links on the index pages. No additional webpages were found for calcium supplementation by following this procedure. Totally, 56 webpages from 23 websites were included for further analysis. Figure 6 and 7 show the flow charts for the search results for each topic.

Figure 6: ***Flow chart for the low glycemic diets search results***

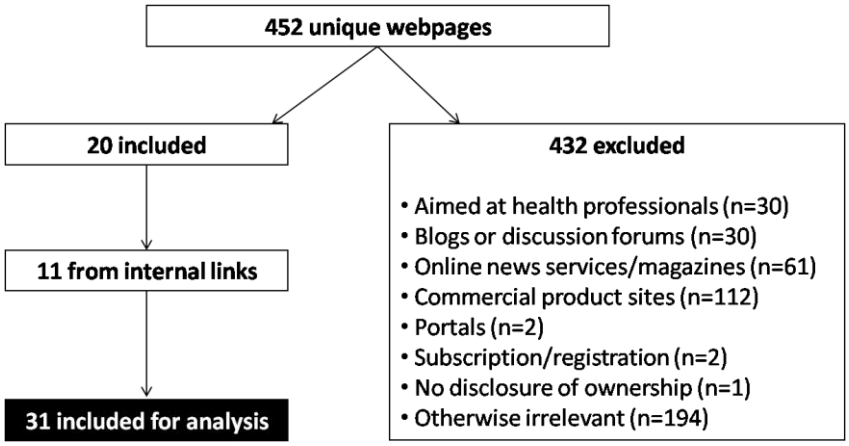
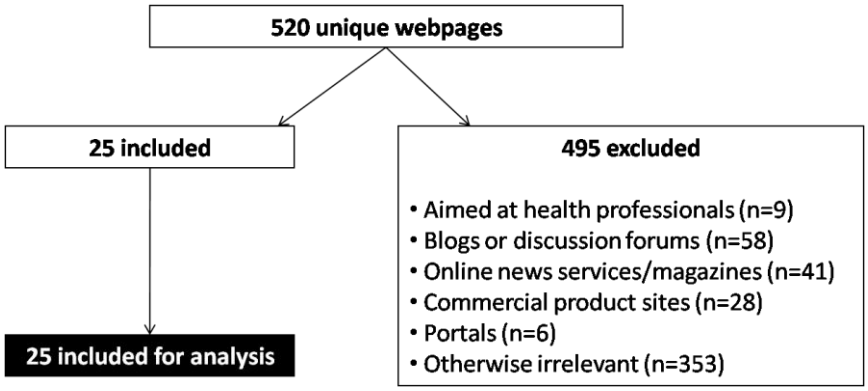


Figure 7: ***Flow chart for the calcium supplementation search result***



Four websites overlapped; they provided information about both low glycemic diets and calcium supplementation. Nineteen of the 23 websites were provided by commercial organisations, three sites by non-for-profit organisations, and one site by a public agency. On webpage level, 49 (88%) of the included pages were hosted by commercial providers, six (11%) by non-for-profit organisations, and one (2%) by a public agency. Please find an overview of the included websites and the distribution of webpages across sites in Appendix 9. The references to the webpages cited in this chapter are found in Appendix 10. Throughout the chapter, the citations are numbered and kept in brackets (e.g. *[#1]*).

6.2 ACCURACY OF WEB STATEMENTS (INFORMATION CREDIBILITY)

In this section I describe how the information (Web statements) on the different webpages scored in terms of accuracy. In the first two sub-sections I report the analysis and coding of the Cochrane review results (criterion standard) and the webpages respectively. In the third and final sub-section, the evaluations of the webpages are compared to the criterion standard. The complete data set and evaluations are available from the author on request.

6.2.1 CODING OF THE CRITERION STANDARD

Agreement was generally high among the different raters with regard to coding the results of the reviews, although some differences occurred. For instance, I first rated the short-term effect for low glycemic diets in weight reduction as «effective». After input from my colleagues, I changed the coding to «probably effective». In the review, the participants that received a low glycemic diet lost significantly more weight than did those on other diets. The p-values were below 0.05 for all weight measures (Thomas et al., 2007). However, my colleagues emphasised that the effect difference was relatively small, with the low glycemic group losing only one kilogramme more on average than the others. Other aspects that moderated the final effect coding was the relatively small number of trial participants in the overall analysis (n=202), and that a range of comparator diets was used across studies.

The second example of variations in coding related to the effect significance of calcium supplementation for the high risk pregnancy sub-group. All my colleagues coded the calcium supplements to be «effective» for this sub-group, while my own coding was «probably effective». I decided to keep the more moderate effect code, «probably effective», because the high-risk subgroup analyses involved only a few hundred women from five small trials (Hofmeyr et al., 2006). In comparison, the other subgroups (i.e. women with low calcium intake) included many thousand women in the overall analysis. Because the smaller trials in the analysis showed the largest benefits, the association between high-risk women and greater efficacy may have been overestimated in the review.

The final coding of the review results is shown in Table 4 (low glycemic diets) and 5 (calcium supplementation).

Table 4: ***Final coding of findings from the Cochrane review on low glycemic diets***

	Effect category
Short term effect (less than six months) of low glycemic diets for weight reduction in overweight and obesity	Probably effective
Long term effect (12 months or more) of low glycemic diets for weight reduction in overweight and obesity	Unknown effectiveness
Effects of low glycemic diets for overweight or obesity in childhood	Unknown effectiveness
Adverse effects of low glycemic diets in overweight and obesity	Unknown effectiveness

Table 5: ***Final coding of findings from the Cochrane review on calcium supplementation***

	Effect category
Effect of taking calcium supplementation in a mixed population of pregnant women	Probably effective
Effect of taking calcium supplementation in pregnant women with a low calcium intake (less than 900 milligrams a day)	Effective
Effect of taking calcium supplementation in pregnant women with an adequate calcium intake (900 milligrams or more per day)	Not effective
Effect of taking calcium supplementation in pregnant women at high risk[a]	Probably effective

a) In the Cochrane review, high risk was defined as teenage pregnancies, pre-eclampsia in previous pregnancies, and pre-existing high blood pressure.

According to the final coding of the criterion standard, webpages about low glycemic diets were considered accurate («met») if their statements indicated that:

- In the short-term, overweight or obese people who want to lose weight can probably benefit from choosing a low glycemic diet («probably effective»).

- The long-term effect of low glycemic diets in overweight and obesity is not known («unknown effectiveness»).
- The effect of low glycemic diet for overweight or obesity in childhood is not known («unknown effectiveness»).
- The adverse effects of being on a low glycemic diet when overweight or obese are not known («unknown effectiveness»).

As stated in Section 5.3.2, webpages that mentioned «weight reduction» were included, even if they did not mention overweight or obese people specifically.

Webpages about calcium supplementation in pregnancy would be accurate if statements indicated that:

- Pregnant women with an adequate calcium intake, that is 900 milligrams or more per day, do not have any extra benefit from taking calcium supplementation («not effective»).
- Pregnant women with a low calcium intake, that is less than 900 milligrams per day, do benefit from taking calcium supplementation («effective»)
- Overall, when not distinguishing between low and adequate calcium intake, pregnant women do probably benefit from taking calcium supplementation («probably effective»).
- Pregnant woman at high risk of pre-eclampsia do probably benefit from taking calcium supplementation («probably effective»)

As I stated in Section 5.3.2, I did include webpages even if they did not mention the specific outcomes defined in the Cochrane review (high blood pressure, pre-eclampsia, etc.).

6.2.2 CODING OF THE WEBPAGES

During the coding process the context of a Web statement was taken into consideration. However, in many cases the assessment of statements was not straightforward. Some challenges were general, while others applied to the specific topics. More general issues

related to the use of the Norwegian verbal auxiliary «kan», which corresponds to the English auxiliaries «can» and «may»:

> Det kan imidlertid være nødvendig med tilskudd hvis du skal få i deg minst 900 mg kalsium hver dag som gravid. [#1].
>
> *However, it might be necessary to take a supplement to ensure a daily intake of 900 mg of calcium a day during pregnancy.*
>
> Lav GI betyr altså at du får en lavere økning av blodsukkeret etter et måltid, og at du kan forbedre kroppens følsomhet for insulin - et viktig hormon i reguleringen av blodsukkeret. Det gjør det lettere å kontrollere kolestrolnivået, det demper appetitten og kan hjelpe folk å gå ned i vekt. [#2].
>
> *Low GI leads to smaller fluctuations in your blood glucose levels after you've finished a meal and can improve your body's sensitivity to insulin - an important hormone for the regulation of blood sugar. This makes it easier to control the cholesterol level; it reduces the appetite, and can help people to lose weight.*

There was agreement among the panel of raters that the statements above indicated a beneficial effect of the interventions. However, assessments on the *degree* of effect varied between us, and in some cases even within the same rater for similar statements. Some raters interpreted the use of «kan» as somewhat modifying the effect and consequently coded the statements as «probably effective». Others did not assign any modifications and coded the statements to indicate a definite effect («effective»). Nevertheless, because the majority of such statements were assigned the code «probably effective» across raters, I applied this as a rule in the final coding of statements.

Interpretations did also vary between my own coding and that of my colleagues for statements in which low glycemic diets or calcium supplementation were listed as an option, or as one of various interventions, for example:

> Barn og overvekt: Et råd til foreldre kan være å oppmuntre til mer fysisk aktivitet - ikke mase om slanking og mat, men bare endre familiens kosthold - f.eks. til mat med lav glykemisk indeks. [#3]
>
> *Overweight in children: Parents could be advised to encourage more physical activity – to not be persistent in terms of dieting and food, but rather just change the family's diet, for example by choosing low glycemic index food.*

> Under svangerskapet er det viktig å passe på at man får i seg nok kalsium, blant annet av hensyn til barnets benbygning. Anbefalt inntak for gravide er 900 milligram pr. dag, ifølge nordiske næringsstoffanbefalinger som Statens ernæringsråd har vært med på å utarbeide. Kalsiumbehovet kan dekkes med to-tre glass melk og 50 gram gulost pr. dag. Hvis man ikke drikker melk, må man sørge for å få dekket kalsiumbehovet på annen måte, f.eks. gjennom ost, yoghurt eller eventuelt kalsiumtabletter. [#4]
>
> *During pregnancy it is important to get enough calcium, for your baby to grow strong bones for one thing. According to the Nordic nutritional guidelines (...) the recommended dietary intake is 900 mg a day. 2-3 glasses of milk and 50 g of cheese every day meets the requirements. If you don't drink milk, you must make sure to meet the requirements in other ways, for example by eating cheese, yoghurt, or possibly by taking calcium supplements.*

Initially I interpreted the statements above to not express an effect. In the first example, I considered the statement to indicate that low glycemic food is one option when choosing a diet for overweight children, but there could also be other diet options (which are not specified). In the second example, I evaluated the statement to express calcium supplements as one of various options for meeting the requirements. However, my colleagues considered the statements above, and those similar, to express an effect and used the coding system to assign the degree of effect indicated. Their input became decisive in the final coding.

I also faced some challenges when interpreting information on the specific topics. For example, assessment sometimes proved difficult with regard to the distinction between short-term and long-term effects of low glycemic diets. Often, this was not specified on the different webpages:

> Lav GI betyr altså at du får en lavere økning av blodsukkeret etter et måltid, og at du kan forbedre kroppens følsomhet for insulin - et viktig hormon i reguleringen av blodsukkeret. Det gjør det lettere å kontrollere kolestrolnivået, det demper appetitten og kan hjelpe folk å gå ned i vekt. [#2]
>
> *Low GI leads to smaller fluctuations in your blood glucose levels after you've finished a meal and can improve your body's sensitivity to insulin, which is an important hormone for the regulation of blood sugar. This makes it easier to control the cholesterol level; it reduces the appetite, and can help people to lose weight.*

The statement above was assessed as expressing low glycemic diets to be «probably effective» in weight reduction, which is in line with the conclusions of the Cochrane review. However, it lacks information about effect duration. I decided to rate this and similar statements as «met» because they are not wrong (low glycemic diets can help people lose weight), yet the information is *incomplete*. The authors of the Cochrane review did not find any long-term studies on the effect on low glycemic diets, hence they could only conclude about short-term efficacy. Nevertheless, where effect duration was specified, statements were only coded as «met» if they were consistent with the findings of the review. Accordingly, the following statement was rated as «not met»:

> Flere langtidsstudier viser at det ikke er noen forskjell mellom vektnedgangen til de som har spist lavglykemisk og de som har fulgt en vanlig mager kost. [#5]
>
> *Various long-term studies show that there are no difference in weight loss between those on a low glycemic diet and those who have followed a standard energy-reduced diet.*

In comparison, the granularity of information was vital in webpages about calcium supplementation that distinguished between pregnant women with and without a sufficient dietary calcium intake. The specified limit value in the Cochrane review was 900 milligrams per day, which also corresponds to the Norwegian national nutritional recommendations on calcium intake in pregnancy (Directorate of Health and Social Affairs, 2005b). In the final coding, statements that specified the limit value, or its equivalent in food, were rated as «met» (statement was accurate):

> Det er viktig å få i seg nok kalsium, blant annet til barnets benbygning. Anbefalt inntak for gravide er 900 mg per dag. Dette behovet dekkes med 2–3 glass melk og 50 g gulost hver dag. Drikker du ikke melk, må du sørge for å få dekket behovet på annen måte, med ost, yoghurt eller kalsiumtabletter. [#6].
>
> *It is important to get enough calcium, for your baby to grow strong bones for one thing. The recommended intake in pregnancy is 900 mg a day. 2-3 glasses of milk and 50 g of cheese every day meet with the requirements. If you don't drink milk, you must make sure to meet the requirements in other ways, by eating cheese, yoghurt or taking calcium supplements.*

Otherwise, if a statement only contained words such as «adequate» or «sufficient» without any further specification, I considered it to be «unclear» and therefore «not met» (statement was inaccurate):

> Oppbyggingen av skjelettet begynner allerede i mors liv (…). Viktig er det at mor får i seg nok melk og melkeprodukter. Kalktabletter kan hun ta hvis hun ikke får nok kalsium gjennom kosten. [#7]
>
> *Bone development and growth starts while in utero (...). It is important that the mother has a sufficient intake of milk and dairy products. She can take calcium supplements if she does not get enough calcium through her diet.*

Similar to the webpages on low glycemic diets, there were also many examples of incomplete, yet not wrong, information on calcium supplementation. Several statements were advice on supplements targeted at those with a sufficient calcium intake, without clearly defining the opposite case (those with an insufficient intake). As stated in Chapter 5, webpages where statements targeted only one subgroup of pregnant women were evaluated to be accurate if they were consistent with the Cochrane review result for that particular subgroup. In practice, such statements would still be incomplete due to the lack of information on the other sub-groups. For instance, I coded the statement below as accurate because it complies with the conclusions of the Cochrane review for those with a sufficient calcium intake:

> Gravide har også økt behov for kalsium. Det var lenge vanlig å anbefale gravide å drikke mye melk. Norsk kosthold inneholder forholdsvis mye melk og melkeprodukter (ost, yoghurt og lignende), og gir om lag 900 mg kalsium daglig. Anbefalt daglig inntak av kalsium i svangerskapet er 900 mg. (..) 2-3 glass melk og 50g hvit ost daglig sikrer kalsiumtilførselen i svangerskapet. (…). Hvis du passer på å drikke melk og spise ost, er det unødvendig å spise ekstra kalktabletter eller ekstra kalsiumrik mat i tillegg til den daglige kosten. [#8]
>
> *Pregnant women have an increased need for calcium. It was long recommended to drink lots of milk during pregnancy. The Norwegian diet holds fairly amounts of milk and dairy products (cheese, yoghurt etc.), providing about 900 mg of calcium a day. The recommended calcium intake in pregnancy is 900 mg (...). 2-3 glasses of milk and 50 g of cheese meets the calcium requirements during pregnancy (...). If you make sure to drink milk and eat cheese it is unnecessary to take extra calcium supplementation or eat calcium-fortified food in addition to your daily diet.*

When I compared assessments between raters, our interpretations varied as to whether the statement indicates that: (a) those with an insufficient intake should take calcium supplements («effective»); (b) it is not clear what the statement says about calcium supplements for those with an insufficient intake («unclear»); or (c) the statement does not say anything about the effect of supplementation for those with an insufficient calcium intake («no statement of effect»). After discussion most raters agreed on the first option. Therefore, I coded this and similar statements to be accurate.

