Comprehension of Online Cancer Information: A Propositional Assessment of Readability, Inferences, and Coherence

by

Rachelle Ta-Min

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Health Studies and Gerontology

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This study was designed to examine the comprehension processes used when reading cancer information found on the Internet and its relation to readability. The focus of the research was on the utilization of an alternative methodology, propositional analysis, to provide indices of textual difficulty that complement standard readability formulas. Kintsch's (1998) framework about discourse comprehension was used to distinguish between textbase and situational factors influencing comprehension.

This study analysed the verbal protocols of 16 community dwelling older adults. They each read a pair of either breast, prostate, or colorectal cancer web pages at low and high readability levels. Propositional density and coherence were measured for the web pages; recall, concepts, inferences, and coherence were measured for the protocols. Coherence was also captured using network representations. The readability formulas used were the Flesch-Kincaid (F-K), Flesch Reading Ease (FRE) and SMOG. Results showed that propositional density did not agree with readability scores, and that the readability formulas did not even produce consistent results among themselves. Results also showed that readability was not associated with web page coherence. Analysis of the protocols revealed marked individual differences for the kinds of information recalled, the types of inferences made, and the coherence of mental models. Variations in background and personal interest appeared to influence whether superordinate or subordinate propositions were recalled. Dependence on prompted versus non-prompted interview questions also varied by individual.

The findings of this study suggest that propositional analysis should be considered as a complementary methodology to readability formulas. Relying solely on these formulas as an indictor of comprehension may mislead online health providers that their information will be understood. The

findings also highlight that that individual create distinct and personalized mental models when presented with web pages that are influenced by text and situation based factors.

Acknowledgements

My foremost gratitude goes to my family, especially my parents, who have always provided me with the greatest worth – opportunity. I would like to say thanks to my supervisor, Dr. Jose F. Arocha for his expertise, guidance, and support regarding this project. My sincere appreciation also goes to Dr. Laurie Hoffman-Goetz and Dr. Paul Thagard for their feedback and critical points of view, and to Daniela Friedman for her pivotal contribution.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: Introduction

1.1 The World Wide Web

The use of the Internet continues to grow in Canada, where an estimated 7.9 million (64%) of the 12.3 million Canadian households with access had at least one member who used the Internet regularly in 2003 ("Statistics Canada", 2004). This steady growth of the Internet has been through the World Wide Web, which supports multimedia through a graphical interface that allows sophisticated text formatting, graphics, and embedded hypertext links to other locations on the Web and contains vast amounts of consumer health and medical information. Much of this information is posted by government agencies, medical foundations, universities, medical schools, individual physicians, health insurance companies, health care providers, individuals' personal Web pages, special interest support groups, and many health and medical-related companies such as pharmaceutical companies and medical supply firms. According to the 2003 Household Internet Survey, searching for medical or health information ranks fourth amongst the most popular uses for accessing the Internet at approximately 65%, which has steadily increased since 1999.

1.1.1 A Channel for Health Information

One strength of the World Wide Web as a communication medium is that health messages can be delivered to receivers through multiple communication channels such as text, graphics, photos, animations, audio, and video. Bernhardt and Cameron (2003) noted that non-text channels may be more accessible and understandable by people with low literacy than text-based messages alone. Another strength is that the information can be tailored to each receiver's characteristics, needs, and preferences. This would involve surveying individuals on predetermined psychosocial and communication constructs and then use computer programs to generate and deliver these tailored messages. In this vein, the Internet can be viewed as a facilitator of message delivery, featuring

flexibility and multi-channel capabilities. Tailoring has been used in the past to effectively communicate important public health issues such as physical activity (Bull, Kreuter, & Shariff, 1999), smoking cessation (Strecher et al., 1994), and cancer screening (Skinner, Strecher, & Hospers, 1994).

Jimison and Sher (2000) outlined some of the basic potential advantages of a computer-based approach to patient education over traditional approaches:

- Consistent content and delivery
- Potentially more easily available than a health educator
- Potentially cost-effective (compared to routine and consistent education by staff)
- Enables privacy of communication
- Promotes active learning

The computer-based approach, however, also brings new challenges for patient-provider relationships. One current challenge is health literacy in patient education. Patient comprehension is a prerequisite to adherence to health care treatment goals, and the conventional techniques for patient education often fall short in this regard (French & Larrabee, 1999). Currently, the most commonly used format for patient education is the brochure or pamphlet, making heavy use of text at a readability level too high for the general public (Glazer, Kirk, & Bosler, 1996; Payne, Large, Jarrett, & Turner, 2000; H. Smith, Gooding, Brown, & Frew, 1998). Most studies measuring the understanding of patient education materials have found that approximately 50% of the patients are unable to understand the written materials (Davis, Crouch, Wills, & Miller, 1990; Doak, Doak, Friedell, & Meade, 1998; Le Bas, 1989).

A variation of the readability problem has to do with the patient's ability to understand medical language. A highly educated patient may have very little medical background and be unfamiliar with much medical terminology. Conversely, a hospital employee with less formal education and lower reading ability could obtain a fairly high level of understanding of medical

terminology. Even conducting searches on the Internet requires a certain degree of computer literacy. In short, online health information providers need to recognize factors that can potentially limit users' comprehension.

1.1.2 Breast, Prostate, and Colorectal Cancer

According to the National Cancer Institute of Canada, in 2003, men and women's lifetime probability of developing any cancer was approximately 41% and 38% respectively. Breast cancer is the most common cancer among Canadian women. In 2005, an estimated 21,600 women were diagnosed with breast cancer and 5,300 died of it (Canadian Cancer Society, 2005). For men, prostate cancer is the number one cancer threat. The Prostate Cancer Research Foundation of Canada (2006) estimates 20,500 men will be afflicted with prostate cancer this year alone.

Colorectal cancer is the second and third most prevalent cancer for men and women respectively. In contrast to the national screening programs for breast and prostate cancer, routine colorectal cancer screening is not yet a standard practice in Canada. As such, health consumers have had to become increasingly vigilant regarding risk factors and symptoms of colorectal cancer. Given the current projection that Internet is to grow as a leading source of health information, there is a legitimate concern that any cancer information found online must be comprehensible to consumers.

1.2 Health Literacy

Literacy is sometimes measured in terms of comprehension skills, vocabulary, and the ability to communicate effectively across a wide variety of contexts. The term "health literacy" was first used in health education about thirty years ago (Ratzan, 2001). The U.S Department of Health and Human Services (2000) defined it as "the capacity to obtain, interpret and understand basic health information and services and the competence to use such information and services to enhance health." In a medical situation, the health literacy of patients directly influences their access to crucial

information about their right and their health care, whether it involves following instructions for care, taking medicine, comprehending disease-related information, or learning about disease prevention and health promotion (Rudd, Moeykens, & Colton, 1999).

The problem with educating the patient through written materials is compounded when the target population includes a high percentage of patients with less than adequate literacy skills. The importance of matching patient education materials with patient reading comprehension levels has been documented in diverse areas such as smoking cessation (Meade & Byrd, 1989), cancer education (Griffiths & Leek, 1995; Meade, Diekman, & Thornhill, 1992), and diabetes education (McNeal, Salisbury, Baumgardner, & Wheeler, 1984).

Glazer et al. (1996) examined the reading comprehension level of patient education pamphlets concerning prevention, detection and treatment of breast cancer for low literacy women. Though they did not provide a definition for "low literacy", they referred to an earlier study (Glazer-Waldman, Hall, & Weiner, 1985) that found that only 40% of patients seeking care at a local hospital could read at a sixth grade level. They found that the 19 pamphlets they analysed had an average readability level at grade nine, making much of the available literature inappropriate for the target population. Most of the studies examining the match between reading level of health materials and that of those expected to read them document a clear difference. Many studies in the literature focus on the disparity between the reading abilities of cancer patients and the reading level of the educational materials written for them (Cooley et al., 1995; Doak et al., 1998; Michielutte, Bahnson, Dignan, & Schroeder, 1992). Cooley and colleagues (1995) found that the reading levels of 27 percent of cancer outpatients in one study were well below that of any of the thirty cancer pamphlets analysed using the Flesch Reading Ease Formula as the index of comprehensibility.

CHAPTER 2: Literature Review

Health education can take many forms, with the Internet being among the newer channels of delivery. One fundamental principle for effective health education is that the receiver must be able to understand the arguments being made in order to make decisions, perform suitable health behaviours, or simply to build a body of knowledge. In terms of text comprehension, the representation of text is formed by continually connecting incoming text information with prior knowledge (Kools, Ruiter, van de Wiel, & Kok, 2004). Ideally, the reader forms a coherent mental representation of the text's context as reading proceeds. At times, however, inference processes are needed when the overlap between concepts during reading is insufficient and text coherence is low.

2.1 Readability and Comprehension

Reading comprehension is the process of acquiring information. The characteristics of the meaning of a text as well as the processes involved in deriving this meaning are critical in determining readability. What are the current practices for determining readability? Traditional methods have included using previous formulas to validate a new one and having subjects rate the readability of texts.

According to Kintsch and Vipond (1977), interest in readability peaked in the 1950s but fell out of favour in the 1970s. The recent resurgence of interest in readability is in part due to the development of computerized reading management software (School Renaissance Institute, 2000). Readability is of considerable practical significance to educators and publishers of educational materials. For those who depend upon communication through the printed word, a pertinent question is whether the material will be read and, if read, understood by those for which it was intended. In the domain of the Internet, health information providers want to know in advance whether a particular text is likely to be read with understanding, or read at all.

To meet this need, readability formulas were invented. More than 40 readability formulas have been developed over the years (Klare, 1974-5). The formulas differ from one another in the number, weights, and types of variables used. The most common methods include the Fry, Flesch, Fog, and SMOG indices, which are generally based on sentence length, word difficulty, and word length (>3 syllables). Each formula provides an estimate of reading age, which approximates the educational level required to read and understand written materials (Bauman, 1997). An early formula, the McCall-Crabbs Standard Test Lessons in Reading (reviewed in Mitchell, 1985), looked at the number of questions regarding prose passages that could be correctly answered by children of different ages. There was, however, an absence of a theoretical framework to relate the nature of the questions from either the structure or content of the text. Similarly, in the absence of a theory of text structure and text processing, researchers selected a number of fairly obvious surface features of the text (e.g., average word length). They then used linear regression to determine to what extent these variables predicted question-answering scores.

Currently almost every formula has a word variable and sentence variable. Klare (1974-5) has noted that word length has a consistently higher predictive value than sentence length for assessing difficulty. It is, however, important to keep in mind that formulas only have predictive validity. Kintsch and Vipond (1977) have noted that if long words and long sentences were the singular sources of comprehension difficulty, then shortening them would remove the difficulty, which is obviously false. This is because comprehension is comprised of many factors beyond the length of words and sentences, such text coherence and prior knowledge.

2.2 Limitations of Readability Formulas

Many readability formulas have been developed as a result of research into factors within writing that correlated highly with style difficulty. Two of the most commonly used measures are the Flesch Reading Ease formula and the Flesch-Kincaid Grade Level because they are contained in

Microsoft Word and other widely used software. These conventional accounts of readability have certain shortcomings, in part because they are concerned with word and sentence properties at a superficial level. At best, they are correlated with whatever makes a sentence easy or hard, but they are not the sole determinants. The problem is that the predictors commonly used do not directly reflect either the content or the organization of a text, both of which are important elements of comprehension. For example, it is said that an abstract, complex discussion is accompanied by many conjunctions (Kintsch & Vipond, 1977). Clearly, the conjunctions are not the root of the difficulty, but are a "surface level" symptom of it.

Readability formulas also cannot capture the cohesion or coherence of a text. Research has clearly shown that readers have less difficulty reading cohesive texts (Britton & Gülgöz, 1991; McNamara, Kintsch, Songer, & Kintsch, 1996; McNamara & Kintsch, 1996). Traditional readability measures can run orthogonal to cohesion measures. Average sentence length and average number of syllables per word alone cannot sufficiently predict coherence and therefore understanding of a text. In Table 1, versions of a low vs. high cohesion sentence are compared in terms of two common measures of readability. The output of the Flesch Reading Ease formula is a number from 0 to 100, with a higher score indicating easier reading; the Flesch-Kincaid Grade Level formula converts the Reading Ease Score to a grade-school level. A popular grading formula for cancer material is the SMOG, which computes reading level as a function of the number of polysyllabic words (words with three or more syllables), using 10 sentences from the beginning, middle, and end of the text source. Here, more polysyllabic words correspond to high reading levels (Romano, 1979).

Table 1. Relationship between Readability and Cohesion

Cohesion Level	Sample Text	Flesch Reading Ease	Flesch-Kincaid Grade
Low	The woman was experiencing chest pains. She went to the emergency room.	59.7	6.4
High	The woman was experiencing chest pains so she went to the emergency room.	69.9	6.7

The high cohesion version of the sentence requires a higher grade level to be understood, despite having a higher reading ease score. Although a text should generally contain 200 words before the Flesch Reading Ease and Flesch-Kincaid Grade Level scores can successfully be applied, this example serves to demonstrate that more than surface features need to be considered in measuring the coherence and comprehensibility of a text.

Kintsch and van Dijk (1978) developed a theory of reading comprehension with the aim of supplementing current predictor variables. This is based upon the model for the representation of meaning in memory developed by Kintsch in 1974. They noted that factors such as the number of questions answered correctly, the ease in which a text can by typed or translated, and various eye movement statistics may be useful as practical indicators of comprehension, but that they only reflect comprehension indirectly. They also pointed out that texts equated for readability scores may still vary considerably in behavioural measures of reading ease.

2.3 Theory for the Representation of Meaning in Knowledge and Memory

A number of theories about text comprehension exist in which different parts of the reading process are described: recognizing letters and words, syntactic parsing of sentences, understanding the meaning of words and sentences, and incorporating the meaning of the text to other present

knowledge about the same topic. One of the most influential theories is Kintsch and van Dijk's theory for the representation of meaning in knowledge and memory (Kintsch & van Dijk, 1978). This theory describes the complete reading process, from recognizing words until constructing a representation of the meaning of the text. The emphasis of the theory is on understanding the meaning of a text. In 1988, Kintsch extended the theory with the construction-integration model (Kintsch, 1988), which he followed up by an updated version in 1998 (Kintsch, 1998).

2.1.1.1 Main Assumptions

Text comprehension is the active process of constructing mental representations of the text information. The process of constructing a situation model is called the "comprehension process". Kintsch and van Dijk (1978) assume that readers of a text build three different mental representations of the text: a verbatim representation of the text, a semantic representation that describes the meaning of the text, and a situational representation of the situation to which the text refers. The propositional representation consists initially of a list of propositions that are derived from the text. After having read a complete sentence, this list of propositions is transformed into a network of propositions. If the text is coherent, all nodes of the network are connected to each other.