Table 6 gives an overview the effect codes with some examples of statements from the various webpages.

Table 6: ***Some examples of coded statements***

Statements coded as «effective»	Å spise mat med lav GI er et smart trekk for å få til en varig vektnedgang og større overskudd, både på kort og lang sikt. [#9] *Eating low GI food is a good idea to accomplish sustainable weight reduction and more energy, in the short as well as the long term.* Siden behovet for kalsium under amming og graviditet er høyere enn vanlig, vil jeg anbefale deg å ta et kalsiumtilskudd. [#10] *Due to the increased need for calcium during breastfeeding, I recommend you to take a calcium supplement.*
Statements coded as «probably effective»	Dietter basert på glykemisk indeks kan se ut til å virke en smule bedre enn slankekurer som ikke tar hensyn til hvordan kjemien i maten påvirker blodsukkeret, mener forskere fra University of Sydney. [#11] *Diets based on low glycemic index tend to be a little more effective than diets that don't pay attention to how nutritional chemistry influences the blood sugar, according to researchers at University of Sydney.* Det kan imidlertid være nødvendig med kalsiumtilskudd hvis du skal få i deg minst 900 mg kalsium hver dag som gravid. [#1] *It can/may, however, be necessary to take a calcium supplement to ensure a daily intake of 900 mg of calcium a day during pregnancy.*
Statements coded as «unknown effect»	Når det gjelder langtidsstudier og effekten av lav GI på vektreduksjon, er holdepunktene langt færre. [#12] *When it comes to long-term studies and the effect of low GI and weigh loss, there are far less facts to build upon.* Følgende tiltak synes å begrense risikoen for å bli rammet av svangerskapsforgiftning (…): Tilskudd av kalsium, magnesium, vitamin C eller vitamin E. Nytten av ovennevnte tiltak er foreløpig ikke tilstrekkelig avklart. Spesielt vet man ikke nok om hvilke kvinner som bør få en eller flere av disse tilskuddene. [#13] *The following interventions seem to reduce the risk of pre-eclampsia (…); Supplementation of calcium, magnesium, vitamin C or vitamin E. The benefits of these interventions are not sufficiently clear. We don't know enough about what women should take one or more of these supplements.*
Statements coded as «not effective»	Studien viser at de som spiste mest proteiner var de som beholdt vekten best, og at lav GI ikke hadde noen som helst påvirkning. - Vi kunne ikke se noen sammenheng mellom lav GI og vektreduksjon. [#14]. *The study shows that those with a high protein intake kept their weight reduced, and low GI did not have any effect. – We were not able to detect any correlation between low GI and weight reduction.* Anbefalt daglig inntak av kalsium i svangerskapet er 900 mg. (..) 2-3 glass melk og 50g hvit ost daglig sikrer kalsiumtilførselen i svangerskapet. (…). Hvis du passer på å drikke melk og spise ost, er det unødvendig å spise ekstra kalktabletter eller ekstra kalsiumrik mat i tillegg til den daglige kosten. [#15] *The recommended calcium intake in pregnancy is 900 mg (…). 2-3 glasses of milk and 50 g of cheese meets the calcium requirements during pregnancy (…). If you make sure to drink milk and eat cheese it is unnecessary to take extra calcium supplementation or eat calcium-fortified food in addition to your daily diet.*

Table 6: ***Some examples of coded statements (continued)***

Statements coded as «probably not effective»	[*Name of person*] er kritisk til de populære ernæringsrådene fra dr. Lindberg. Lindberg mener at vi skal spise mat med lav glykemisk indeks (GI). - Mat med mye fett og proteiner har lav GI. Selv om denne maten gir lav blodsukkerstigning, vil den kunne bidra med mye energi som er ugunstig for personer som vil ned i vekt. [#16] *[Name of person] is critical to the popular dietary recommendations from dr. Lindberg. Lindberg recommends eating foods with a low glycemic index (GI). Fatty and protein-rich food has a low GI. Even if this type of food leads to smaller fluctuations in blood glucose levels, it could provide high levels of energy which is unbeneficial for people who want to lose weight.*
Statements coded as being unclear	Når du drikker mye melk og i tillegg har et sunt og variert kosthold, er det kanskje ikke nødvendig med ekstra tilskudd av kalk [#17] *If you drink a lot of milk and have a healthy and balanced diet, it might not be necessary to take any extra calcium supplementation.* Personer som ikke klarer å dekke kalsiumbehovet gjennom kosten bør ta et kalsiumtilskudd. Dette kan gjelde personer som ikke spiser meieriprodukter, for eksempel på grunn av melke- eller laktoseintoleranse, som spiser lite, er gravide eller ammer, eller er i overgangsalderen (kvinner). [#18] *People who are not able to get a sufficient intake of calcium through their diet should take a calcium supplement. This may apply to people who don't eat dairy products, for example due to milk or lactose intolerance, people who don't' eat a lot, women who are pregnant or breastfeeding, or menopausal women.*

6.2.3 COMPARISON OF WEB STATEMENTS TO THE COCHRANE REVIEWS

In Table 7 the distribution of accurate («met») and inaccurate («not met») webpages for the two topics separately and combined. Forty-one percent of webpages were coded as «met», leaving 59% of the webpages to be evaluated as inaccurate. Above two-thirds (68%) of the low glycemic diets webpages were classified as inaccurate. The distribution of accurate and inaccurate webpages about calcium supplementation was almost half-by-half, with 13 of the 25 webpages (52%) being accurate.

Table 7: ***Accuracy of content on the different webpages***

	Total no. pages (sites)	Accurate webpages	Inaccurate webpages
Low glycemic diets	31 pages (13 sites)	10 (32%)	21 (68%)
Calcium supplementation	25 pages (14 sites)	13 (52%)	12 (48%)
Combined	56 pages (23 sites)	23 (41%)	33 (59%)

Taking the two topics together, the most frequent reason for inaccuracy was webpage statements being coded as «unclear». Sometimes this was due to little agreement among raters, as illustrated by the statements below. The distribution of the raters' coding is specified in brackets:

> Har du har kjøpt lister over mat med høy/lav glykemisk indeks, men synes det er vanskelig? Glem GI - tenk heller grovt! Mye enklere og fører både god helse og lavere vekt (…). GI ikke hele løsningen. Skal du ned i vekt er det helt essensielt at du spiser på en slik måte at det gir deg færre kalorier enn tidligere og at dette er et kosthold du kan klare å holde på over tid. Du kan velge å la deg inspirere av nye trender – eller du kan holde deg til et norsk tradisjonelt kosthold! For begge tilfeller gjelder det at du kan glemme dette med GI (glykemisk indeks) – bare du tenker GROVT, GRØNT og GUNSTIG! [#19]
> (Raters' coding: «probably effective» = 1; «not effective» = 2; «unknown effect» = 1; «unclear» = 2).
>
> *Have you bought overviews of food with high/low glycemic index, yet having difficulties? Forget about GI –focus on fiber! It is much simpler and results in good health and reduced weight (…). GI is not the only solution. If you want to lose weight it is essential that you consume fewer calories and that your diet is sustainable over time. You can choose to be inspired by new trends – or you can hang on to a traditional Norwegian diet! In any case you can forget about GI (glycemic index) – as long as you focus on FIBER, GREENS and BENEFITS!*
>
> Det kan være lurt å ta et kosttilskudd når du er gravid, men med et sunt kosthold er det strengt tatt ikke nødvendig. Når du drikker mye melk og i tillegg har et sunt og variert kosthold, er det kanskje ikke nødvendig med ekstra tilskudd av kalk. [#17]
> (Raters' coding: «probably effective» = 1; «probably not effective» = 2; «not effective» = 1; «unclear» = 1).

> *It could be an idea to take a nutritional supplement when you are pregnant, although strictly speaking it is not necessary with a healthy diet. If you drink a lot of milk and have a healthy and balanced diet, it might not be necessary to take any extra calcium supplementation.*

The reasons for being classified as inaccurate also differed between the two topics. Statements being coded as «unclear» as the most frequent reason for inaccuracy among webpages on calcium supplementation. Sometimes this was related to missing information about the calcium intake limit value of 900 milligrams. Seven of the 25 web pages lacked information about the limit value. However, five of them were also inaccurate due to other statements that were not consistent with findings of the reviews

Providing unclear information was also a common reason for inaccuracy among the low glycemic diets web pages. However, a similar proportion of webpages were inaccurate because their statements were evaluated to indicate diets to be «effective»:

> Det er viktig å huske at dersom alle ingredienser til en rett har lav GI, vil retten også ha lav GI. Det er en sikker vei å gå så lenge man ønsker å gå ned i vekt. [#20]
>
> *It is important to remember that as long as all ingredients have a low GI, the dish itself will also have a low GI. That will surely work when you want to lose weight.*

Only four webpages distinguished between short-term and long-term effects of diets, and only two pages correctly pointed to the lack of evidence about long-term effects («unknovn effectiveness»). Yet, as previously stated, webpages that did not make this distinction were still considered accurate if their effect statements were in line with the coded results of the review («probably effective»).

Table 8 and 9 summarises the reasons for statement inaccuracies on the webpages for each of the topics respectively. As one webpage could include more than one inaccurate effect statement, the numbers do not correspond to the total number of inaccurate webpages.

Table 8: ***Reasons for inaccuracy for the low glycemic diets webpages***

Web statements	Code	No. of pages
Low glycemic diets are effective for weight reduction	Effective	8
Low glycemic diets are probably effective for weight reduction in children	Probably effective	1
It is uncertain/we don't know if low glycemic are effective for weight reduction	Unknown effect	2
Low glycemic diets are not effective for weight reduction	Not effective	1
Low glycemic diets are probably not effective for weight reduction	Probably not effective	1
The statement is unclear about the effect of low glycemic diets	Unclear	9

Table 9: ***Reasons for inaccuracy for the calcium supplementation webpages***

Web statements	Code	No. of pages
Pregnant women in general should take calcium supplementation	Effective	1
Pregnant women with a low dietary calcium intake should probably take calcium supplementation	Probably effective	3
It is uncertain/we don't know if pregnant in general women should take calcium supplementation	Unknown effect	1
Pregnant women at high risk should take calcium supplementation	Effective	1
The statement is unclear about the effect of calcium supplementation	Unclear	9

In some cases, I found inconsistent information on webpages *within* the same website. The following recommendations on calcium supplementation were given by the same dietician with a few days intervals on an online «Ask the expert»- service:

> Det er nylig kommet ut nye nordiske anbefalinger for næringsinntak. Her er anbefalingene for gravide følgende: (…) 900mg (milligram) for kalsium (…). Bortsett fra tran og folat, anbefales det ikke noe ekstra under graviditeten. Hvis du spiser lite melkeprodukter, er jeg enig i at et kalsiumtilskudd hadde vært lurt. Det finnes flere kalsiumtilskudd uten vitamin D, som [Produkt A] 250mg, [Produkt B] og [Produkt C] 500mg. [#21].
>
> *The new Nordic nutrition recommendations have recently been published. For pregnant women the recommendations are: (...) 900 mg (milligrams) of calcium (...). With the exception of fish oil and folic acid, there are no recommendations on extra supplementation. If your intake of dairy products is low, I agree that a calcium supplement would be a good idea. There are various calcium supplements without extra vitamin D, such as [Product A] 250 mg, [Product B] and [Product C] 500 mg.*
>
> Siden behovet for kalsium under amming og graviditet er høyere enn vanlig, vil jeg anbefale deg å ta et kalsiumtilskudd. Hvis du drikker lite melk og spiser lite melkeprodukter foreslår jeg 1000mg per dag, ellers 500mg. [#10]
>
> *Due to the increased need for calcium during breastfeeding, I recommend you to take a calcium supplement. If your intake of milk and dairy products is low I suggest 1000 mg a day, otherwise 500 mg"*

In the first case, I interpreted the statement to express that a calcium supplement is recommended only when the dietary calcium intake during pregnancy is low (i.e. «effective» in women with low dietary calcium intake; «not effective» in women with adequate intake). In the latter, I interpreted the statement to express that calcium supplementation is recommended irrespective of dietary intake (i.e. «effective» in women with low as well as adequate dietary calcium intake).

Another website provided contradictory information:

> Matvarer som har høy GI gir høy blodsukkerstigning, og kroppen produserer dermed mye insulin. Flere studier tyder på at disse to effektene har negative følger for helsen din generelt, og for utvikling av overvekt. Å spise mat med lav GI – er derfor et smart trekk for å få til en varig vektnedgang og større overskudd, både på kort og lang sikt. [#9]
>
> *High glycemic index foods increase the blood sugar levels, which results in a high insulin production. Several studies suggest that these aspects have negative consequences for your health in general, and for the development of overweight. Eating low GI food is a good idea to accomplish sustainable weight reduction and more energy, in the short as well as the long term.*

> Studien viser at de som spiste mest proteiner var de som beholdt vekten best, og at lav GI ikke hadde noen som helst påvirkning. - Vi kunne ikke se noen sammenheng mellom lav GI og vektreduksjon. [#14]
>
> *The study shows that those with a high protein intake kept their weight reduced, and low GI did not have any effect. – We were not able to detect any correlation between low GI and weight reduction.*

6.3 COMPARING ACCURACY WITH THE TECHNICAL CRITERIA

In the next sections, I describe how many of the webpages that complied with the technical criteria. Moreover, I report how compliance with the criteria was distributed between the accurate and inaccurate webpages. Results are reported for the web operator, content, and design criteria respectively. The complete data set and technical evaluations are available from the author on request.

6.3.1 WEB OPERATOR: STRUCTURAL CRITERIA

I report the distribution of accurate and inaccurate webpages for the four structural web operator criteria in Table 10. Results are displayed for all webpages combined and the two topics separately.

Table 10: ***Webpages' compliance with the structural web operator criteria***

		No. of webpages	(%)	Accurate webpages	(%)	Inaccurate webpages	(%)
Disclosure of ownership (S)							
All webpages	**Yes**	**56**	**(100)**	**23**	**(100)**	**33**	**(100)**
	No	**0**		**0**		**0**	
Statement of purpose (S)							
All webpages	**Yes**	**46**	**(82)**	**21**	**(91)**	**25**	**(76)**
	No	**10**	**(18)**	**2**	**(9)**	**8**	**(24)**
Low glycemic diets	Yes	24	(77)	8	(80)	16	(76)
	No	7	(23)	2	(20)	5	(24)
Calcium supplementation	Yes	22	(88)	13	(100)	9	(75)
	No	3	(12)	0		3	(25)
General disclaimers provided (S)							
All webpages	**Yes**	**47**	**(84)**	**19**	**(83)**	**28**	**(85)**
	No	**9**	**(16)**	**4**	**(17)**	**5**	**(15)**
Low glycemic diets	Yes	25	(81)	9	(90)	16	(76)
	No	6	(19)	1	(10)	5	(24)
Calcium supplementation	Yes	22	(88)	10	(77)	12	(100)
	No	3	(12)	3	(23)	0	
Feedback mechanism provided (S)							
All webpages	**Yes**	**56**	**(100)**	**23**	**(100)**	**33**	**(100)**
	No	**0**		**0**		**0**	

Note: **S** indicates evaluated at website level; **P** indicates evaluated at webpage level.

Because I only included webpages from host sites that clearly indicated who owned the presented information, all webpages complied with the «disclosure of ownership» criterion. In addition, all websites provided contact details. Accordingly, their respective webpages complied with the criterion «feedback mechanism provided».

A higher number of the accurate webpages (91%) came from sites which stated their aims and target audience compared to the inaccurate webpages (76%). In calcium supplementation, all accurate webpages were hosted by sites that provided information about their statement of purpose, compared to 75% of the inaccurate pages. There were no real differences in compliance with the criterion when comparing accurate and inaccurate webpages about low glycemic diets, the corresponding percentages were 80 and 76 respectively.

Many websites described their main purpose to be a source where lay people could find reliable and useful information:

> Lommelegens målsetting er å bidra til helse for alle gjennom opplysning, veiledning og informasjon. Vi tilstreber å tilby så god og oppdatert informasjon som mulig på alminnelig og lett forståelig norsk. [#22]
>
> *The objective of Lommelegen is to contribute to the public's health through education, counseling and information. We aim to provide, in the best possible way, good and updated health information in a non-jargon and easy-to-understand language.*
>
> Pasienthåndboka er et generelt medisinsk oppslagsverk for legfolk. Pasienthåndboka er utviklet for å gjøre oppdatert og forskningsbasert kunnskap lett tilgjengelig for folk flest. Målet er å bidra til økt kunnskap om medisin og helse for å kunne styrke pasient og pårørendes medvirkning i medisinske beslutninger. [#23]
>
> *Pasienthåndboka is a general medical compendium aimed at lay people. The purpose of Pasienthåndboka is to make updated and research based knowledge easily available. Our aim is to contribute to increased knowledge on medical and health topics to strengthen patients and their relatives' participation in medical decisions.*
>
> Babyverden er et nettsted for deg som prøver å bli gravid, er gravid eller har små barn. Her skal du enkelt finne svar på spørsmål som dukker opp i denne begivenhetsrike tiden, og alltid finne ny, oppdatert og riktig informasjon. [#24]
>
> *Babyverden is a website for those of you who are trying to get pregnant, are pregnant, or have small children. Here you can easily find answers to questions, and you will always find new, updated and accurate information.*

Sixteen of the 23 websites (70%) had a general disclaimer (not shown in table). On webpage level, 47 of the included webpages (84%) came from sites that provided such a disclaimer. On the whole, there were no differences in how the accurate and inaccurate webpages complied with the disclaimer criterion. 83% of the accurate and 85% of the inaccurate webpages did so. In calcium supplementation, all inaccurate webpages originated from sites that fulfilled the criterion, compared to 77% of the accurate pages. In contrast, fewer of the inaccurate pages about low glycemic diets (76%) met with the criterion compared to the accurate pages (90%).

Most disclaimers were concerned with how the users of the site acted on the information provided. Statements such as «please seek the guidance of a healthcare provider before attempting to use any of this information», and «we will not be responsible for actions based on information found on our website» were commonly seen across all websites. A few

websites, particularly those who covered a range of health topics, also expressed general reservations about the accuracy of the information provided:

> Tilsiget av kunnskap fra pågående forskning, nye erfaringer fra klinisk praksis, manglende enighet blant faglige autoriteter og muligheten for menneskelige feil i arbeidet med et så omfattende nettsted som dette, gjør at verken redaktørene eller Lommelegen kan garantere at alle opplysningene på nettstedet er nøyaktige og/eller fullstendige i alle henseender. [#22]
>
> *Due to regular advances in research, new experiences in clinical practice, different opinions among professionals in the field and the possibility of human error, the editors or Lommelegen can't guarantee that all the information provided on our website is accurate and/or complete in all regards.*

6.3.2 WEB OPERATOR: DEVELOPMENT CRITERIA

Eighteen of the 23 websites gave details on how the information was produced and its quality checked («editorial review process»). The findings on webpage level for the development criteria are displayed in Table 11. Forty-nine of the included webpages originated from the 18 sites (not shown in table); accordingly a majority of the webpages (n=49; 88%) complied with the editorial review criterion. Although a larger proportion of the calcium supplementation webpages complied with the criterion, so did the majority of webpages about low glycemic diets. There were no real differences between accurate and inaccurate webpages in complying with the editorial policy criterion. Overall, 87% of the accurate webpages were hosted by sites with a policy, compared to 88% of the inaccurate webpages.

Table 11: ***Webpages' compliance with the development criteria***

		No. of webpages	(%)	Accurate webpages	(%)	Inaccurate webpages	(%)
Editorial review process (S)							
All webpages	**Yes**	**49**	**(88)**	**20**	**(87)**	**29**	**(88)**
	No	**7**	**(12)**	**3**	**(13)**	**4**	**(12)**
Low glycemic diets	Yes	25	(81)	8	(80)	17	(81)
	No	6	(19)	2	(20)	4	(19)
Calcium supplementation	Yes	24	(96)	12	(92)	12	(100)
	No	1	(4)	1	(8)	0	
Literature search conducted (S)							
All webpages	**Yes**	**0**		**0**		**0**	
	Other	**56**	**(100)**	**23**	**(100)**	**33**	**(100)**
Independent critical appraisal conducted (S)							
All webpages	**Yes**	**0**		**0**		**0**	
	Other	**56**	**(100)**	**23**	**(100)**	**33**	**(100)**

Note: **S** indicates evaluated at website level; **P** indicates evaluated at webpage level; **Other** = No or Not Applicable (i.e. when no author was identified, then author credentials were identified as not applicable).

Using experts to author and/or review information content was the most frequent development method mentioned across sites. Twelve of the 18 sites with an editorial policy explicitly mentioned that they consulted experts for their content quality assurance. Some examples:

> Alle våre tilknyttede eksperter utgjør Norges fremste på sine områder. Du kan derfor trygt bruke informasjonen du finner på Baby. [#25]

> *Our associated experts are leading within their professional areas. Thus, you can confidently use the information provided by Baby.*

> For å sikre at den informasjonen dere finner på Babyverden skal være den riktige, bruker vi et stort panel av fagfolk som konsulenter og til dels forfattere. [#24]

> *We use a large panel of specialists as consultants and partly as authors to ensure that the information you find on Babyverden is accurate.*

The other six websites listed an editorial board and provided their professional details without any additional information about how the content was developed. None of the websites specifically mentioned literature searching or critical appraisal as part of their content production method. However, the providers of a large medical online textbook did indicate that such procedures were part of their content development process:

> I alle sammenhenger bestreber vi oss på å publisere forskningsbaserte oversiktsartikler etter anerkjente internasjonale prinsipper (såkalt «evidence-based medicine») (…). Der det er mulig, vil informasjon i Pasienthåndboka være støttet av tydelige referanser til kildedata. I teksten er det for de fleste referanser indikert kvalitet på studien (Ia - metaanalyse, Ib - randomisert, kontrollert studie, IIa - kontrollert studie uten randomisering, IIb - annen type godt utformet studie, III - ikke eksperimentelle studier (epidemiologiske studier), IV - konsensusuttalelser). [#23]

> *In all areas we endeavor to publish research based review articles that follow internationally recognised principles (so-called «evidence-based medicine») (...). Where possible, the information on [Name of website] will be accompanied by explicit references to sources, and if possible. For most references in the text the quality of the study is indicated (Ia – meta-analysis, Ib – randomised controlled trial, IIa controlled study without randomisation, IIb – another type of well conducted study, III – non-experimental studies (epidemiological studies), IV – consensus statements).*

6.3.3 WEB CONTENT: AUTHORSHIP CRITERIA

The findings related to the authorship criteria are displayed in Table 12. Forty-four (79%) of the webpages gave information about author(s) of the presented information. In general, more webpages on calcium supplementation fulfilled the authorship criteria than did the low glycemic index pages. Eighty-eight percent of the calcium webpages disclosed who authored the information, compared to 71% of the low glycemic diets pages.

As explained in Section 5.4.3 the criteria related to affiliation and credentials are attributes to the authorship criterion. If no author was listed, then information about affiliation and credentials was not available. In Table 12 the results for the attribution criteria are calculated

for the total Web population only. For example, details on authors' affiliation were available in 34 of the webpages. This comprised 61% of the total webpage population. However, it comprised 77% (34/44) of the web pages that disclosed authorship (not shown in table). Furthermore, information on authors' credentials was provided 31 of the webpages; which comprised 70% (31/44) of the webpages disclosing authorship. Eight webpages were authored by a physician (14% of all webpages; 18% (8/44) of webpages disclosing authorship).

Table 12: ***Webpages' compliance with the authorship criteria***

		No. of webpages	(%)	Accurate webpages	(%)	Inaccurate webpages	(%)
Disclosure of authorship (P)							
All webpages	**Yes**	**44**	**(79)**	**19**	**(83)**	**25**	**(76)**
	No	**12**	**(21)**	**4**	**(17)**	**8**	**(24)**
Low glycemic diets	Yes	22	(71)	7	(70)	15	(71)
	No	9	(29)	3	(30)	6	(29)
Calcium supplementation	Yes	22	(88)	12	(92)	10	(83)
	No	3	(12)	1	(8)	2	(17)
Authors affiliation disclosed (S)							
All webpages	**Yes**	**34**	**(61)**	**13**	**(57)**	**21**	**(64)**
	Other	**22**	**(39)**	**10**	**(43)**	**12**	**(36)**
Low glycemic diets	Yes	12	(39)	5	(50)	11	(52)
	Other	19	(61)	5	(50)	10	(48)
Calcium supplementation	Yes	18	(72)	8	(62)	10	(83)
	Other	7	(28)	5	(38)	2	(17)
Authors' credentials disclosed (S)							
All webpages	**Yes**	**31**	**(55)**	**14**	**(61)**	**17**	**(52)**
	Other	**25**	**(45)**	**9**	**(39)**	**16**	**(48)**
Low glycemic diets	Yes	13	(42)	6	(60)	7	(33)
	Other	18	(58)	4	(40)	14	(67)
Calcium supplementation	Yes	18	(72)	8	(62)	10	(83)
	Other	7	(28)	5	(38)	2	(17)
Physicians' credentials disclosed (S)							
All webpages	**Yes**	**8**	**(14)**	**2**	**(9)**	**6**	**(18)**
	Other	**48**	**(86)**	**21**	**(91)**	**27**	**(82)**
Low glycemic diets	Yes	2	(6)	1	(10)	1	(5)
	Other	29	(94)	9	(90)	20	(95)
Calcium supplementation	Yes	6	(24)	1	(8)	5	(42)
	Other	19	(76)	12	(92)	7	(52)

Note: **S** indicates evaluated at website level; **P** indicates evaluated at webpage level; **Other** = No or Not Applicable (i.e. when no author was identified, then author credentials were identified as not applicable)

More of the accurate webpages disclosed authorship (83%) than did the inaccurate pages (76%). This was also the tendency for webpages disclosing authors' credentials (61% versus 52%). However, more of the inaccurate webpages disclosed authors' affiliation (64%) than did the accurate webpages (57%). The number of pages that displayed physicians' credentials is so low that meaningful comparisons cannot be made.

There were almost no differences between the accurate and inaccurate low glycemic diets webpages in how they complied with the authorship criteria, except that more of the accurate pages (60%) disclosed authors' credentials than did the inaccurate pages (33%). The opposite was the case regarding calcium supplementation, where more of the inaccurate webpages disclosed authors' credentials (83%) compared to the accurate webpages (62%).

Where authors' credentials were disclosed, most information was authored by health professionals. In many cases the authors were dieticians. The distribution of accurate and inaccurate webpages where information was authored by health professionals was almost even. Although there was a tendency of the information written by non-health professionals being more inaccurate, the number of pages is too low to make any meaningful comparisons.

Table 13: A*ccuracy related to information authored by health professionals and non-health professionals*

	Total no. pages	Accurate webpages	Inaccurate webpages
Health professionals*	25	12 (48%)	13 (52%)
Non-health professionals	6	2 (33%)	4 (67%)

*Included physicians (n=8); clinical dieticians (n=14); nurses (n=2); midwives (n=1); pharmacists (n=1)

6.3.4 WEB CONTENT: CURRENCY CRITERIA

I report the distribution of accurate and inaccurate webpages for the two currency criteria in Table 14. In total, 63% of the webpages provided a content creation date. The corresponding percentages were 48 for the low glycemic diets webpages and 80 for the calcium supplementation pages. Moreover, one-fourth (23%) of all webpages indicated when the content was last revised. There were no differences between the two topics in this respect.

In general, more of the accurate webpages displayed a creation date and/or an update date than did the inaccurate webpages. The difference was most visible for the low glycemic diets web pages. Regarding calcium supplementation, there was no real difference between accurate and inaccurate webpages for this criterion. The number of pages complying with the criterion «date of last update disclosed» is too small to make any comparisons.

Table 14: ***Webpages' compliance with the currency criteria***

		No. of webpages	(%)	Accurate webpages	(%)	Inaccurate webpages	(%)
Date of creation disclosed (P)							
All webpages	**Yes**	**35**	**(63)**	**16**	**(70)**	**19**	**(58)**
	No	**21**	**(37)**	**7**	**(30)**	**14**	**(42)**
Low glycemic diets	Yes	15	(48)	6	(60)	9	(43)
	No	16	(52)	4	(40)	12	(57)
Calcium supplementation	Yes	20	(80)	10	(77)	10	(83)
	No	5	(20)	3	(23)	2	(17)
Date of last update disclosed (P)							
All webpages	**Yes**	**13**	**(23)**	**8**	**(35)**	**5**	**(15)**
	No	**43**	**(67)**	**15**	**(65)**	**28**	**(85)**
Low glycemic diets	Yes	7	(23)	4	(40)	3	(14)
	No	24	(77)	6	(60)	18	(86)
Calcium supplementation	Yes	6	(24)	4	(31)	2	(17)
	No	19	(76)	9	(69)	10	(83)

Note: **S** indicates evaluated at website level; **P** indicates evaluated at webpage level.

I also compared the currency of the webpages for the two topics to the publication date of their respective Cochrane reviews. In Table 15 I display the results for the two topics combined. Overall, twenty-one (38%) of the webpages provided information that was created or updated before the publication of the review. Of these, a higher proportion was inaccurate (62%) than accurate (38%). Furthermore, of the 14 webpages created or updated after the publication date of their respective Cochrane reviews, more webpages (64%) were evaluated to be accurate.

Table 15: ***Currency of webpages compared to the publication date of Cochrane reviews***

	N	Accurate webpages	Inaccurate webpages
Webpage created/updated before publication of review	**21**	8 (38%)	13 (62%)
Webpage created/updated after publication of review	**14**	9 (64%)	5 (36%)
No date	**21**	7 (33%)	14 (67%)

6.3.5 WEB CONTENT: ATTRIBUTION CRITERIA

As shown in Table 16 below, 32 webpages (57%) backed their claims up with some type of source («Sources clear»). Again, compliance was better for the calcium supplementation webpages (76%) than the low glycemic webpages (42%).

More of the accurate webpages (65%) reported a source than did the inaccurate sources (52%). This was also the case for the calcium supplementation pages, the percentages were 85 and 67 for accurate and inaccurate webpages respectively. There were no notable differences between the accurate (40%) and inaccurate webpages (43%) about low glycemic diets in complying with the source criterion.