Most thinking entails relating concepts to one another in a meaningful fashion. One way of doing this is by forming propositions. Propositions are the smaller units of meaning that can stand alone as an assertion and be judged true or false. The concept *dog* is neither true nor false; nor is *mammal*. The proposition dogs are mammals, however, relate the two concepts in a meaningful way and can be judged for its accuracy. Mental images, mental models, concepts, and propositions are thus the building blocks of thought. Furthermore, when two different sentences say the same thing, they express the same proposition. The notion of a proposition is useful for stating the equivalence of sentences in different languages as well: "Ontario is a province in Canada" expresses the same proposition in French ("l'Ontario est une province en Canada") as it does in English.

The purpose of propositional analysis is to analyse the meaning of text by abstracting from its surface feature (words and terms used). Along with network representations, it is used to analyse both the text read and the mental model (i.e., verbal protocol) constructed from reading the text. Mental models are the conceptions of a text that develop in the mind of the reader. This allows the mapping of what the text says onto the reader's mental models of the text meaning, which grants a very detailed examination of the text coherence and the reader's comprehension. In this way, cognitive analyses are complementary to surface measures of readability such as Cloze tests.

Propositional analysis is a form of representation of a semantic network in memory (Patel & Arocha, 1995). Propositional measures have been used in the past to investigate how complex texts and explanations are in terms of their underlying meaning in a variety of contexts from doctors' memory for clinical cases (Patel, Groen, & Frederiksen, 1986) to political discourses (Ghiglione, Landré, & Bromberg, 1990). Propositional analysis has also been applied to stories as well as to texts described as technical prose, and with regard to recall protocols, has been applied to (1) gist recall protocols from prose memory experiments in which reproductive recall only is scored, and (2) responses produced in a "main idea" task, where after reading a passage the participant must produce a one-sentence statement of the "main idea" (Bovair & Kieras, 1985). For recall protocols, the material to be propositionalized is normally a passage and the propositions form a scoring key.

Format of a Propositional Representation. Propositions are composed of concepts; each proposition must include first a predicate or relational concept, and one or more arguments. The latter may be concepts or other embedded propositions. Predicates can be in the form of verbs, adjectives, adverbs, and sentence connectives such as "and", "but", "because", "although", "yet", "then", "next", and so on (Kintsch & van Dijk, 1978).

2.1.2 Propositional Density

Propositional analysis provides a measure of propositional density, or how many propositions are in the abstract textbase underlying a text. P-Density is different from word length, and seems to have a direct effect on difficulty. The more propositions there are in the textbase, the harder the text is in the sense that more propositions take longer to read. Kintsch established this point early in his research, when he compared reading speed on two passages of equal length with different numbers of propositions (Kintsch & Keenan, 1973). The passage with more propositions took longer to read, though it should be noted that all propositions were not equally difficult to remember: superordinate propositions were recalled better than propositions which were structurally subordinate. Propositional density also provides an index of the quantity of information contained in expository text independent of its word length. It is also known to increase processing difficulty and to reduce memory performance in terms of proportional recall (Kintsch & Keenan, 1973; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975). Stine and Wingfield (1990) have studied the extent to which working memory deficits contribute to age differences in discourse memory. Overall, older adults demonstrated lower levels of recall than younger adults for the sentence span task. Additionally, age differences in recall were slightly exaggerated by increased propositional density, but not by increased passage length. Stine and Wingfield concluded that age effects could not be completely explained in terms of age differences in working memory span because working memory capacity provided greater explanatory power when texts were simpler.

2.1.3 Inferences and Prior Knowledge

Language users are able to provide, during comprehension, the unstated relationships between propositions on the basis of their general or contextual knowledge of the facts. A reader's knowledge determines to a large extent the meaning that he or she derives from the text. If the

knowledge base is lacking, the reader will not be able to derive the same meaning that a person with adequate knowledge, reading the same text, would obtain (Kintsch & van Dijk, 1978).

Barclay, Bransford and Franks (1974) discovered that initial understanding of a text depends on applying relevant prior knowledge that is not in the text (for a detailed discussion, see Bransford & Johnson, 1972). Constructive processes are a crucial part of text comprehension, in particular to the making of many inferences in the course of comprehension. Texts rarely explicitly spell out everything needed for successful comprehension, so prior knowledge helps the reader fill in contextual gaps within the text and to develop a better global understanding or situation model of the text (McNamara & O'Reilly, 2002).

Kintsch's theory of text comprehension makes the distinction between two levels of representation, the textbase and the situation model. The textbase is defined in terms of propositions and relations among propositions (van Dijk & Kintsch, 1983). It has a local structure (the microstructure) as well as a global structure (the macrostructure). The process of transformation from words to meaning units involves a certain amount of inferential activity. The reader must add nodes and establish links between nodes from his or her own knowledge and experience to make the structure coherent. For example, the referents of pronouns must be identified, synonymous terms must be matched, and gaps in the coherence of the network must be identified by bridging inferences (Kintsch, 1988; van den Broek, Virtue, Graddy, Tzeng, & Sung, 2002). If successful, the result is a locally and globally well-structured memory representation of the text.

A reader is not, however, ensured a deeper understanding of the text by knowing it at the level of the textbase. Generally, the reader must establish links from his or her own knowledge and experience to make the structure coherent, and to complete and interpret it (Kintsch, 1998). This inferencing results in Kintsch's second level of representation, the situation model. This model incorporates knowledge about the language, about the world, and about the specific communication

situation. Such sources of information may all be needed to complement the textual information and to transform what by itself is only an isolated memory structure into something that relates to and is integrated with the reader's personal store of knowledge and experience. For example, the local coherence of the text "The woman was experiencing chest pains" followed by "They raced to the hospital, sirens blaring" is underwritten, to most readers, that the woman was being taken to the hospital in an ambulance. The combination of explicit text information with relevant background knowledge produces a situational representation that is the hallmark of constructivist processing.

Prior knowledge and text structure have been found to influence readers' comprehension, browsing, and perceived control in a hypertext environment (Calisir & Gurel, 2003; Potelle & Rouet, 2003). Potelle and Rouet (2003) noted that without adequate prior knowledge, readers are limited in the constructive processes. They conducted a study looking at the effects of content representation and reader' prior knowledge on the comprehension of hypertext. Hypertext is generally defined as any text that contains links to other documents. More specifically, Potelle and Rouet used Kintsch's theory of text comprehension as a framework to interpret the effect of various design options on students' comprehension of instructional hypertexts. They found that the hierarchical map improved comprehension for the low knowledge participants at the global, but not the local level. There was no effect of content representation on the comprehension of high prior knowledge students. This suggests that there may be an optimal content representation for each category of readers. Content representations displaying only basic relationships may be beneficial to low knowledge readers while more sophisticated representations may be appropriate to high knowledge readers.

2.1.4 Coherence

Comprehension research has shown that coherence makes a text easier to understand by constraining the amount and types of inferences that are made during interpretation. Studies of text comprehension (Britton & Gülgöz, 1991; Kintsch & Vipond, 1979) have demonstrated that the

quality of understanding increases as the coherence of the text increases. This is because different types of propositions bridge various parts of the text, making more salient the inferences that need to be made to capture the intended meaning.