Table 16: ***Webpages' compliance with the attribution criteria***

		No. of webpages	(%)	Accurate webpages	(%)	Inaccurate webpages	(%)
Sources clear (P)							
All webpages	**Yes**	**32**	**(57)**	**15**	**(65)**	**17**	**(52)**
	No	**24**	**(43)**	**8**	**(35)**	**16**	**(48)**
Low glycemic diets	Yes	13	(42)	4	(40)	9	(43)
	No	18	(58)	6	(60)	12	(57)
Calcium supplementation	Yes	19	(76)	11	(85)	8	(67)
	No	6	(24)	2	(15)	4	(33)
References provided (P)							
All webpages	**Yes**	**6**	**(11)**	**3**	**(13)**	**3**	**(9)**
	No	**50**	**(89)**	**20**	**(87)**	**30**	**(91)**
Low glycemic diets	Yes	4	(13)	2	(20)	2	(10)
	No	27	(87)	8	(80)	19	(90)
Calcium supplementation	Yes	2	(8)	1	(8)	1	(8)
	No	23	(92)	12	(92)	11	(92)

Note: **S** indicates evaluated at website level; **P** indicates evaluated at webpage level.

A conventional reference was provided in six (11%) of the included webpages. There were various examples of webpages containing phrases like «research has shown that...» and «studies reveal that....» without any reference in the text. This was especially the case for the low glycemic diets webpages. For example the following statements were not accompanied by a reference:

> Flere langtidsstudier viser at det ikke er noen forskjell mellom vektnedgangen til de som har spist lavglykemisk og de som har fulgt en vanlig mager kost. [#6]
>
> *Various long-term studies show that there are no difference in weight loss between those on a low glycemic diet and those who have followed a standard energy-reduced diet.*
>
> Dietten er relativt godt dokumentert gjennom forskning, og har blant annet vist seg å gi en bedre stoffskifteregulering for de som lider av insulinresistens og overvekt. [#26]
>
> *The diet is scientifically rather well documented, and has shown to improve metabolism in insulin resistance and overweight.*

An overview of the different sources used across webpages is displayed on Table 17. Because some webpages referred to more than one type of source, the number of webpages in the table does not correspond with the total number of included webpages.

Table 17: ***Types of evidence used across webpages***

	No. of webpages	Accurate webpages	(%)	Inaccurate webpages	(%)
Guideline					
All webpages	**10**	**8**	**(80)**	**2**	**(20)**
Low glycemic diets	-	-		-	
Calcium supplementation	10	8	(80)	2	(20)
Systematic review					
All webpages	**2**	**-**	**-**	**2**	**(100)**
Low glycemic diets	-	-	-	-	
Calcium supplementation	2	-	-	2	(100)
Single study					
All webpages	**3**	**2**	**(67)**	**1**	**(33)**
Low glycemic diets	3	2	(67)	1	(33)
Calcium supplementation	-	-	-	-	
Expert or textbook					
All webpages	**24**	**10**	**(42)**	**14**	**(58)**
Low glycemic diets	9	1	(11)	8	(89)
Calcium supplementation	15	9	(60)	6	(40)
Personal opinion					
All webpages	**18**	**7**	**(39)**	**15**	**(61)**
Low glycemic diets	13	4	(31)	9	(69)
Calcium supplementation	5	3	(60)	2	(40)
Other					
All webpages	**3**	**-**	**-**	**3**	**(100)**
Low glycemic diets	1	-	-	1	(100)
Calcium supplementation	2	-	-	2	(100)
Not available (no source or author)					
All webpages	**7**	**4**	**(57)**	**3**	**(43)**
Low glycemic diets	5	3	(60)	2	(40)
Calcium supplementation	2	1	(50)	1	(50)

The most frequently used sources were experts and personal opinion, in that order. Twenty-four of all webpages referred to experts, of which 10 (42%) were evaluated to be accurate. Eighteen webpages used personal opinion, of which 7 (39%) were considered accurate. The overall results for the two source types are influenced by the results for the low glycemic diets. In calcium supplementation, webpages using these sources to a larger extent were accurate than inaccurate.

None of the webpages on low glycemic diets referred to guidelines. In calcium supplementation, national nutritional guidelines were mentioned in ten of the included webpages. Most of these pages were accurate (80%). Furthermore, two webpages referred to a systematic review. Both pages gave information about calcium supplementation and both referred to the specific Cochrane review used as a criterion standard in this study (Hofmeyr et al., 2006), yet both were inaccurate. The inaccuracies were due to their statements about the effects of calcium supplements in high risk pregnancies. We evaluated the statements to express supplements to be «effective» for this sub-group. The Cochrane review results for the high-risk sub-group were coded «probably effective». Consequently, the statements did not meet with the criterion standard.

Three webpages, all of them about low glycemic diets, cited and provided references to single studies. Two of the pages were evaluated to be accurate.

6.3.6 WEB DESIGN CRITERIA

I evaluated two web design criteria; the results for the included webpages are displayed in Table 18.

Table 18: ***Webpages' compliance with the design criteria***

		No. of web pages	(%)	Accurate web pages	(%)	Inaccurate web pages	(%)
Disclosure of advertising (P)							
All web pages	**Yes**	**37**	**(66)**	**16**	**(70)**	**21**	**(64)**
	No	**19**	**(34)**	**7**	**(30)**	**12**	**(36)**
Low glycemic diets	Yes	18	(58)	8	(80)	10	(48)
	No	13	(42)	2	(20)	11	(52)
Calcium supplementation	Yes	19	(76)	8	(62)	11	(92)
	No	6	(24)	5	(38)	1	(8)
Internal search engine present (S)							
All web pages	**Yes**	**48**	**(86)**	**20**	**(87)**	**28**	**(85)**
	No	**8**	**(14)**	**3**	**(13)**	**5**	**(15)**
Low glycemic diets	Yes	26	(84)	9	(90)	17	(81)
	No	5	(16)	1	(10)	4	(19)
Calcium supplementation	Yes	22	(88)	11	(85)	11	(92)
	No	3	(12)	2	(15)	1	(8)

S indicates evaluated at web site level; **P** indicates evaluated at web page level.

As stated in Section 6.1 most of the identified webpages came from commercial providers (88 %). Although many of the commercial websites clearly distinguished between content and advertisements, some webpages were heavily condensed with banner ads and other types of sponsored information. In one-third (34%) of the webpages, advertisements were not clearly distinguished from the text. This was more frequently seen in the low glycemic diets webpages (42 %) than in the calcium supplementation pages (24%).

Although the accurate pages proved better than the inaccurate pages at disclosing advertisements, the difference is not substantial: 70% of the accurate webpages complied with the criterion compared to 64% of the inaccurate pages. Nevertheless, when looking at the two topics separately, there are some contrasting results. In the case of low glycemic diets, eight

of ten webpages (80%) evaluated as accurate complied with the advertisement criterion, compared to just under half (48%) of the inaccurate webpages. In calcium supplementation it was the opposite way, with more inaccurate (92%) than accurate pages (62%) disclosing advertisements.

Most websites provided a search engine and accordingly a large portion of the included webpages complied with this design criterion (86%). There were no differences between the accurate and inaccurate webpages in this respect.

6.4 THE HEALTH PORTAL "SMIL" VERSUS SEARCH ENGINES

One of my objectives was to see if the webpages, and accordingly information on these webpages, identified via the health World Wide Web portal SMIL was more accurate than webpages retrieved by the search engines. My searches identified nine relevant webpages from SMIL. Six of these webpages overlapped the search engine results, all of them were evaluated as being inaccurate. Of the three unique webpages retrieved from SMIL, two were evaluated as inaccurate. Because of the small sample of pages identified via the SMIL portal, a comparison to the search engine results was not practical.

6.5 PUBLIC AGENCY SITES VERSUS COMMERCIAL SITES

Table 19 displays the distribution of accurate and inaccurate webpages across provider types. Forty-nine webpages (88%) were hosted by commercial providers, 21 (43%) of them were evaluated to be accurate. Six webpages were hosted by a total of three non-for-profit organisations. The non-for-profit organisations included two patient interest groups and one

large umbrella organisation for different patient groups and associations. None or the non-for-profit webpages were evaluated to be accurate.

Only one of the included webpages was hosted by a public agency. Accordingly, there was no opportunity to make meaningful comparisons between provider types.

Table 19: ***Accuracy by provider types***

Provider type	N	Accurate webpages	Inaccurate webpages
Commercial	**49**	21	28
Public agency	**1**	1	0
Non-for-profit organisations	**6**	0	6

7. DISCUSSION

My search identified a total of 56 webpages that were evaluated with regard to their information credibility, more specifically the accuracy of information compared to a criterion standard. They were also evaluated on their surface credibility by using a set of validated technical criteria commonly used to assess *transparency*. A large majority of the webpages were hosted by commercial providers; only one of the included pages came from a public agency website. Consequently, it was not possible to evaluate whether public agencies provide more accurate information than commercial agents.

A possible explanation to the predominance of commercial web sites could be the selection of topics. Dieting and weight loss is a "commercial" topic that sells in the news as well as on the World Wide Web. The results from my search engine searches add to this conclusion. A substantial number of the excluded webpages were commercial product sites. Moreover, many of the webpages were articles about dieting and fitness from generic online news services, more so than was the case in calcium supplementation. The use of calcium supplementation in pregnancy does not qualify as a "hot topic". Nevertheless, pregnant women increasingly use the Web to Search information regarding their own as well as the child's health and well-being. This is reflected by the number of commercial websites and discussion forums for pregnant women available via the Web. If I had chosen more clinical topics, for example specific interventions related to serious conditions such as cancer and heart disease, or chronic diseases such as rheumatic arthiritis and asthma, I might have retrieved a larger number of public agency sites.

The number of webpages retrieved via the health portal SMIL did not allow for any comparisons with webpages retrived via the general search engines. This could be due to the lack of updating of the portal during the time of my investigation.

In the next sections I will discuss my findings related to information credibility and surface credibility on Norwegian lay health webpages in more detail.

7.1 ACCURACY OF WEB PAGES (INFORMATION CREDIBILITY)

Many of the included webpages in is this study were hosted by sites with a clear aim of providing visitors with accurate and comprehensible information. Despite these good intentions, only above 40% of the webpages were evaluated to be accurate. In other words, the information credibility was generally low. Accordingly, my findings are similar to other studies that have used systematic reviews as a criterion standard for evaluating Web accuracy (Coulter et al., 2006; Eysenbach et al., 2002).

There were some differences between the two topics though. The low glycemic webpages tended to be more inaccurate than the calcium supplementation pages; just one-third of the pages was evaluated as accurate. This is in line with previous research, where prevalence figures have suggested that information on the Web about healthy eating tend to be more inaccurate than information on other issues, such as cancer therapy or management of rheumatoid arthritis (Coulter et al., 2006; Eysenbach et al., 2002). However, cautions must be made about generalising to the whole population of Norwegian webpages about healthy eating. My study looked at one particular type of diet, and only involved one topic for comparison. Furthermore, even if the proportion of accurate webpages with information about calcium supplements was higher compared with that of low glycemic diets, half the calcium webpages were still inaccurate.

A possible explanation of the frequent inaccuracy in webpages about low glycemic diets could be the lack of consensus on the evidence base regarding dietary interventions (Dommerud 2003; Poleszynski & Mysterud 2008). Thus, the information provided via the World Wide Web and other media varies in the way effects are expressed, and it is often imbalanced. A frequent reason for inaccuracy in the low glycemic diets webpages was that authors overestimated the effect of the diets. At the other end of the scale, a few webpages underestimated the weight reducing effects of low glycemic diets. In one instance, the same website provided contradictory messages about the impacts of diets on weight loss. Moreover, the uncertainties relating to the clinical evidence for long-term effects were hardly ever mentioned. In only two webpages did the Web operator or author correctly emphasise the

«unknown effect» of following a low glycemic diet over time. In other words, the information presented was often incomplete.

A lack of consensus in the evidence base is reflected by the challenges in conducting large, long-term randomised controlled trials for lifestyle interventions, such as healthy eating (Ashcroft et al., 2001, p. 54). Such interventions are demanding, claiming participants to change their lifestyle and also to deal with the potential impact of the lifestyle change in their immediate environment. Accordingly, the evidence base is weaker, because trials are small and only spans over a limited time. The randomised controlled trials in the Cochrane review about low glycemic diets included only a few hundred participants (Thomas et al., 2007). In comparison, the review on calcium supplementation included over 15 000 women (Hofmeyr et al., 2006). Testing a medication, or as in this case, a nutritional supplement, is deemed less complicated due to a large sample of participants to start with, and the relatively easy distribution of the intervention.

I did not find the same examples of overestimating, or underestimating the effect when evaluating the webpages about calcium supplements in pregnancy. Rather, inaccuracy was often due to the unclarity of the effect statements. Sometimes the statements were vague due to their wording; statements contained words equivalent to the English adverbs “maybe” and “perhaps” (in Norwegian: “kanskje” and “kan hende”). The vagueness of statements was sometimes reflected by little agreement among raters. This compares to the findings of the study by Glenton et al. (2005); they found many examples of unclear information across webpages. The various interpretations and difficulties in coding the statements, indicate that others may have similar problems in understanding the information provided by different Norwegian Web sites.

Inaccuracy was also related to the *granularity* of the information provided. A few of the calcium supplementation webpages lacked details about the required calcium intake in pregnancy, defined as 900 milligrams a day. As a result, their effect statements were coded as «unclear». On could of course question my decision to code these webpages as inaccurate. In comparison, webpages about low glycemic diets were not evaluated as unclear even if many of them lacked information about short and long term effects on weight reduction. However, the specified limit value of 900 milligrams was clearly stated in the Cochrane review

(Hofmeyr., et al 2006). It is also in line with the nutritional recommendations on calcium intake in pregnancy from the Directorate of Health and Social Affairs (2005b). Moreover, statements like «if you don't drink enough milk, you should take a calcium supplement» or «if your calcium intake is low, make sure to take a calcium supplement» are not only incomplete, they are also unclear. In turn, they raise new questions as to what is enough milk and what is a low calcium intake.

Unclarity and lack of detailing in the statements was due to that most webpages described effects qualitatively, rather than quantitatively. For example, no webpages about low glycemic diets included quantitative statements such as: «people who follow a low glycemic diet for three months are likely to lose one more kilogramme of weight than if they choose other diets». This may be due to that Web operators or authors of information assume a qualitative presentation is easier to understand, or that they do not want to express more than they know. Yet, as Glenton et al. (2005) point out, qualitative descriptions mean different things to different people. This was apparent in the variations of coding among the raters for some of the statements.

There was high agreement among raters when coding the results of the review. The coding of Web statements proved more difficult. It may be that involving a larger panel of people in coding the statements would have led to more consistent results. Variations in coding could also be explained by different interpretations of the coding scheme. For example, there is a thin line between the codes «effective» and «probably effective»; and opposite, «not effective» and « probably not effective». This was reflected during the coding process, where raters interpreted the auxiliary verb "kan" differently. Statements including terms like «kan være nyttig» («can/may be useful»), «kan hjelpe» («can/may help») were by some evaluated as expressing a definite effect («effective»). Others evaluated them to express a more moderate effect («probably effective»). A further validation and operationalisation of the coding scheme, preferably with lay people involved in the validation process, could help prevent inconsistent coding.

The results of this study are subject to bias because they involved subjective assessments with regard to how the webpages were selected, how the results of the Cochrane reviews were coded, and how the different Web statements were evaluated. I selected and included

webpages without input from others. Furthermore, even if I involved my colleagues in the rating, I decided on the final coding of the reviews and the webpages. The outcomes of this study might have been different if someone else had performed the searches, selected pages for inclusion, and accomplished the final coding. Moreover, the coding of the webpages have looked different if lay people had been involved in the process.