The process of incorporating propositions into the textbase is a process of maintaining coherence. Propositions that have overlapping arguments, and thus are semantically related, create coherence. However, if the current proposition being processed does not share arguments with propositions in short-term memory, then a bridging inference must be made by the reader in order to maintain coherence. To construct a coherent representation, the reader must interpret each element of the text and identify meaningful connections to other elements in the text and in semantic knowledge. The resulting representation consists of nodes, which capture the elements in or related to the text, and connections, which capture the semantic relations between text elements (van den Broek et al., 2002). Together, these nodes and connections form a network. The more interconnected the representation, the more coherent it is. If readers do not have the necessary background knowledge to fill in the gaps left by the author of the text, they will be unable to form a coherent representation (Kintsch, 1988). Propositional analysis allows the identification of propositions that express single ideas and those that express ideas which in turn reference other concepts. Coherence is thus defined for our purposes as a measure of the degree of connectedness among propositions explicitly provided in the text.

2.1.5 Individual Characteristics

The comprehension process cannot be discussed without considering the goals and control processes that govern it. The reader's purpose and schema, as well as the reader's knowledge base in acquiring new information are crucial components of the entire process. A schema is an integrated pattern of knowledge stored in memory that organizes information and guides the acquisition of new information. It can be manipulated easily enough by instructions, titles, or implicit task demands.

There is a large body of research that shows that consideration of purpose can override text structure effects. Kozminsky (1977) showed how the pattern of recall changes predictably as a function of which of two possible titles were assigned to it. This organization determines the pattern of rehearsal for the various propositions and hence the pattern of recall.

A reader's goals also influence the comprehension process. For instance, suppose a person reads a text not with the general goal of comprehending all of it as well as possible, but with some specific purpose in mind. This would suggest that a related special-purpose textbase is constructed in which propositions related to the reader's goals are emphasized. Since only special-purpose propositions would be focused upon, this would result in an incoherent textbase. The propositions that would remain unrelated to the main network, however, would be irrelevant to the reader's goals, so it would not matter if they were lost.

2.1.6 Text Characteristics

Texts may be comparable in word length, but they may differ in the number of propositions expressed by these words. Kintsch and Keenan (1973) systematically varied the number of propositions in a textbase while keeping constant the number of words in the text. They observed that the longer the texts were, the more time was required for the processing of each additional proposition.

Another reason two texts that have comparable readability scores can have differences in the ease in which they are comprehended is based on the number of different arguments that are used in a textbase. Kintsch, Kozminsky, Streby, McKoon, and Keenan (1975) constructed texts that were equal in the number of words as well as the number of propositions. The texts differed, however, in that in some paragraphs the same concepts were used over and over again as propositional arguments, whereas in others, new concepts were continuously introduced. Intuitively, one might think that the paragraphs with repeated concepts would be more difficult to comprehend because of the necessarily

complex embedding at both the propositional and syntactic levels. However this is not the case, as Kintsch et al. found that word concept recall increases as a function of the number of repetitions of that concept in the textbase.

2.1.7 Summary and Implications

While readability formulas are useful for measuring surface features of text difficulty, it is clear that a more in-depth approach is needed to ascertain users' comprehension of cancer information on the Internet. The advantages of providing health information online include the ability to: update as new information becomes available, reach a wide audience, be cost-effective, target special sub-groups of the population, and so on. These advantages can, however, be diminished, if people experience difficulty comprehending the information provided. It is therefore necessary to understand the relationship between the readability and comprehensibility of breast, colorectal, and prostate cancer information that are available on the Internet.

CHAPTER 3: Rationale

There is general consensus that the readability levels of health information on the Internet are too high. It is not enough, however, to improve the comprehension of users by simply lowering the readability levels of the information (Klare, 1976), but it is a necessary condition. This is because comprehensibility of information is a function of a number of factors other than text readability, in particular the quantity of information contained in the text (i.e., its propositional density) and its coherence, as well as the reader's familiarity with the information.

Although cognitive semantic analyses and the case-based paradigm have been used to investigate people's understanding of health and disease (Arocha & Patel, 1995; Groen & Patel, 1988; Patel, Kaufman, & Arocha, 2000), they have never been applied to the evaluation of Web-based cancer information. The goal of the case-based paradigm is to describe a complex phenomenon in detail in order to have an understanding of the phenomenon in all its complexity (Arocha, 2005). In contrast to group-based research, which requires large numbers of participants to make use of statistical techniques to generalize findings to the population, this paradigm investigates individuals rather than groups and tries to capture details that are particular to each unique individual (Arocha, 2005; Runkel, 1990).

Quantitative experimental research bases its scientific rigour largely on statistical tests such as the correlation. It is known, however, that predictions about individuals from statistical data about the population are generally unreliable. Kennaway (1998) mathematically demonstrated that unless the correlation is at least 0.99, it is not possible to reliably estimate even the sign of the variable relative to its mean. The argument is that correlations are useless for making reliable predictions in individual cases. This example is presented here not to suggest that statistical assessment methods should cease to be used, but serves to simply demonstrate what they can and cannot capture. The

complexity of human behaviour is such that it should encourage methodological pluralism and that the methods used should suit their purpose. That is, the investigation of individual behaviour versus between-group comparisons should employ different methods because they each explicitly seek to describe different phenomena.

Protocols are one mean of gaining information about the course of an individual's cognitive processes by verbal methods. Propositional analysis seeks to understand in detail the mechanisms and internal structure of comprehension though the examination of semantic relations. This type of analysis does not require a large sample size because it focuses on understanding individual cases instead of on group differences that can be generalized to the population. Case studies satisfy the three tenets of the qualitative method: describing, understanding, and explaining. It can be thought of a specialized type of qualitative research that lends itself to generating, rather than testing, questions and hypotheses. Farand, Lafrance, and Arocha (1988) have noted that the internal validity of case studies is related to the quality and the level of detail of the underlying theoretical framework, as well as to the correspondence between the type of analyses and the framework. Kintsch's theory of comprehension and its method of propositional analysis describe highly specific processes and thus should be viewed as a suitable theoretical framework.

This study will contribute to the growing body of knowledge about comprehension and readability. The general issues that will be addressed are the following:

- Is propositional density similar to the results of standard readability formulas?
- What difficulties does the lay person encounter when trying to understand breast, colorectal, and prostate cancer information: general conceptual problems (i.e.., are certain concepts, e.g., screening procedures, risk factors, medical terminology, and so forth, omitted from the protocols?), local and/or global text coherence problems, inference problems?
- What is the effect of gender for different types of cancer information?

Although readability formulas are correlated with text difficulty, their results could be more informative if they are accompanied by results from a semantically-based tool. The formulas rely on calculable factors such as the average word length and the average number of syllables per word as determinants of reading difficulty. The first objective of this study will be to test the Flesch Reading Ease (FRE), Flesch-Kincaid (F-K), and SMOG formulas to see if they correspond with (a) the theory-based readability measure of propositional density; and (b) a comprehension measure based on the semantic methods of propositional analysis and semantic network representations. By doing this, one can assess the extent to which the FRE, F-K, and SMOG readability formulas really measure difficulty.

Question 1A: Will the FRE and F-K scores be associated with propositional density? This question asks whether a text with a high FRE score indicating easier reading will correspond with a low propositional density score, or whether a text with a low FRE score indicating difficult reading will correspond with a high propositional density score. Similarly, the question asks whether the related F-K formula which converts the FRE to a U. S. grade-school level will indicate an increased propositional density score as grade level increases, or if propositional density will decrease as the grade level decreases; this question also applies to the SMOG.

This question is based on the literature about the limitations of readability formulas. Readability formulas may underestimate the level of health oriented information because even short words (one or two syllables) that are technical may be unfamiliar to the lay reader (e.g., polyps). Conversely, the use of health terminology in web pages may artificially raise the readability score due to the use of many polysyllabic terms (e.g., adenomatous polyposis) and test and procedure names such as sigmoidoscopy, barium enema, and fecal occult blood test.

Contrarily, the literature about propositional density is straightforward: The more propositions (i.e., idea units) there are per sentence, the harder it is to comprehend. Studies have

found that as propositional density increases, so does reading time (Kintsch & Keenan, 1973), age differences for recall (Stine & Wingfield, 1990), and demands on processing efficiency (Kemper & Sumner, 2001). Looking at comprehension from the level of the proposition serves as a complementary measure to readability formulas, as propositional density remains comparatively independent from surface factors. By virtue of the difference in properties of the FRE, F-K, SMOG, and propositional density, it is reasonable to question whether there will be an association between the measures.

Question 1B: Will levels of local and global coherence, as discerned by the deconstruction of the text to its propositional textbase and then reconstructed into semantic network representations, have an association with FRE and F-K scores? This question asks whether a high FRE score indicating easier reading and its translated F-K score of a low grade-level will be associated with high levels of coherence. Similarly, the question asks whether a low FRE score indicating more difficult reading and its translated F-K score of high grade-level will be associated with low levels of coherence; this question also applies to the SMOG scores.

This question is based on research about the properties of readability formulas. One can lower grade level estimates by reducing word and sentence lengths, however this results in short and choppy sentences with minimal cohesion. Paradoxically, texts with shorter sentences run the risk of being more difficult to comprehend, particularly for readers with low reading proficiency. This is because there are fewer cues for cohesion that specify how the sentences should be conceptually related.

The semantic method of propositional analysis and network representations reduces the concept of coherence to argument repetition between propositions and the referential ties between nodes of the network. It is a simplistic yet objective and identifiable method of indexing the existence of a relationship. For this reason, it is expected that this method will detect inconsistencies in the relationship between coherence and the FRE, F-K, and SMOG.

The second objective of this project is to examine the extent to which gender influences comprehension of cancer information by comparing readers' understanding of a gender neutral (i.e., colorectal cancer) versus a gender specific cancer (i.e., males and females will read about prostate and breast cancer respectively).

Question 2: Will participants produce more coherent and detailed protocols, and make a greater use of inferences for gender specific rather than for gender neutral cancer information?

This question is based on the literature about prior knowledge and personal relevance and how these factors relate to comprehension. One of the most consistent findings of research in the late 1970s and early 1980s is that there is a strong relationship between prior knowledge and reading comprehension (Fielding & Pearson, 1994). In terms of Kintsch's theory of comprehension, information obtained purely from the textbase is directly cued by the text without the addition of anything else. It is not until the reader establishes nodes from his or her own knowledge is the information synthesized, made coherent, and more likely to be transferred to long-term memory (McNamara & Kintsch, 1996). Furthermore, the value of personal relevance is well-established as a memory enhancer (Rogers, Kuiper, & Kirker, 1999). When incoming information is encoded as self-relevant, an individual can elaborate further, making more rich connections to their associative networks of personal experiences. It is therefore reasonable to question whether the coherence for the semantic networks of gender specific cancer will be greater than that for gender neutral cancer.

CHAPTER 4: Method

This research was part of a larger research program devoted to the study of comprehension of Web-based cancer information (Friedman, Hoffman-Goetz, & Arocha, 2004). The selection of the web pages, the recruitment of participants, as well as the collection of the data, were conducted as part of this research program, by another researcher.

4.1 Data Source

The following sections describe the selection of web pages, participants, and procedure of the experimental session that constituted the primary data collection of the data source (Friedman, Hoffman-Goetz, & Arocha, in press).

4.1.1 Selection of Web Pages

Briefly, the selection process of the websites was the following:

- Keywords, such as "breast cancer", "prostate cancer" and "colorectal cancer" were entered into each of the ten most popular Internet search engines based on Nielsen statistics on audience reach
- Search engines were ranked based on a scoring system so that overall scores for the top 100 websites for each cancer type reflected an averaged ranking across all search engines
- Table 2. Inclusion and Exclusion Criteria for the Selection of Web Pages

	Criteria
Inclusion:	1. Pages were located within the given website 2. Pages were accessible within five clicks on the website 3. Websites had information on or links to breast, prostate, and/or colorectal cancers 4. Pages had > 10 sentences in the article

Exclusion:	1.	Web page hits that were part of a website with the same URL as a previous hit on the same search engine
	2.	Dead URLs

Six web pages (i.e., two from each cancer type) were chosen in collaboration with the research team from the final top 100 websites for each of the three types of cancer. Each pair of web pages varied in terms of their readability difficulty, one low and one high. Also, graphical elements such as pictures were removed so that only the text was remained for the participants to read. Appendix A contains the web pages used in this study.

In total, participants read through three web pages of cancer information. The first two web pages were chosen amongst pre-selected pairs of web pages and were randomly assigned according to cancer type. Participants were asked to read about gender consistent or neutral cancer information. That is, men either read about prostate or colorectal cancer, and women read about either breast or colorectal cancer. Colorectal cancer is considered to be neutral because it does not differentiate according to gender. The third web page read was selected by the participants on an individual basis, using search engines that they were taught during an Internet workshop.

4.1.2 Participants

A sub-sample of 16 participants, eight each of males and females, was selected from the overall sample of 44 participants on the basis on how much of the protocol made direct reference to web page content. The mean age of this group was 63.4 yrs, $\underline{SD}$ = 6.53. One participant's specific age was missing, however it was known to be aged 50 or older, which was a requirement stated in recruitment materials. All participants were native English speakers or had been living in Canada for more than 30 years and possessed a high school education or greater. Fourteen participants reported having a personal computer and Internet access in their home.

4.1.2.1 Recruitment

The convenience sample was recruited from Kitchener-Waterloo-Cambridge and surrounding areas. Recruitment was facilitated by the Kitchener Public Library (KPL) and the Cancer Prevention and Early Detection Network of the Waterloo region (CPEDN-WR) via oral announcements, message boards, print publications, and websites. All recruitment procedures were approved by the University of Waterloo Institutional Review Board and the KPL.

4.1.3 Procedure

Once participants agreed to take part in the study, information letters were sent containing an overview of the study and appointment particulars such as date, time, and payment details. The KPL hosted a series of workshops, free of charge, on how to search for cancer information using the Internet. Participants were expected to visit the KPL on two separate occasions:

1) To participate in the Internet workshop led by the consumer health librarian.
2) To participate in the comprehension testing session.

4.1.3.1 Collection of Comprehension Data

The data were collected using semi-structured interviews. Questions about the articles were both specific (e.g., What type of cancer, prevention techniques, screening, etc., are contained within the text?) and non-specific (Pretend you are telling a friend or a family member about the information you read on this web page. Please describe what you would tell them about this article). Prompts such as “Please go on”, “Is there anything else you would add?” were given to the participant during the response to the non-specific question.

All interviews were tape recorded and transcribed. The verbal protocols elicited by both the specific and non-specific questions were used in the present study. The use of transcripts was approved by the University of Waterloo Institutional Review Board prior to beginning the analysis.

All verbal protocols were cleaned prior to analysis. This means that all non-lexical fillers, such as ""uh", "umm", "duh", and so on, and all non-related verbal utterings (i.e., discourse that is not study related) were excluded from the transcripts.