7.2 COMPARING ACCURACY AND COMPLIANCE WITH TECHNICAL CRITERIA

In general, there were no clear associations between accuracy and the technical, surface criteria, although for many of the web content criteria there was a tendency of better compliance among the accurate web pages.

All web pages, irrespective of being accurate or not, complied with the criteria «Disclosure of ownership» and «Feedback mechanism provided». I could probably have omitted the former criterion from the analysis because I only included webpages hosted by sites that disclosed ownership. Nevertheless, my search only identified one relevant Web page with no ownership information. Specifying who's behind a Web site seems to be common procedure among non-private Web site operators. Thus, the ownership criterion may not be useful for discriminating between accurate and inaccurate Web Pages hosted by these operators.

The criteria «Editorial review process», «General disclaimers provided», or «Internal search engine» were not found to be appropriate discriminators of accurate and inaccurate information. I found no real differences in compliance with these criteria when I compared the accurate webpages to the inaccurate pages. It is not surprising that functionality features, such as providing a search engine, is no real indicator of accurate information. This has already been suggested by other researchers (O'Grady, 2006; Stanford et al., 2002). One might think that Web sites that have a clear editorial policy will provide more accurate content, particularly where external experts and specialists are involved in the peer review process. The relatively large proportion of inaccurate webpages in total, and the even proportions of accurate and inaccurate pages complying with the editorial criterion, indicates that this might not be the case.

For the criteria «statement of purpose», «disclosure of authorship», «authors credentials disclosed», «date of creation disclosed», «sources clear», and «disclosure of advertising» there was higher compliance among the accurate webpages compared to the inaccurate pages. This may indicate that webpages that clearly display the authors' name, the creation date, details of the sources used, and have a less commercial «touch» are more credible. Nevertheless, the differences between the proportions of accurate and inaccurate webpages that met with the above mentioned criteria were not substantial. Thus, I will be careful in drawing any firm conclusions. Other studies that have used more rigorous methods for identifying associations, did not find any significant relationships between accuracy and technical, surface criteria (Bernstam et al., 2008; England & Nicholl, 2004).

When looking at the two topics separately, the webpages about calcium supplementation complied better with the technical criteria than did the low glycemic diets web pages. Yet, for calcium supplementation, compliance was in favour of the inaccurate webpages for several of the criteria. For example, a higher proportion of the inaccurate webpages displayed a general disclaimer and disclosed authors' credentials compared to the accurate webpages. Nevertheless, it would be senseless to suggest that webpages providing a general disclaimer are inaccurate. Rather, the conclusion would be that the disclaimer criterion is not a good discriminator for identifying accurate information.

There was a tendency among the accurate low glycemic index webpages to better comply with the technical criteria, when compared to the inaccurate webpages. For instance, the accurate pages were better at distinguishing between advertisements and information content than were the inaccurate pages. There were also positive associations between displaying authors' credentials and accuracy. This may indicate that a webpage providing information about dieting and weight loss will be more accurate if advertisements are clearly differentiated from content, and the content is written by a qualified author. Yet, for many of the criteria there were no real differences between accurate and inaccurate webpages. Moreover, the number of accurate webpages was low (n=10). Thus, any differences in favour of the accurate webpages could be due to chance. This also applies to the population of calcium webpages, where the number of accurate and inaccurate pages was 13 and 12 respectively.

Despite the lack of any apparent associations between the surface credibility criteria and the credibility of the actual information content (information credibility), the surface criteria are considered to be important for a number of reasons. Firstly, their presence is deemed ethical by a number of organisations that focus on credibility of health information (Commission of the European Communities, 2002). Secondly, they are helpful when it comes to creating context and avoiding misunderstandings (e.g. «statement of purpose» and «general disclaimers»). Thirdly, they enable lay people to validate the information themselves (e.g. «sources clear», «feedback mechanism provided»). However, it may not be sufficient to evaluate the accuracy of information (information credibility) simply by investigating appearance features (surface credibility).

Because the focus of my study was on the informational content of lay health webpages, I will discuss the criteria in the web content category in some more depth.

7.2.1 WEB CONTENT: AUTHORSHIP

Most webpages provided enough information about who authored the information («disclosure of authorship»). A majority of these webpages disclosed the authors' affiliation and credentials. My study show more positive results in this respect compared to a similar study that used the same criteria (Bernstam et al., 2008). I decided to evaluate the author attributions' criteria on website, rather than webpage level. This may explain the relatively high compliance among the webpages in my study.

Where details about authors' credentials were available, most information was written by health professionals. In only a small proportion of the webpages (n=8), the information was authored by a physician. This probably reflects the fact low glycemic diets and calcium supplementation are dietary interventions. Thus, in many instances the information was authored by a dietician rather than a physician. Accordingly the criterion « physicians' credentials disclosed» was not applicable.

There is an assumption that information written by health professionals can be trusted. Although there was a small difference in favour of the accurate webpages and compliance with the authors' credentials criterion, there was no association between accuracy and the information being written by a health professional. Altogether, this indicates that health information on the Web is not necessarily accurate even if written by someone with a health background. In one instance, the same dietician gave inconsistent advice about calcium supplementation in pregnancy, with only a few days interval.

7.2.2 WEB CONTENT: CURRENCY

Only a moderate proportion (63%) of the included webpages disclosed when the information was produced («Date of creation disclosed»). Furthermore, very few webpages displayed a date of last update (n=13). Although more of the accurate web pages displayed a creation date, the difference compared to the inaccurate webpages was not substantial. Accordingly, cautions should be made in suggesting that displaying a creation date indicates the content being accurate. Due to the small number of pages, it was not possible to compare accuracy to compliance with the criterion «Date of last update».

When I compared the currency of the Web pages to the publication of their respective Cochrane reviews, I found that a higher proportion of the inaccurate webpages were created before the review was published. The opposite was the case for the webpages created or updated after their respective review was published. It may have been wrong to compare information on older webpages to the results of relatively new Cochrane reviews. However, this indicates that there are many out-dated web pages on the Web that is not in line with results from the summarised research evidence. A few of the older webpages were links to questions-and-answers-services. Obviously, such webpages will not be updated. However, providers of such services should put a "date-stamp" on older responses, warning end-users that medical practice and knowledge might have changed after the advice was published. The fact that links to these services appear on the first results page of search engines, indicates that lay people will come across them when seeking out health information on the Web.

Another point is the value of a currency criterion that simply asks for the appearance of a date. Regarding the currency of information, the criterion "Does the webpage display the date on which it was last updated?" is not as valuable as "has the site been updated in the last six months?" In the LIDA-instrument it is suggested that for treatment information, the ideal target is six monthly updates; yet for diagnosis and background information, including risk factors, it can be longer (Minervation, 2007).

7.2.3 WEB CONTENT: ATTRIBUTION

Only a little over half the webpages disclosed the sources behind the information provided. There were more accurate web pages than inaccurate pages that backed their claims up with some type of source («sources clear»). Yet, as for the other Web content criteria, there were no substantial differences in this respect. Again, *transparency* with regard to the sources used does not necessarily result in accurate information. Nevertheless, it helps the people visiting the site to get a better idea about how the materials were compiled.

Very few webpages that displayed a source did so in terms of a conventional reference, despite some obvious examples where references should have been used to back up claims research findings. I came across this more often on the low glycemic diets pages than the calcium supplements pages. Where references were used, they referred to single studies. Sometimes the same webpage provided contradictory information. Referrals to research could be in favour of the intervention («low glycemic diets are probably effective for weight reduction»), or they were at the opposite end. On one hand, presenting information from different angles could be viewed as a way of gaining credibility; this may lead to lay people perceiving the Web operator as balanced and nuanced. On the other hand, a fragmented reporting of research results with and without references, may lead to more confusion than clarification for those in search of information.

Experts and personal opinion was used as the main source of information in most pages, with experts ranging at the top. In the cases where experts were used as a source or as an author of information, more web pages were found than inaccurate. This result may challenge the belief

in field experts as credible sources of information. However, the low number of webpages means that the results should be interpreted with caution.

Only two webpages used systematic reviews as a source of information. In both cases, the referenced review was the Cochrane review on calcium supplementation (Hofmeyr et al., 2006). In both cases the information was evaluated as inaccurate. This indicates that even if summarised research evidence is used to compile the information, accuracy is depending on how the results are presented.

7.3 CONTENT DEVELOPMENT PROCEDURES

A majority of the included webpages were hosted by sites that used experts or specialists to author or review content. When I examined the websites' editorial policies, I found statements that clearly linked the use of experts to the credibility of the content. This could be linked to reputed credibility; if the information is written by recognised experts or professionals with domain knowledge, then it is more likely that it will be credible. However, the use of experts may not always result in accurate information. Inaccuracies were frequently seen in webpages that were authored by experts, or where experts were used as a source of information.

Keeping information up to date can be a challenge given the pace at which new research results is published in some clinical fields, and there is a risk that writers who do not conduct systematic literature searches will fail to include the most up-to-date evidence. A lot of the webpages were created before the publication of their respective systematic review, and none of the websites explicitly mentioned how they searched for new evidence. Again, relying on a single expert's knowledge may not be sufficient to ensure up to date information. Only one of the included websites in this study specifically mentioned that they used an evidence-based approach to compile and revise information. Nevertheless, the information on the respective web pages was not accompanied by reference. One might expect that employing an evidence-based policy to content development at least would results in referencing the statements. Nevertheless, both webpages from this site were evaluated as accurate. They were also two of the few pages that provided both a date of creation and of update.

It should be noted that details about the content development process of a Web operator, including literature searching and critical appraisal of research evidence, may not be possible to identify merely by looking at the site. Rather, the Web operators need to be contacted directly to find out in detail how their content production methods are. I only investigated the editorial information on the different Web operators' sites. Telephone interviews with the operators, as was done by Coulter et al. (2006), would probably have identified more details around their content development procedures.

8. CONCLUSION

To my knowledge this is the first study that has evaluated the credibility of Norwegian-language lay health information on the World Wide Web by using a criterion standard. Furthermore, I used systematic reviews as a criterion standard because they are considered to be a reliable and accurate source of information. This study was limited in scope in that only two small topics were investigated. However, the conclusions reached are broadly in line with other studies of lay health Web pages that have compared information credibility to the results of systematic reviews.

Overall, I only found 40% of the webpages to provide accurate information. The information was often unclear, incomplete, and inconsistent. Moreover, webpages were frequently outdated. However, their compliance with commonly used technical criteria such as authorship, disclosure of purpose, and disclosure of advertising, was moderate to high. Compliance with the technical criteria was evenly distributed among the accurate and inaccurate webpages. Even if compliance was sometimes better among the accurate webpages, the differences were not substantial. For some criteria there were no differences between accurate and inaccurate web pages in how they met with the criteria. This indicates that investigating appearance features (surface credibility) may not be the best solution to determine accuracy (information credibility) of lay health content on the Web.

It was not possible to evaluate whether public agencies provide more accurate information than commercial providers because my search only retrieved one relevant, public webpage. Moreover, the low number of webpages retrieved from SMIL did not allow any comparisons with the webpages retrieved from the search engines searches.

Most webpages were hosted by sites that were highly expert-based; experts or specialists were used to author or review content. The number of inaccurate webpages indicates that the use of experts may not always indicate that the information is credible.

8.1 IMPLICATIONS FOR PRACTICE

Many have pointed to the fact that lay people need to be educated about Web credibility and health information. Many librarians use technical, domain-independent criteria related to authorship, currency, attribution, and other disclosures when educating lay people in identifying credible information. However, evaluating surface credibility may not be sufficient to identify accurate information. Independent comparison with the most up to date facts obtained elsewhere is preferable.

A source of comparison would be systematic reviews. In Norway, health professionals have access to sources like the Cochrane Library and Clinical Evidence via the National electronic Health Library. These sources are also available to patients and other lay people. A new strategy for patient education and library instruction would be to introduce these sources to lay people. However, systematic reviews are often lengthy and difficult to understand. Patients and the general public should be provided with tools to enable them to understand and evaluate systematic reviews. In that way, information obtained from general search engines could be compared to the results of summarised research, even by lay people.

8.2 IMPLICATIONS FOR RESEARCH

This was a small-scale study which only involved two specific topics for evaluation. My findings might have looked different with a larger sample of web pages, and various comparison topics. Future studies on Norwegian lay health information on the Web should involve more topics to be compared. Moreover, it would be useful to select topics that will identify information from public agencies such as hospitals and health authorities. My study only involved one public webpage. Maybe the topics need to be more clinically oriented

Future studies should also use better methods for establishing relationships between surface credibility criteria and information credibility. Such methods include an analytical approach, rather than the descriptive cross-sectional approach used in this study. Moreover, a longitudinal study of webpages and websites will be able to track changes over time.

The coding scheme used for assessing information about the effectiveness of interventions on the Web needs further validation. Validation should include lay peoples' assessments of Web sites.

REFERENCES

Air, M., Roman, S. A., Yeo, H., Maser, C., Trapasso, T., Kinder, B., et al. (2007). Outdated and incomplete: a review of thyroid cancer on the World Wide Web. *Thyroid, 17*(3), 259-265.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall

Andreassen, H. K., Bujnowska-Fedak, M. M., Chronaki, C. E., Dumitru, R. C., Pudule, I., Santana, S., et al. (2007). European citizens' use of E-health services: a study of seven countries. *BMC Public Health, 7*, 53.

Andreassen, H. K., Sandaune, A. G., & Gammon, D. (2002). Norwegian's use of Internet health services. *Tidsskrift for Den Norske Laegeforening, 122*(17), 1640-1644.

Andreassen, H. K., Wangberg, S., Wynn, R., Sørensen, T., & Hjortdahl, P. (2006). Health-related use of the Internet in the Norwegian population. *Tidsskrift for Den Norske Laegeforening, 126*(22), 2950-2952.

Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA, 268*(2), 240-248.

Ashcroft, R., Hope, T., & Parker, M. (2001). Ethical issues and evidence-based patient choice. In A. Edwards & G. Elwyn (Eds.), *Evidence-based patient choice: inevitable or impossible?* Oxford: Oxford University Press.

Badgett, R. G., & Lawrence, J. C. (1998). Internet health ratings systems: knowledge vs Babel. *JAMA, 280*(8), 697-698.

Benbassat, J., Pilpel, D., & Tidhar, M. (1998). Patients' preferences for participation in clinical decision making: a review of published surveys. *Behavioral Medicine, 24*(2), 81-88.

Bernstam, E. V., Sagaram, S., Walji, M., Johnson, C. W., & Meric-Bernstam, F. (2005). Usability of quality measures for online health information: Can commonly used technical quality criteria be reliably assessed? *International Journal of Medical Informatics, 74*(7-8), 675-683.

Bernstam, E. V., Shelton, D. M., Walji, M., & Meric-Bernstam, F. (2005). Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *International Journal of Medical Informatics, 74*(1), 13-19.

Breckons, M., Jones, R., Morris, J., & 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,* e3, from http://www.jmir.org/2008/1/e3/

Centre for Reviews and Dissemination. (2001). Systematic reviews: CRD's guidance for undertaking reviews in health care. Retrieved February 14, 2009 from http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews_.pdf

Chalmers, I. (2001). Invalid health information is potentially lethal. *BMJ, 322*(7292), 998.

Childs, S. (2004). Developing health website quality assessment guidelines for the voluntary sector: outcomes from the Judge Project. *Health Information & Libraries Journal, 21*, 14-26.