4.2 Analytical Model

The web pages and the participants' transcripts were analysed using two methods: 1) propositional analysis; and 2) semantic network representations. These two analyses are complementary in that the first documents the semantic content of a text in the form of propositions, while the latter provides a graphical representation of the structure and coherence of a text. Propositional analysis provides:

- A list of concepts
- The propositions to be mapped out in the semantic network representations
- The variables needed to calculate propositional density
- The comparison of inferences
- The basis for proposition categorization

Figure 1 is an overall summary of the analysis; the primary study components are on the left, while the list of factors that each component provided are on the right. In the case of Component #3, the comparison provided by propositional density is outlined.

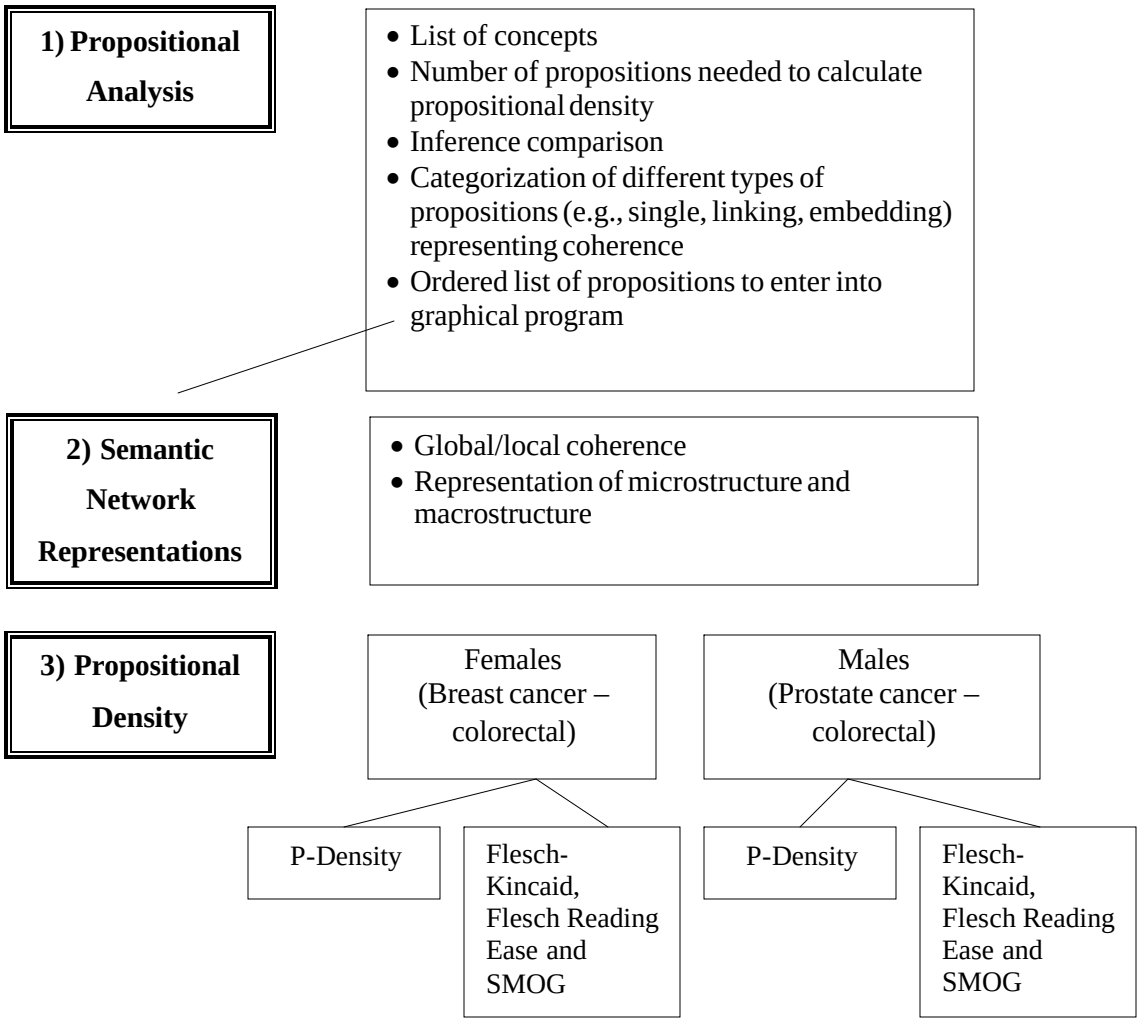


Figure 1. Overall summary of study analysis which includes propositional analysis, semantic network representations, and propositional density.

4.3 Data Analyses

Two methods of semantic analysis, propositional analysis and network representations were used to analyse (a) the six pre-selected web pages the participants were asked to read and (b) their corresponding verbal protocols. The third web page read by participants (i.e., the self-selected site)

was excluded from analyses. In total, the protocols of 16 participants were analysed, yielding 32 protocols (16 participants * 2 web pages per participant).

4.3.1 Web Pages

Descriptive statistics were calculated for the following variables: number of propositions and concepts, number of unique and repeated concepts, as well as the propositional density of the text contained within each web page. The propositional density was then compared to the Flesch-Kincaid, Flesch Reading Ease, and SMOG scores to see if they reflected a similar pattern of information complexity.

4.3.2 Propositional Analysis

Kintsch's propositional system was used to represent both the texts read and the participants' recall protocols (Bovair & Kieras, 1985; Kintsch & Vipond, 1977). In the system, the first step in the analysis is to convert a given text into its propositional textbase. The predicate of a proposition is always written first, followed by its arguments. A predicate can be an action (e.g., affect), an object (e.g., disease), or a relation connecting two or more propositions (e.g., DURATION-OF to describe an action occurring over some period of time). Word concepts are capitalized to distinguish them from the words and they are normally singular. Predicates and arguments are separated by spaces and the whole proposition is enclosed within parentheses. Each proposition is labelled with a number to make individual propositions easy to refer to. An example will best illustrate the general method and format of this representation:

E.g., The new disease affects young children.
P1 (AFFECT DISEASE CHILDREN)
P2 (MOD DISEASE NEW)
P3 (MOD CHILDREN YOUNG)

The proposition P1 is a simple verb frame consisting of the predicate AFFECT and two arguments: the logical subject, the word concept DISEASE, and the logical object, the word concept CHILDREN.

The web pages and the verbal protocols were coded by the student researcher. The coding was then qualitatively reviewed by another researcher to ensure Kintsch's system was followed using Bovair and Kieras'(1985) manual. With propositional analysis, there exist several different coding schemes that vary slightly in terms of format (i.e., order and emphasis). For example, another typically used coding scheme uses the format (argument1 PREDICATE argument2). This format would change the representation of the sentence presented earlier to the following:

E.g., The new disease affects young children.
P1 (disease AFFECT children)
P2 (new MOD disease)
P3 (young MOD children)

As one can see, the representation has changed, but the underlying semantic relations remain the same. The coding scheme selected is not of particular significance; what is important is that there is consistent analysis across all protocols (Bovair & Kieras, 1985). This study, however, used Kintsch's coding system, which is the most widely used in this area of research.

Scoring Recall Protocols. The scoring methods usually used in the propositional framework involve all-or-nothing scoring; a proposition is counted either as recalled, or as not recalled. No partial credit is given (Bovair & Kieras, 1985). Consistency in scoring is the most important aspect of scoring from the experimenter's point of view. It follows then that scoring recall protocols demands that criteria be established so that scoring is consistent.

Criteria can be strict, allowing only close reproductions of propositions, or they can be liberal, for example, "gist" scoring. Bovair and Keiras (1985) have reported that strict scoring is much

easier to perform and results in more consistency than liberal scoring. They also have noted that although scorings of intermediate degrees of strictness and liberality can also be done, there appears to be no advantage in doing so.

This study marked as a correct recall item every proposition in the protocol that corresponded exactly to the message base of the original text. Because participants rarely produce protocols that are word for word exactly like the original text, some degree of latitude and interpretation is necessary. Therefore, close reproductions such as synonyms were marked as correct recall. Synonyms were determined using dictionary definitions.

4.3.2.1 Concepts

Each proposition is made up of word concepts, which are also abstract, and which are different from words. A word concept in a proposition may appear in a real sentence as a word or a phrase. Because of the abstract nature of propositional analysis, a proposition can be represented in several ways in text form. For example, in the sentence "The doctor attended to the patient", ATTEND is the predicator, and DOCTOR and PATIENT are the arguments of the proposition. The sentence, therefore, contains three concepts: ATTEND, DOCTOR, and PATIENT. However, the text may represent this proposition in any of the following forms:

1. The doctor attended to a patient.
2. The doctor is attending to a patient.
3. A patient is being attended by the doctor.
4. The attending of a patient by the doctor.
5. The doctor's attending of a patient.

Concept analysis was conducted after the propositional analysis, because propositional analysis is needed to transform the text read into an ordered list of propositions before concepts can be quantified. The number of concepts contained within the textbase and corresponding recall

protocols was analysed using descriptive statistics to see which concepts were most frequently mentioned and those that were consistently omitted.

4.3.2.2 Inferences

Aside from comparing the textbase to the verbal protocol that a participant provides, one can also determine the parts of the recall which are modifications or inferences. Analysis of text propositions that are transformed in the participant's protocol provides an indication of how many concepts and propositions are from the text or modified by the participants' knowledge.

Inferences are reflected as propositional transformations made on a text (Patel, Arocha, Diermeier, Greenes, & Shortliffe, 2001). They are made on the basis of the participant's prior knowledge and as prior knowledge increases, explanation protocols tend to contain more high-level inferences. An example of a low-level inference is the linking of an anaphor and its antecedent. In contrast, a high-level inference could be defined as a summary proposition derived from the content of a text. For instance, if a text describes a person suffering from a runny nose, sneezing, a sore throat, and a cough, a high-level inference would be that the person has a cold because it is based on all the information provided.

Just as a participant's protocol scoring involves marking as a recall every item in the protocol that corresponds exactly to original text, such transformations made by the participant are scored as inferences. Inference scoring can also be done by establishing semantic categories such as logical deduction, analogies, and part-whole.

Inference scoring was conducted by comparing the ordered list of propositions from the text read to the ordered list of propositions produced by the verbal protocols. Any transformations made were divided into three categories: (1) general to specific (G-S); (2) specific to general (S-G); and (3) same level (SL). Although scoring was conducted by the student researcher, results were discussed with another member of the research team to ensure that inference judgements were acceptable.

4.3.2.3 Propositional Density

While prior studies have defined propositional density as the average number of propositions per 100 words (Hoskyn & Swanson, 2003; Kemper & Sumner, 2001; Kintsch & Keenan, 1973), the present study defined it as the number of propositions divided by the number of words in a sentence. This change in definition was to avoid the problem of multicollinearity, a case of multiple regression in which the predictor variables are themselves highly correlated. It is assumed that longer sentences will contain more propositions and as such, there might be a high correlation between sentence length and number of propositions (Park, 2000).

Propositional density was calculated for the text contained within the six web pages (see Objective 1A). This was done by dividing the text by sentences, recording the number of words per sentence, then dividing the number of words in the sentence by the number of propositions within it. Propositional density was necessarily calculated after the propositional analysis of the text was completed because the propositional analysis supplied the number of propositions contained within each sentence.

After the propositional density was calculated for each sentence, the average P-Density for each web page was computed by summing the P-Density scores of all the sentences and dividing by the total number sentences. Then the P-Density score for the entire document was calculated by finding the average of the individual page P-Density scores.

4.3.2.4 Coherence

Propositional analysis also allows the identification of various types of propositions that serve to indicate the level of coherence of the texts, based on its degree of connectedness (Patel et al., 2001). Propositions can be categorized into three types based on the degree to which they provide coherence to a text: single, embedding, and linking.

1) Single

    - Propositions that express only one idea
    - Self-contained (i.e., they do not refer to other propositions)
    - E.g., “She lost weight”

2) Embedding

    - Identified by the presence, in the argument, of one or more concepts or proposition numbers that refer to other propositions
    - e.g., “She lost weight, resulting in an improvement of the control of her blood sugar”
    - Analysed as two propositions “she lost weight” and “improvement in control of her blood sugar”, plus the link by a resultive relation
    - Propositions about information within paragraphs

3) Linking

    - Propositions containing a relation (i.e., a term in propositional grammar, such as causal, temporal, or logical relations) as a predicate, and at least two propositions or concepts as part of the argument
    - Propositions linking information across paragraphs

Individuals who possess a background knowledge in the area of a text in which they are trying to comprehend tend to increase the number of linking and embedding propositions in their mental model. In the present study, individuals who are familiar with cancer information, relative to those who are unfamiliar, are likely to produce a greater number of these types of propositions. Specifically, males’ verbal protocols about prostate cancer are likely to contain more linking and embedding propositions than for colorectal cancer; similarly for females, this will be the case for

breast cancer as relative to colorectal cancer information. The result of this process is a decrease in the number of inferences needed to interpret the text. Fewer inferences reduce the variability in interpretation, thus making it more easily recalled and understood.

The process of proposition categorization was necessarily conducted after propositional analysis was administered on the text read and corresponding verbal protocols. It was done by the student researcher in consultation with a member of the research team to ensure proper categorization.

4.3.2.5 Semantic Network Representations

Network representations allow the representation, in a graphical form, of the ideas in the text. Because a propositional representation consists of a list of coded propositions, it is difficult to visualize the whole structure of a text or the coherence of an explanation. The network allows the relationships among propositions to be shown, including inferences (i.e., semantic ties between propositions). Appendix B contains an example of a semantic network representation. In general, a node-link-node triplet represents a single proposition in the propositional analysis. Within the structure, propositions that describe attribute information form the nodes of the network and those that describe the relational information form the links.

GraphViz (GraphViz, 2005), an open source graph generation software, was used to construct the semantic network representations. This software is also used in the generation of medical ontologies, for example, in the Protégé Project (Protégé, 2000). Propositional analysis was conducted prior to the construction of the semantic network representations, because each line of code used by GraphViz can be thought of as a single proposition. In this way, the list of ordered propositions produced by a propositional analysis can be thought of as lines of code for GraphViz. Appendix C contains an example of a text, its ordered list of propositions, and its semantic representation as created by GraphViz using each proposition as a node-link-node triplet.

Aside from capturing the semantic meaning of individual propositions, GraphViz's "cluster" feature was used to represent the macrostructure of the text read and the verbal protocols. There is no precise methodology for accomplishing this; it is a more intuitive division or clustering of concepts that fit within a meaningful category. For example, a web page about breast cancer may begin by describing the early signs of breast cancer then move on to citing statistics and facts, after which an early breast cancer detection plan is outlined, and conclude with instructions of how to conduct a breast self-examination. Here, the text and its underlying propositions fall into four categories and can be represented having ranking nodes so that they appear graphically distinct. By doing this, the microstructure of individual elements if preserved, however, the relationships between the microstructure and macrostructure are made evident. The analysis is based on a medical record of a patient experiencing chest pain. The representation is divided into three clusters: (1) description of the chest pain; (2) possible conditions underlying the chest pain; and (3) observable symptoms of the patient. One can see how each category relates with one another, as well as the internal structure within each category.