Cochrane Collaboration. An introduction to Cochrane reviews and The Cochrane Library. Retrieved February 14, 2009 from http://www.cochrane.org/reviews/clibintro.htm

Commission of the European Communities. (2002). eEurope 2002: Quality criteria for health related websites. *Journal of Medical Internet Research, 4*, 72-91, from http://www.jmir.org/2002/3/e15/

Coulter, A., Ellins, J., Swain, D., Clarke, A., Heron, P., Rasul, F., et al. (2006). Assessing the quality of information to support people in making decisions about their health and healthcare. Retrieved February 14, 2009, from http://www.pickereurope.org/Filestore/Downloads/Health-information-quality-web-version-FINAL.pdf

Crocco, A. G., Villasis-Keever, M., & Jadad, A. R. (2002). Two wrongs don't make a right: harm aggravated by inaccurate information on the Internet. *Pediatrics, 109*(3), 522-523.

Davies, H. T. (2001). What are confidence intervals? *Bandolier, 3*, 8. Retrieved October 10, 2008, from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_are_Conf_Inter.pdf

Dawes, M., Summerskill, W., Glasziou, P., Cartabellotta, A., Martin, J., Hopayian, K., et al. (2005). Sicily statement on evidence-based practice. *BMC Medical Education, 5*. Retrieved Jan 5, from http://www.biomedcentral.com/1472-6920/5/1

Dickersin, K., Straus, S. E., & Bero, L. A. (2007). Evidence based medicine: increasing, not dictating, choice. *BMJ, 334 Suppl 1*, s10.

Directorate of Health and Social Affairs. (2005a). *...And it's going to get better: national strategy for quality improvement in social and health services (2005-2015)* [Report in Norwegian]. Retrieved February 14, 2009, from http://www.helsedirektoratet.no/vp/multimedia/archive/00004/IS-1162_4390a.pdf

Directorate of Health and Social Affairs. (2005b). *Norwegian recommendations for nutrition and physical activity [Report in Norwegian]*. from http:http://www.helsedirektoratet.no/vp/multimedia/archive/00002/IS-1219_2606a.pdf

Directorate of Health and Social Affairs. (2006). *National strategy for quality improvement in social and health services. Priority area Strengthening the basis for continuous quality improvement in health and social professions education: working group recommendations [Report in Norwegian]*. Retrieved February 14, 2009, from http://www.ogbedreskaldetbli.no/237/6_SHdir_SFK_Styrke_Forbedringskunnskapen.pdf

Dommerud, T. (2003). Ja til brød og potet [Yes to bread and potatoes] [. *Dagens medisin*. Retrieved February 14, 2009 from http://www.dagensmedisin.no/nyheter/2003/02/13/ja-til-br-d-og-potet/index.xml

Eddy, D. M. (2005). Evidence-based medicine: a unified approach. *Health Affairs, 24*(1|), 9-17.

Eiring, O., & Tvedten, A. (2007). Internett på resept: Hvilke nettsteder kan vi anbefale pasientene? [Internet on prescription: which home pages can we recommend to patients?]. *Utposten, 36*(4), 25-29.

Elwyn, G., & Edwards, A. (2001). Evidence-based patient choice? In A. Edwards & G. Elwyn (Eds.), *Evidence-based patient choice: inevitable or impossible?* . Oxford: Oxford University Press.

England, C. Y., & Nicholls, A. M. (2004). Advice available on the Internet for people with coeliac disease: an evaluation of the quality of websites. *Journal of Human Nutrition & Dietetics, 17*(6), 547-559.

Entwistle, V., & O'Donnell, M. (2001). Evidence-based health care: what roles for patients. In A. Edwards & G. Elwyn (Eds.), *Evidence-based patient choice: inevitable or impossible?* . Oxford: Oxford University Press.

Entwistle, V. A., Sheldon, T. A., Sowden, A., & Watt, I. S. (1998). Evidence-informed patient choice. Practical issues of involving patients in decisions about health care technologies. *International Journal of Technology Assessment in Health Care, 14*(2), 212-225.

European Commission. (2005). eUser population survey 2005 (eHealth Chart 1). Retrieved February 14, 2009, from http://www.euser-eu.org/Document.asp?MenuID=68

Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis)information. *American Journal of Medicine, 113*(9), 763-765.

Eysenbach, G., & Diepgen, T. L. (2001). The role of e-health and consumer health informatics for evidence-based patient choice in the 21st century. *Clinics in Dermatology, 19*(1), 11-17.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ, 324*, 573-577.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA, 287*(20), 2691-2700.

Farrell, K. D., Robinson, L. M., Baydock, S. A., Farrell, S. A., Irving, L. E., & O'Connell, C. M. (2006). A survey of Canadian websites providing information about female urinary incontinence. *Journal of Obstetrics & Gynaecology Canada: JOGC, 28*(8), 700-712.

Fogg, B. J. (2003). *Persuasive technology: using computers to change what we think and do*. San Fransisco: Morgan Kaufmann

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001). What makes web sites credible: a report on a large quantitative study. *CHI 3*(1), 61-68.

Fox, S. (2006). Online Health Search 2006. *PEW Internet & American Life Project* Retrieved February 14, 2009, from http://www.pewinternet.org/PPF/r/190/report_display.asp

Gagliardi, A., & Jadad, A. R. (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), 569-573.

Glenton, C. (2006). *Enabling evidence-informed patient choice.* Unpublished Ph.D dissertation, Faculty of Medicine, University of Oslo, Oslo.

Glenton, C., Nilsen, E. S., & Carlsen, B. (2006). Lay perceptions of evidence-based information: a qualitative evaluation of a website for back pain sufferers. *BMC Health Services Research, 6*, 34, Retrieved February 14, 2009,from http://www.biomedcentral.com/1472-6963/6/34

Glenton, C., Paulsen, E. J., & Oxman, A. D. (2005). Portals to Wonderland: health portals lead to confusing information about the effects of health care. *BMC Medical Informatics & Decision Making, 5*, 7. Retrieved Mar 15, from http://www.biomedcentral.com/1472-6947/5/7

Glenton, C., Underland, V., Kho, M., Pennick, V., & Oxman, A. D. (2006). Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. *Journal of Clinical Epidemiology, 59*(8), 770-778.

Griffiths, K. M., & Christensen, H. (2002). The quality and accessibility of Australian depression sites on the World Wide Web. *Medical Journal of Australia, 176 Suppl*, S97-S104.

Griffiths, K. M., & Christensen, H. (2005). *Website Quality Indicators for Consumers*. Retrieved February 14, 2009 from http://www.jmir.org/2005/5/e55/

Guyatt, G., Cook, D., & Haynes, B. (2004). Evidence based medicine has come a long way. *BMJ, 329*(7473), 990-991.

Harland, J., & Bath, P. (2007). Assessing the quality of websites providing information on multiple sclerosis: evaluating tools and comparing sites. *Health Informatics Journal, 13*(3), 207-221.

Haynes, B. (2007). Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evidence-Based Nursing, 10*(1), 6-7.

Health on the Net Foundation. (1997). HON Code of Conduct (HONcode) for medical and health Web sites. Retrieved February 14, 2009, from http://www.hon.ch/HONcode/Conduct.html

Higgins, J., & Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1*. Retrieved February 14, 2009, from www.cochrane-handbook.org

Hofmeyr, G. J., Atallah, A. N., & Duley, L. (2006). Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. *Cochrane Database of Systematic Reviews*(3).

Houle, T. T., & Stump, D. A. (2008). Statistical significance versus clinical significance. *Seminars in Cardiothoracic and Vascular Anesthesia, 12*(1), 5-6.

Jadad, A. R., & Gagliardi, A. (1998). Rating health information on the Internet: navigating to knowledge or to Babel? *JAMA, 279*(8), 611-614.

Jamtvedt, G., & Nortvedt, M. W. (2008). Kunnskapsbasert ergoterapi: et bidrag til bedre praksis [Evidence-based occupational therapy: a contribution to better practice]. *Ergoterapeuten* 10-18.

Jenkins, D. J., Wolever, T. M., Taylor, R. H., Barker, H., Fielden, H., Baldwin, A., et al. (1981). Glycemic index of foods: a physiological basis for carbohydrate exchange. *American Journal of Clinical Nutrition, 34*, 362-366.

Jørgensen, K. J., & Gøtzsche, P. C. (2004). Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ, 328*(7432), 148.

Lavoie, B., & Nielsen, H. F. (1999). Web characterization terminology and definitions sheet. Retrieved February 14, 2009, from http://www.w3.org/1999/05/WCA-terms/#page

Mantzoukas, S. (2008). A review of evidence-based practice, nursing research and reflection: levelling the hierarchy. *Journal of Clinical Nursing, 17*(2), 214-223.

McMullan, M. (2006). Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. *Patient Education and Counseling, 63*(1-2), 24-28.

Meric, F., Bernstam, E. V., Mirza, N. Q., Hunt, K. K., Ames, F. C., Ross, M. I., et al. (2002). Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. *BMJ, 324*(7337), 577-581.

Miles, A., Bentley, P., Polychronis, A., & Grey, J. (1997). Evidence-based medicine: why all the fuss? This is why. *Journal of Evaluation in Clinical Practice, 3*(2), 83-86.

Minervation. (2007). The LIDA instrument: Minervation validation instrument for health care web sites. Version 1.2. Retrieved February 14, 2009, from http://www.minervation.com/mod_product/LIDA/minervalidation.pdf

Moher, D., Tsertsvadze, A., Tricco, A. C., Eccles, M., Grimshaw, J., Sampson, M., et al. (2008). When and how to update systematic reviews. *Cochrane Database of Systematic Reviews*(1).

Monsen, T. (2007). Spis med livet som innsats [Eat for your life]. *Forskning.no* Retrieved February 14, 2009, from http://www.forskning.no/artikler/2007/september/1189514164.24

Morahan-Martin, J. M. (2004). How internet users find, evaluate, and use online health information: a cross-cultural review. *Cyberpsychology & Behavior, 7*(5), 497-510.

Muir Gray, J. A. (2001a). *Evidence-based healthcare. how to make health policy and management decisions* (2nd ed. ed.). London: Churchill Livingstone.

Muir Gray, J. A. (2001b). Evidence-based medicine for professionals. In A. Edwards & G. Elwyn (Eds.), *Evidence-based patient choice: inevitable or impossible?* . Oxford: Oxford University Press.

Muir Gray, J. A. (2008). Viva the revolution. *Health Information & Libraries Journal, 25*(Suppl 1), 96-98.

Munthe, A., & Nordnes, S. E. (2004). SMIL - Scandinavian Medical Information for Laypersons. *Scandinavian Public Library Quarterly*(2), 22-24.

National Committee for Research Ethics in the Social sciences and the Humanities. (2003, September 24). Research ethics guidelines for internet research Retrieved February 14, 2009, from http://www.etikkom.no/English/Publications/internet03/view_publikasjon

Nielsen, J. (2000). *Designing web usability: the practice of simplicity.* Indianapolis: New Riders Publishing.

Nilsson-Ihrfelt, E., Fjallskog, M.-L., Blomqvist, C., Ahlgren, J., Edlund, P., Hansen, J., et al. (2004). Breast cancer on the Internet: the quality of Swedish breast cancer websites. *Breast, 13*(5), 376-382.

Norem, J., & Moen, M.-A. (2004). The Websites of Norwegian hospitals: do they meet national guidelines and patient's expectations? *Journal of Telemedicine and Telecare, 10*(5), 272-276.

Norwegian Copyright Act. (1961). Act No. 2 of 12 May 1961 relating to copyright in literacy, scientific and artistic works, etc. Retrieved February 14, 2009, from http://www.kopinor.org/opphavsrett/norwegian_copyright_act

Norwegian Knowledge Centre for the Health Services. (2008). Retrieved February 14, 2009, from http://www.kunnskapssenteret.no/

O'Grady, L. (2006). Future directions for depicting credibility in health care web sites. *International Journal of Medical Informatics, 75*(1), 58-65.

Patients' Rights Act. (1999). The Act of 2 July 1999 No. 63 relating to Patients' Rights. Retrieved February 14, 2009, from http://www.regjeringen.no/upload/kilde/hod/red/2005/0079/ddd/pdfv/242636-patients'_rights_act.pdf

Poleszynski, D., & Mysterud, I. (2008). Karbohydratfattig kosthold må anbefales [Carbohydrate-restricted diet is to be recommended] *Tidsskrift for Den Norske Laegeforening, 128,* 1422.

Powell, J. A., Lowe, P., Griffiths, F. E., & Thorogood, M. (2005). A critical analysis of the literature on the Internet and consumer health information. *Journal of Telemedicine and Telecare, 11* (5), 41-43.

Powell, R., & Connaway, L. (2004). *Basic research methods for librarians* (4th ed.). Westport: Libraries unlimited.

Provost, M., Koompalum, D., Dong, D., & Martin, B. C. (2006). The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. *International Journal of Medical Informatics, 75,* 42-57.

Raben, A. (2002). Should obese patients be counselled to follow a low-glycaemic index diet? No. *Obesity Reviews, 3*(4), 245-256.

Renahy, E., & Chauvin, P. (2006). Internet uses for health information seeking: A literature review. *Revue d Epidemiologie et de Sante Publique, 54*(3), 263-275.

Rosenbaum, S. E., Glenton, C., & Cracknell, J. (2008). User experiences of evidence-based online resources for health professionals: user testing of The Cochrane Library. *BMC Medical Informatics and Decision Making, 8,* 34, from http://www.biomedcentral.com/1472-6947/8/34

Ruland, C. M. (2005). Shared decision making and nursing informatics research in Norway. *Applied Nursing Research, 18*(2), 70-72.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312*(7023), 71-72.

Santesso, N., Glenton, C., Strømme Nilsen, E., Rosenbaum, S. E., Rader, T., Pardo, J., et al. (2007). *Evaluation of a Plain Language Summary template for Cochrane Reviews*. Paper presented at the XV Cochrane Colloquium, Sao Paolo.

Seidman, J. J., Steinwachs, D., & Rubin, H. R. (2003a). Conceptual framework for a new tool for evaluating the quality of diabetes consumer-information Web sites. *Journal of Medical Internet Research, 5,* e29. Retrieved February 14, 2009, from http://www.jmir.org/2003/4/e29/

Seidman, J. J., Steinwachs, D., & Rubin, H. R. (2003b). Design and testing of a tool for evaluating the quality of diabetes consumer-information Web sites. J*ournal of Medical Internet Research,* 5(4):e30. Retrieved February 14, 2009, from http://www.jmir.org/2003/4/e30/

Shepperd, S., Charnock, D., & Cook, A. (2002). A 5-star system for rating the quality of information based on DISCERN. *Health Information & Libraries Journal, 19*(4), 201-205.

Shon, J., Marshall, J., & Musen, M. A. (2000). The impact of displayed awards on the credibility and retention of Web site information. *Proceedings / Amia Symposium*, 794-798.

Sieving, P. C. (1999). Factors driving the increase in medical information on the web--one American perspective. *Journal of Medical Internet Research, 1,* E3. Retrieved February 14, 2009, from http://www.jmir.org/1999/1/e3/

Stanford, J., Tauber, E. R., Fogg, B. J., & Marable, L. (2002). Experts vs. online consumers: a comparative study of health and finance web sites. Retrieved February 14, 2009, from http://www.consumerwebwatch.org/pdfs/expert-vs-online-consumers.pdf

St.meld. nr. 39 (1999). Forskning ved et tidsskille [White Paper 39/1999]. Retrieved February 14, 2009, from http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/19981999/Stmeld-nr-39-1999-.html?id=192405

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, B. (2005). *Evidence-based medicine : how to practice and teach EBM* (3rd ed.). Edinburgh Elsevier Churchill Livingstone.

Thomas, D. E., Elliott, E. J., & Baur, L. (2007). Low glycaemic index or low glycaemic load diets for overweight and obesity. *Cochrane Database of Systematic Reviews*(3).

Tonelli, M. R. (2001). The limits of evidence-based medicine. *Respiratory Care, 46*(12), 1435-1440; discussion 1440-1431.

Turner, A., Fraser, V., Muir Gray, J. A., & Toth, B. (2002). A first class knowledge service: developing the National electronic Library for Health. *Health Information & Libraries Journal, 19*(3), 133-145.

Twycross, A., & Shields, L. (2004a). Validity and reliability--What's it all about? Part 1. Validity in quantitative studies. *Paediatric Nursing, 16*(9), 28.

Twycross, A., & Shields, L. (2004b). Validity and reliability--what's it all about? Part 2 reliability in quantitative studies. *Paediatric Nursing, 16*(10), 36.

White, P. J. (2002). Evidence-based medicine for consumers: a role for the Cochrane Collaboration. *Journal of the Medical Library Association, 90*(2), 218-222.

Wyatt, J. C. (1997). Commentary: measuring quality and impact of the World Wide Web. *BMJ, 314*(7098), 1879-1881.

APPENDICES

APPENDIX 1: MEDLINE SEARCH STRATEGY

Database: Medline (via Ovid)

Search date: 04.12.2007

Strategy:

1. Patients/
2. Consumer satisfaction/
3. Consumer participation/
4. Patient education/
5. Health education/
6. Health Behavior/
7. Self-help groups/
8. (patient$1 or consumer$1 or lay people or lay use$2 or lay usage or public).tw.
9. evidence based patient choice.tw.
10. or/1-9
11. World Wide Web/
12. (World Wide Web or world wide web or www or website or website or web portal$ or ehealth or e-health).tw.
13. "web 2.0".tw.
14. (online adj3 health adj3 information).tw.
15. (online adj3 medical adj3 information).tw.
16. (electronic adj3 health adj3 information).tw.
17. (electronic adj3 medical adj3 information).tw.
18. (digital adj3 health adj3 information).tw.
19. (digital adj3 medical adj3 information).tw.
20. or/11-19
21. Quality control/
22. Quality Indicators, Health Care/
23. Trust/
24. (quality or evaluat$ or trust or credibility or accuracy or reliability or assessment).tw.
25. or/21-24
26. 10 and 20 and 25
27. limit 26 to yr="2002 - 2008"
28. limit 27 to (danish or english or norwegian or swedish)

APPENDIX 2: COCHRANE LIBRARY SEARCH STRATEGY FOR PREGNANCY REVIEWS

The following search strategy was performed in Cochrane Library, Issue 3 2008:

#1 MeSH descriptor Pregnancy explode all trees

#2 MeSH descriptor Postpartum period all trees

#3 (pregnan* or gestation* or labo?r? or birth? Or childbirth? Or parturition* or postpartum* or puerperium* or maternal or maternit*):ti,ab,kw

#4 #1 OR #2 OR #3

In addition all sections (n= 29) under **By Topic → Pregnancy and childbirth** were investigated to ensure retrieval of reviews not identified by the search strategy described above.

APPENDIX 3: CODING SCHEMES FOR THE COCHRANE REVIEWS

Cochrane review: Low glycemic diets for overweight and obesity

Resultatene i tabellen nedenfor er hentet fra en systematisk oversikt som ser på effekten av å gå på en lavglykemisk diett for å gå ned i vekt når du er overvektig[1]. I oversikten sammenlignet forfatterne lavglykemiske dietter med andre dietter. Deltakerne var menn og kvinner i alle aldre. Overvekt og fedme ble klassifisert etter validerte kriterier. Vurder graden av effekt for de ulike utfallsmålene i oversikten. Bruk følgende koder:

A. Effective – Lavglykemiske dietter virker for å gå ned i vekt

B. Probably effective - Lavglykemiske dietter virker sannsynligvis for å gå ned i vekt

C. Unknown effectiveness - Vi vet ikke/forskning finnes ikke om lavglykemiske dietter virker for å gå ned i vekt

D. Not effective - Lavglykemiske dietter virker ikke for å gå ned i vekt

E. Probably not effective - Lavglykemiske dietter virker sannsynligvis ikke for å gå ned i vekt

Utfallsmål	Resultater	Effekt-kode
Ulike vektmål (kg, KMI m.m.)	**Korttidseffekt (mindre enn seks måneders oppfølging):** Personer på en lavglykemisk diett gikk signifikant mer ned i vekt enn de på andre dietter. De som gikk på en lavglykemisk diett gikk i gjennomsnitt ned ett kilo mer, og reduserte i gjennomsnitt 1-2 KMI-enheter mer enn de på andre dietter. Resultater i tall for ulike vektmål: • Kroppsmasse (kg): 163 personer i 4 studier, gjennomsnittsforskjell -1.1 kg (-2.0 – -0.2), p=0.02. • Kroppsmasseindeks (KMI/BMI): 48 personer i 2 studier, gjennomsnittsforskjell -1.3 KMI-enheter (-2.0 – -0.5), p=0.009. • Total fettmasse (kg): 147 personer i 4 studier, gjennomsnittsforskjell -1.1 kg (-2.0 – -0.2), p=0.003.	
	Effekt mellom 6 og inntil 12 måneder: Ingen studier målte effekt over seks måneder.	
	Langtidseffekt (12 måneder eller mer): Ingen studier målte langtidseffekt.	
	Effekt på barn: Én studie med 16 barn viste en signifikant forskjell i vektreduksjon mellom dem som gikk på en lavglykemisk diett sammenlignet med en energiredusert diett. Resultater i tall for ulike vektmål: • Kroppsmasseindeks: gjennomsnittsforskjell -4.20 KMI-enheter (-7.4 – -1), ingen oppgitt p-verdi. • Total fettmasse: gjennomsnittsforskjell -1.8 kg (-3.8 – -0.2), ingen oppgitt p-verdi.	
Bivirkninger	Ingen av studiene definerte og målte bivirkninger.	

Thomas DE, Elliott EJ, Baur L. Low glycaemic index or low glycaemic load diets for overweight and obesity. Cochrane Database of Systematic Reviews 2007, Issue 3. URL: http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD005105/frame.html

Cochrane review: Calcium supplementation in pregnancy

Resultatene i tabellen nedenfor er hentet fra en systematisk oversikt som ser på effekten av å ta kalsiumtilskudd under graviditeten for å forebygge høyt blodtrykk, svangerskapsforgiftning m.m[1]. I studiene som er inkludert i oversikten fikk de gravide et tilskudd på 1 500 til 2 000 mg per dag. Vurder graden av effekt for de ulike utfallsmålene i oversikten, både for gravide generelt og for de ulike undergruppene av gravide (gravide med tilstrekkelig kalsiuminntak osv.). Bruk følgende koder:

A: Effective – Å ta kalsiumtabletter forebygger høyt blodtrykk osv.

B: Probably effective - Å ta kalsiumtabletter forebygger sannsynligvis høyt blodtrykk osv.

C: Unknown effectiveness - Vi vet ikke/forskning finnes ikke om kalsiumtabletter forebygger høyt blodtrykk osv

D: Not effective - Å ta kalsiumtabletter forebygger ikke høyt blodtrykk osv.

E: Probably not effective - Å ta kalsiumtabletter forebygger sannsynligvis ikke høyt blodtrykk osv.

Utfallsmål	Resultater	Effekt-kode
Høyt blodtrykk	**Alle gravide:** Det var signifikant færre kvinner i kalsiumgruppen som utviklet høyt blodtrykk under graviditeten sammenlignet med placebo: • 14 946 kvinner i 11 studier: RR 0.70 (0.57 – 0.86), p=0.0006	
	Gravide med et tilstrekkelig kalsiuminntak[2]: Det var signifikant færre kvinner i kalsiumgruppen som utviklet høyt blodtrykk under graviditeten sammenlignet med placebo: • 5 022 kvinner i 4 studier: RR 0.90 (0.81 – 0.99), p=0.04	
	Gravide med et lavt kalsiuminntak[3]: Det var signifikant færre kvinner i kalsiumgruppen som utviklet høyt blodtrykk under graviditeten sammenlignet med placebo: • 9 894 kvinner i 6 studier: RR 0.47 (0.29 – 0,76), p=0.002	
	Gravide med høy risiko[4]: Det var signifikant færre kvinner i kalsiumgruppen som utviklet høyt blodtrykk under graviditeten sammenlignet med placebo. • 327 kvinner i 4 studier: RR 0.47 (0.22 – 0.97), p=0.04	
Svangerskaps-forgiftning	**Alle gravide:** Det var signifikant færre kvinner i kalsiumgruppen som utviklet svangerskapsforgiftning sammenlignet med placebo: • 15 206 kvinner i 12 studier: RR 0.48 (0.33 – 0.69), p=0.0001	
	Gravide med et tilstrekkelig kalsiuminntak[2]: Det var ingen signifikant forskjell mellom kalsiumgruppen og de som fikk placebo • 5 022 kvinner i 4 studier: RR 0.62 (0.32 – 1.20), p=0.2	
	Gravide med et lavt kalsiuminntak[3]: Det var signifikant færre kvinner i kalsiumgruppen som utviklet svangerskapsforgiftning sammenlignet med placebo: • 10 154 kvinner i 7 studier: RR 0.36 (0.18 - 0.70), p=0.002	
	Gravide med høy risiko[4]: Det var signifikant færre kvinner i kalsiumgruppen som utviklet svangerskapsforgiftning sammenlignet med placebo: • 587 kvinner i 5 studier: RR 0.22 (0.12 - 0.42), p<0.00001	

Unngå for tidlig fødsel (før 37. svangerskapsuke)	**Alle gravide:** Det var ingen signifikant forskjell i antall for tidlige fødsler mellom kalsiumgruppen og placebo: • 14 751 kvinner i 10 studier: RR 0.81 (0.64 – 1.03), p=0.09	
	Gravide med et tilstrekkelig kalsiuminntak[2]: Det var ingen signifikant forskjell i antall for tidlige fødsler mellom kalsiumgruppen og placebo: • 5 033 kvinner i 4 studier: RR 0.59 (0.26 – 1.33), p=0.2	
	Gravide med et lavt kalsiuminntak[3]: Det var ingen signifikant forskjell i antall for tidlige fødsler mellom kalsiumgruppen og placebo: • 9 718 kvinner i 6 studier: RR 0.90 (0.80 – 1.02), p=0.1	
	Gravide med høy risiko[4]: Det var signifikant færre kvinner i kalsiumgruppen som fødte for tidlig sammenlignet med placebo: • 568 kvinner i 4 studier: RR 0.45 (0.24 – 0.83), p=0.01	
Innleggelse på nyfødtintensiv enhet	**Alle gravide:** Det var ingen signifikant forskjell i antall innleggelser på nyfødtintensiv blant de nyfødte hvor mor hadde fått kalsiumtilskudd sammenlignet med placebo: • 13 406 kvinner i 4 studier: RR 1.05 (0.94 – 1.18), p=0.4	
	Gravide med et tilstrekkelig kalsiuminntak[2]: Det var ingen signifikant forskjell i antall innleggelser på nyfødtintensiv blant de nyfødte hvor mor hadde fått kalsiumtilskudd sammenlignet med placebo: • 4 336 kvinner i 1 studie: RR 1.09 (0.95 – 1.26), p=0.2	
	Gravide med et lavt kalsiuminntak[3]: Det var ingen signifikant forskjell i antall innleggelser på nyfødtintensiv blant de nyfødte hvor mor hadde fått kalsiumtilskudd sammenlignet med placebo: • 9 070 kvinner i 3 studier: RR 0.98 (0.81 – 1.19), p=0.9	
	Gravide med høy risiko[4]: Det var ingen signifikant forskjell i antall innleggelser på nyfødtintensiv blant de nyfødte hvor mor hadde fått kalsiumtilskudd sammenlignet med placebo: • 63 kvinner i 1 studie: RR 0.29 (0.03 – 2.48), p=0.3	
Dødfødsel eller at barnet dør før utskriving fra fødeklinikken	**Alle gravide:** Det var ingen signifikant forskjell i antall dødfødsler eller dødsfall før utskriving mellom kalsiumgruppen og placebo: • 15 141 kvinner i 10 studier: RR 0.89 (0.73 – 1.09), p=0.2	
	Gravide med et tilstrekkelig kalsiuminntak[2]: Det var ingen signifikant forskjell i antall dødfødsler eller dødsfall før utskriving mellom kalsiumgruppen og placebo: • 5 033 kvinner i 4 studier: RR 1.12 (0.66 – 1.90), p=0.7	
	Gravide med et lavt kalsiuminntak[3]: Det var ingen signifikant forskjell i antall dødfødsler eller dødsfall før utskriving mellom kalsiumgruppen og placebo: • 10 108 kvinner i 6 studier: RR 0.86 (0.69 – 1.06), p=0.2	
	Gravide med høy risiko[4]: Det var ingen signifikant forskjell i antall dødfødsler eller dødsfall før utskriving mellom kalsiumgruppen og placebo: • 512 kvinner i 3 studier: RR 0.39 (0.02 – 9.20), p=0.6	
Hvis du vurderer alle disse utfallsmålene under ett, bør gravide ta kalsiumtabletter under svangerskapet?		

1) Hofmeyr GJ, Atallah AN, Duley L. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. Cochrane Database of Systematic Reviews 2006, Issue 3. URL: http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD001059/frame.html
2) Tilstrekklig kalsiuminntak ble definert som 900 mg kalsium eller mer per dag.
3) Lavt kalsiuminntak ble definert som mindre enn 900 mg kalsium per dag.
4) Dette ble definert av de forfatterne i de enkelte studiene og inkluderte tenåringssvangerskap, svangerskapsforgiftning i tidligere svangerskap, og kvinner med eksisterende høyt blodtrykk før graviditeten.

APPENDIX 4: CODING SCHEMES FOR WEB STATEMENTS

Lavglykemiske dietter for vektreduksjon

Bruk kodene nedenfor for å vurdere hva de enkelte nettsidene sier om effekten av lavglykemiske dietter for å gå ned i vekt.

A. **Effective**: Lavglykemiske dietter virker for å gå ned i vekt

B. **Probably effective**: Lavglykemiske dietter virker sannsynligvis for å gå ned i vekt

C. **Unknown effectiveness**: Det er usikkert om lavglykemiske dietter virker for å gå ned i vekt

D. **Not effective:** Lavglykemiske dietter virker ikke for å gå ned i vekt

E. **Probably not effective:** Lavglykemiske dietter virker sannsynligvis ikke for å gå ned i vekt

F. **Unclear:** Det er ikke mulig å avgjøre hva teksten/utsagnet sier om effekten av lavglykemiske dietter for å gå ned i vekt.

Kalsiumtilskudd under svangerskapet

Bruk kodene nedenfor for å vurdere hva de enkelte nettsidene sier om effekt av å ta kalsiumtilskudd (tabletter) under svangerskapet:

A. Effective: Å ta kalsiumtilskudd under svangerskapet har effekt (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)

B. Probably effective: Å ta kalsiumtilskudd under svangerskapet har sannsynligvis effekt (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)

C. Unknown effectiveness: Det er usikkert/vi vet ikke om å ta kalsiumtilskudd under svangerskapet har effekt (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)

D. Not effective: Å ta kalsiumtilskudd under svangerskapet har ikke effekt (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)

E. Probably not effective: Å ta kalsiumtilskudd under svangerskapet har sannsynligvis ikke effekt (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)

F. Unclear: Det er ikke mulig å avgjøre hva teksten/utsagnet sier om effekten av å ta kalsiumtilskudd

Vurder utsagnene ut fra hva de sier om de ulike undergruppene av gravide. La S1 stå blank dersom utsagnet ikke sier noe om gravide generelt. La feltet S2, S3 osv stå blankt dersom utsagnet ikke omtaler akkurat denne undergruppen gravide.

S1:
Vil gravide generelt ha effekt av å ta ekstra kalsiumtilskudd under svangerskapet (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)?

S2:
Vil gravide med tilstrekkelig[a] kalsiuminntak ha effekt av å ta ekstra kalsiumtilskudd under svangerskapet (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)?

S3:
Vil gravide med lavt kalsiuminntak[b] ha effekt av å ta ekstra kalsiumtilskudd under svangerskapet (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)?

S4:
Vil gravide med høy risiko[c] ha effekt av å ta ekstra kalsiumtilskudd under svangerskapet (bl.a. for å forebygge høyt blodtrykk, svangerskapsforgiftning, m.m.)?

a) Tilstrekkelig kalsiuminntak er definert som 900 mg eller mer per dag
b) Lavt kalsiuminntak er definert som under 900 mg per dag
c) Gravide med høy risiko: Kan være tenåringer, svangerskapsforgiftning i tidligere svangerskap, kvinner med eksisterende høyt blodtrykk etc.

APPENDIX 5: SEARCH STRATEGIES IN SEARCH ENGINES AND SMIL

The following free text searches were performed in Google (www.google.no), Kvasir (www.kvasir.no) Yahoo (www.yahoo.no), MSN (http://no.msn.com/) and SMIL (http://smil.uio.no/):

	Topic: Low glycemic diets for overweight	Topic: Calcium supplementation in pregnancy
1.	glykemisk indeks overvekt	1. gravid kalsium
2.	glykemisk indeks overvektig	2. graviditet kalsium
3.	glykemisk indeks fedme	3. svangerskap kalsium
4.	glykemisk indeks slanking	4. baby kalsium
5.	glykemisk indeks vektreduksjon	5. svangerskapsforgiftning kalsium
6.	glykemisk indeks vekt	6. gravid kalsiumtilskudd
7.	glykemisk index overvekt	7. graviditet kalsiumtilskudd
8.	glykemisk index overvektig	8. svangerskap kalsiumtilskudd
9.	glykemisk index fedme	9. baby kalsiumtilskudd
10.	glykemisk index slanking	10. svangerskapsforgiftning kalsiumtilskudd
11.	glykemisk index vektreduksjon	11. gravide kalsium
12.	glykemisk index vekt	12. gravide kalsiumtilskudd
13.	glykemisk belastning overvekt	13. svangerskap[a]
14.	glykemisk belastning overvektig	14. gravid[a]
15.	glykemisk belastning fedme	15. kalsium[a]
16.	glykemisk belastning slanking	16. baby[a]
17.	glykemisk belastning vektreduksjon	17. gravid kalk
18.	glykemisk belastning vekt	18. graviditet kalk
19.	lavglykemisk overvekt	19. svangerskap kalk
20.	lavglykemisk overvektig	20. baby kalk
21.	lavglykemisk fedme	21. svangerskapsforgiftning kalk
22.	lavglykemisk slanking	22. gravid kalktabletter
23.	lavglykemisk vektreduksjon	23. graviditet kalktabletter
24.	lavglykemisk vekt	24. svangerskap kalktabletter
25.	lavglykemiske overvekt	25. baby kalktabletter
26.	lavglykemiske overvektig	26. svangerskapsforgiftning kalktabletter
27.	lavglykemiske fedme	27. gravid kalktilskudd
28.	lavglykemiske slanking	28. graviditet kalktilskudd
29.	lavglykemiske vektreduksjon	29. svangerskap kalktilskudd
30.	lavglykemiske vekt	30. baby kalktilskudd
31.	glykemisk indeks diett	31. svangerskapsforgiftning kalktilskudd
32.	glykemisk indeks dietter	32. gravide kalk
33.	glykemisk index diett	33. gravide kalktabletter
34.	glykemisk index dietter	
35.	glykemisk belastning diett	
36.	glykemisk belastning dietter	
37.	lavglykemisk diett	
38.	lavglykemiske dietter	

a) Due to no hits on search 1-12 the terms "svangerskap"; "gravid"; "kalsium" and "baby" were searched separately in SMIL

In SMIL the following predefined topics (Emner) were investigated:

Topic: Low glycemic diets for overweight	Topic: Calcium supplementation in pregnancy
Ernæring: Kosthold (Norwegian sites = 47)	Graviditet og fødsel: generelt (Norwegian sites = 38)
Ernæring: Overvekt(Norwegian sites = 11)	Graviditet og fødsel: Svangerskapsforgiftning (Norwegian sites = 5)
Ernæring: Slanking (Norwegian sites = 6)	

APPENDIX 6: WEB OPERATOR CRITERIA

Criterion	Operational definition[a]	Page/ Site	Options
STRUCTURAL CRITERIA			
Disclosure of ownership	Indication of the entity that owns the information presented on the website (e.g., organization logo)	S	Yes No
Statement of purpose	The general purpose or aims behind the website or organization- may be found on Front Page, About Page, or Contact US.	S	Yes No
General disclaimers provided	Presence of a general disclaimer such as "Not a substitute for professional care", "For educational purposes only" or a link to a disclaimer.	S	Yes No
Feedback mechanism provided	Author, editor or webmaster or other official can be contacted. Presence of e-mail address, telephone, fax, or online form.	S	Yes No
CONTENT DEVELOPMENT CRITERIA			
Editorial review process	Presence of claim of use of an editorial review process (how the information was produced and its quality checked) or the listing of an editorial review committee or medical advisory board. This might be in an About Us, About this Site or Editorial Policy section.	S	Yes No
Literature search conducted[b]	Presence of a statement on what databases/sources were searched and the search strategies used. (NA if no content production method)	S	Yes No NA
Independent critical appraisal conducted[b]	Presence of a statement on whether two or more authors have appraised the literature using validated appraisal tools (checklists). (NA if no content production method)	S	Yes No NA

Notes: **NA =** Not available; **S** indicates evaluated at website level; **P** indicates evaluated at webpage level

a) As defined by Bernstam, Sagram, et al. 2005

b) This criterion is taken from the LIDA instrument.

APPENDIX 7: WEB CONTENT CRITERIA

Criterion	Operational definition[a]	Page/Site	Options
AUTHORSHIP			
Disclosure of authorship	Name of the person(s) or organization(s) present that is attributed as the creator or producer of the presented information.	P	Yes No
Authors' affiliation disclosed	Disclosure of author's affiliations or relationships with relevant entity (NA if no author).	S	Yes No NA
Authors' credentials disclosed	Disclosure of authority and qualification (M.D., Ph.D., N.D., etc.) of author. Disclosure does NOT include "Dr." or "Professor" (NA if no author).	S	Yes No NA
Physicians' credentials disclosed	Disclosure of credentials of physician (M.D. or including area of specialization). (NA if no author or author not physician, or not known if physician).	S	Yes No NA
CURRENCY			
Date of creation disclosed	Date disclosed when information was produced or reported. If just a date is listed (without "date created", it may be assumed as date created — such as with news organizations)	P	Yes No
Date of last update disclosed	Date disclosed of any revision or update	P	Yes No
ATTRIBUTION			
Sources clear	Claims are backed up with a source (e.g., reference, expert opinion, or bibliography)	P	Yes No
References provided	Presence of conventional references or citations relevant to information on the page; link to reference also acceptable.	P	Yes No
Type of evidence used (see explanations below)	If the "sources clear" or "references provided" is coded as "yes", then register the type of evidence used into the category types defined in the right hand column (NA if no source or reference is present).	P	• Guideline • Systematic review • Single study • Expert/textbook • Personal opinion • Other • NA

Notes: **NA =** Not available; **S** indicates evaluated at website level; **P** indicates evaluated at webpage level

a) As defined by Bernstam, Sagram, et al. 2005

Brief explanation of the evidence types

Guideline

Guidelines are "work consisting of a set of directions or principles to assist the health care practitioner with patient care decisions about appropriate diagnostic, therapeutic, or other clinical procedures for specific clinical circumstances"[1]

Systematic review

A systematic review is a review of the evidence on a clearly formulated question that uses systematic, explicit and reproducible methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review[2]

Single study

An original research study (primary research). Examples of single studies are randomised controlled trials, cross-sectional studies, and studies using qualitative methods such as in-depth interviews or focus groups

Experts or textbooks

Use "expert" if the information is authored by a person with some kind of specialisation within a health care discipline (i.e. clinical dietician), but no other sources/references are used. Textbooks are often written by experts; therefore these two types of evidence are classified as one.

Personal opinion

Use "personal opinion" if an author name is present, but no details on the author's credentials are provided and no other sources are referenced".

Other

If a source is provided that is different from the sources defined above, i.e. law articles and links to other websites, then classify the evidence type as "other".

Not available

If no author, source or reference is present the criterion will be coded as "not available"

1) National Library of Medicine (2009). MeSH: Medical Subject Headings. Thesaurus and controlled vocabulary used for indexing the MEDLINE database of medical literature provided by the National Library of Medicine (NLM).

2) Centre for Reviews and Dissemination (2001). Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York.

APPENDIX 8: WEBSITE DESIGN CRITERIA

Criterion	Operational definition[a]	Page/Site	Options
Disclosure of advertising	Clear distinction (visual, or by text) between advertising (paid piece of information or banner ad) and content. (Advertising does not include product listings or a buy now link). NA when no advertising.	P	Yes No NA
Internal search engine present	Presence or absence of a search engine (can be any type of search engine, including a product search engine)	S	Yes No

Notes: **NA =** Not available; **S** indicates evaluated at website level; **P** indicates evaluated at webpage level

a) As defined by Bernstam, Sagram, et al. 2005

APPENDIX 9: INCLUDED WEBSITES AND DISTRIBUTION OF WEBPAGES

Included websites	webpages – low GI diets	webpages – calcium
Apotek 1 - *www.apotek1.no*	-	2
Baby.no - *www.baby.no*	-	1
Babyverden - *www.babyverden.no*	-	1
BarnIMagen - *www.barnimagen.com*	-	1
DinKost - *www.dinkost.no*	9	2
Dr Fedon Lindbergs - *www.drlindbergs.no*	3	-
Helsedirektoratet - *www.helsedirektoratet.no*	-	1
Helsenett - *www.helsenett.no*	-	3
Helsenytt for alle - *www.helsenytt.no*	1	3
Iform.no - *www.iform.no*	2	-
iFORM - *www.iform.nu*	1	-
Landsforeningen for overvekt - *www.overvektige.no*	1	-
Landsforening for Prader-Willis syndrom - *www.praderwilli.no*	1	-
Lettereliv.no - *www.lettereliv.no*	1	-
Lommelegen - *www.lommelegen.no*	4	5
Mammanett - *http//:mammanett.no*	-	1
Opplysningskontoret for meieriprodukter *www.melk.no*	-	1
Nettdoktor - *www.nettdoktor.no*	-	2
Pasienthåndboka - *www.pasienthandboka.no*	1	1
Roche AS - *www.roche.no*	1	-
Slankenett - *www.slankenett.no*	2	-
Trim.no - *www.trim.no*	4	-
Vitus Apotek - *www.vitusapotek.no*	-	1
Total no. sites: 23	***Total no. webpages: 31***	***Total no. webpages: 25***

APPENDIX 10: REFERENCES TO CITED WEBPAGES

1) Lindberg, F., & Hansen-Møllerud, M. (n.d.). Barn i Balanse: om graviditet. Retrieved February 14, 2009, from http://www.barnimagen.com/barnimagen_site/innhold/tema/kosthold_og_ernaering/barn_i_balanse_om_graviditet

2) Pasienthåndboka. (2007, January 1). Hva er glykemisk indeks? Retrieved February 14, 2009, from http://www.pasienthandboka.no/?mode=document&documentid=15340

3) Landsforeningen for overvektige. (n.d.). Barn og overvekt. Retrieved February 14, 2009, from http://www.overvektige.no/

4) Vogt, K. (1997). Gravides og barns kosthold - hvordan forebygge sykdom? *Helsenytt for alle*. Retrieved December 10, 2008, from http://www.helsenytt.no/artikler/gravideskost.htm

5) Arsky, G. H. (2005, May 26). Inulin for vektkontroll. Retrieved February 14, 2009, from www.dinkost.no

6) Moland, M. (n.d.). Å spise for to... Retrieved February 14, 2009, from http://www.babyverden.no/templates/Article____4323.aspx

7) Fosse, E. (1999). Slik kan du sørge for at barna får et sterkt skjelett. *Helsenytt for alle*. Retrieved December 10, 2008, from http://www.helsenytt.no/artikler/sterkt_skjelett.htm

8) Sosial- og helsedirektoratet. (2004). Ernæring i svangerskapet. Retrieved February 14, 2009 from http://www.shdir.no/vp/multimedia/archive/00012/IS-2184_12613a.pdf

9) Eskelund, T. (n.d.). Hva med en kickstart? Retrieved February 14, 2009 from http://www.trim.no/pub/art.php?id=870

10) Andresen, Å. (2005, September 18). Sliten og gravid. Retrieved February 14, 2009, from http://www.lommelegen.no/php/art.php?id=333750

11) Spilde, I. (2008, July 11). Ny forskning taler for GI-dietten. Retrieved February 14, 2009, from http://www.iform.no/pub/art.php?id=720

12) Dinkost. (2000, March 01). Matens glykemiske indeks bestemmer kroppens insulinrespons. Retrieved February 14, 2009 from www.dinkost.no

13) Anker, C. (2004, November 3). Svangerskapsforgiftning: haster! Retrieved February 14, 2009, from www.helsenett.no

14) Synnevåg, M. (n.d.). Protein holder deg slank. Retrieved February 14, 2009, from http://www.trim.no/pub/art.php?id=846

15) Thommesen, M. (2000, January 1). Kalsium og vitamin D i svangerskapet. Retrieved February 14, 2009, from www.dinkost.no

16) Kyed, A. S. (2003, May 15). Sykehusets kostholdsråd. Retrieved February 14, 2009, from http://www.lommelegen.no/php/art.php?id=322391

17) Yndestad, A. (2000, August 20). Jeg er gravid - Hvor mye jern trenger jeg? Retrieved February 14, 2009, from http://www.lommelegen.no/php/art.php?id=325727

18) Helsenett. (2008, May 22). Kalsium. Retrieved February 14, 2009 from www.helsenett.no

19) Thommesen, M. (2005, September 28). Glem GI – tenk grovt! Retrieved Febrary 14, 2009, from www.dinkost.no

20) Lindberg, F. (2008). Spørsmål og svar om GI og GB. Retrieved February 14, 2009 from http://www.drlindbergs.no/no/Sporsmal_og_svar/GI_og_GB/

21) Andresen, Å. (2005, September 8). Vit A, D og kalsium hos gravide. Retrieved February 14, 2009 from http://www.lommelegen.no/php/art.php?id=333749

22) Lommelegen. (2009). Om oss. Retrieved January 10, 2009 from http://www.lommelegen.no/php/art.php?id=322507

23) Pasienthåndboka. (2007, May 23). Retningslinjer for arbeidet med Pasienthåndboka. Retrieved February 14, 2009 from www.pasienthandboka.no

24) Babyverden. (2007). Om Babyverden. Retrieved February 14, 2009 from http://www.babyverden.no/templates/Article____6780.aspx

25) Baby. (2005). Spesialister tilknyttet Baby. Retrieved February 14, 2009 from http://www.baby.no/om/spesialister/

26) Thommesen, M. (2000, January 1). En diett som passer for meg? Retrieved February 14, 2009, from www.dinkost.no