The connectivity of the text read and that of the participants' recall allowed the investigation of global and local coherence. A network containing global coherence is characterized by connections among all nodes of the network without any contradictions or loose ends. A recall protocol exhibiting local coherence without global coherence would include the recall of isolated components of the text read that are not explicitly linked with the rest of the what is recalled (Arocha, Wang, & Patel, 2005).

CHAPTER 5: Results & Discussion

In this chapter, the web page analyses are presented first and protocol analyses second. A discussion section accompanies each analysis, with the exception of the propositional analysis results, which simply contain samples of different propositions.

5.1 Web Pages

5.1.1 Propositional Analysis

Using propositional analysis, the semantic bases of the web pages were compiled. Each textbase consisted of an ordered list of propositions. Appendix D contains a sample of web page propositional analyses. In this section, examples from the web pages will be used to highlight different propositional types and how predicates were used to represent various expressions.

Example 1 is taken from the low readability breast cancer page. It begins the section about the early signs of breast cancer and was originally formatted as a bullet point. “P1.S1.” indicates that the text is from paragraph 1, sentence 1. The string $ in the first proposition is used as a place holder to denote that the agent who detected the lump remains unspecified. The textbase of this sentence contains three modifier predicates (MOD). In propositions three and four, firm and painless are used in reference to the description of the lump. The forth modifier, often, is a predicate used to represent quantity and is used in a multiple modification. The propositions two and seven use the predicate amount-of to represent a number or quantity. Lastly, the eighth proposition indicates that one thing is the same as another (has the same REFerent). Here, the lump that is detected refers to something that is firm and painless.

Example 1

P1.S1. A lump is detected, which is usually single, firm, and most often painless.

P1	DETECT	$	LUMP
P2	AMOUNT-OF	LUMP	SINGLE
P3	MOD	LUMP	FIRM
P4	MOD	LUMP	PAINLESS
P5	MOD	PAINLESS	OFTEN
P6	AMOUNT-OF	OFTEN	MOST
P7	REF	(P3,P4)	P1

Example 2 is taken from the high readability colorectal web page. This text is part of the section about female hormones which focuses on which women should take hormone replacement therapy (HRT) and the therapy's associated risks. Proposition three is an embedding proposition as it refers to proposition four as an argument. Proposition five is a linking proposition because but is used as a logical relation between two propositions; semantically, it can be read as "HRT lowers risk (P1) but HRT increases risk (P6)". The modifier possible in proposition seven indicates that the degree of uncertainty denoted by the meaning of the word may. Propositions nine through 12 provide a list of all the possible risks associated with HRT.

Example 2

P7.S2. HRT also lowers the risk of developing osteoporosis, but it may increase the risk of heart disease, blood clots, and breast and uterine cancer.

P1	LOWER	HRT	RISK
P2	MOD	LOWER	ALSO
P3	OF	RISK	P4
P4	DEVELOP	$	OSTEOPOROSIS
P5	BUT	P1	P6
P6	INCREASE	HRT	RISK
P7	MOD	INCREASE	POSSIBLE
P8	AND	(P9,P10,P11,P12)	
P9	OF	RISK	HEART-DISEASE
P10	OF	RISK	BLOOD-CLOT
P11	OF	RISK	BREAST-CANCER
P12	OF	RISK	UTERINE-CANCER

Example 3 is taken from the low readability prostate cancer web page. It is part of the text that talks about how to ensure that enough antioxidants are included in one's diet. This example serves to demonstrate the wordiness and the grammatical devices that are sometimes present in literary prose. With a sentence like this, the first step of the analysis is to begin with the central proposition. This proposition will often contain a verb as its predictor. Here, it is about increasing tomato consumption. All the other propositions are structured around this central proposition. One can see that proposition two uses the first proposition as an argument, and that proposition 20 uses it as part of its linking argument.

Example 3

P19.S5. It may therefore be a good idea, whilst awaiting further evidence, to try and increase your tomato consumption to one portion a day – this may take any form, either fresh ripe tomatoes or processed tomatoes e.g. tomato juice, soup, tinned chopped or plum tomatoes.

P1	INCREASE	$	TOMATO-CONSUMPTION
P2	TO	P1	PORTION/DAY
P3	NUMBER-OF	PORTION/DAY	ONE
P4	FORM-OF	P5	PORTION
P5	OR	P5	P8
P6	REF	TOMATO-CONSUMPTION	TOMATO
P7	MOD	TOMATO	RIPE
P8	MOD	TOMATO	FRESH
P9	REF	TOMATO-CONSUMPTION	PROCESS-TOMATO
P10	OR	(P1,P12,P13,P14)	
P11	EXAMPLE-OF	PROCESS-TOMATO	TOMATO-JUICE
P12	EXAMPLE-OF	PROCESS-TOMATO	SOUP
P13	EXAMPLE-OF	PROCESS-TOMATO	TIN-CHOPPED
P14	EXAMPLE-OF	PROCESS-TOMATO	PLUM-TOMATO
P15	REF	IDEA	P1
P16	MOD	IDEA	GOOD
P17	MOD	GOOD	POSSIBLE
P18	AWAIT	$	EVIDENCE
P19	MOD	EVIDENCE	FURTHER
P20	WHILE	P1	P18

5.1.1.1 Descriptives

5.1.1.1.1 Results

The number of words contained within each web page is presented in Figure 2. The average number of words for the low and high readability web pages was 908 (range 575-1287) and 1036 (range 578-1552) respectively. The colorectal web pages contained a similar number of words; however, there was a large word difference for the other two types of cancer pages. The low readability breast cancer page contained less than half the number of words as its high readability counterpart (575 vs. 1552), while the low readability prostate cancer page contained more than double the number of words than its high readability version (1287 vs. 578).

The number of propositions within each web page is given in Figure 3. Overall, the high readability web pages contained a greater number of propositions than the low readability pages. The prostate cancer text was the only one where the low readability page exceeded the high. The number of concepts contained in the web pages is presented in Figure 4. It shows that the low readability pages contained more concepts than the high readability pages. Looking individually by cancer type, one can see that the exceptions to this pattern were the prostate cancer web pages. Here, the low readability prostate text contained more than double the number of concepts than the high readability text.

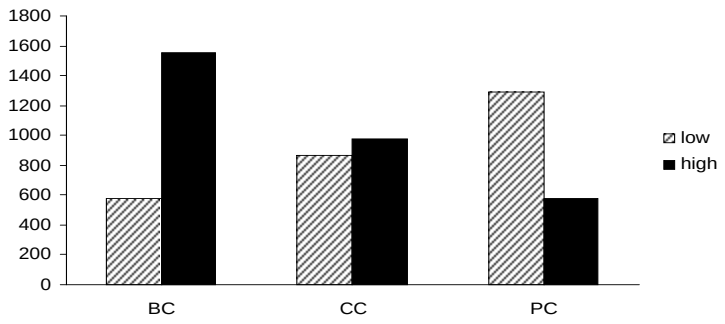


Figure 2. Word count for web pages on breast cancer (BC), colorectal cancer (CC), and prostate cancer (PC) at low and high readability levels.

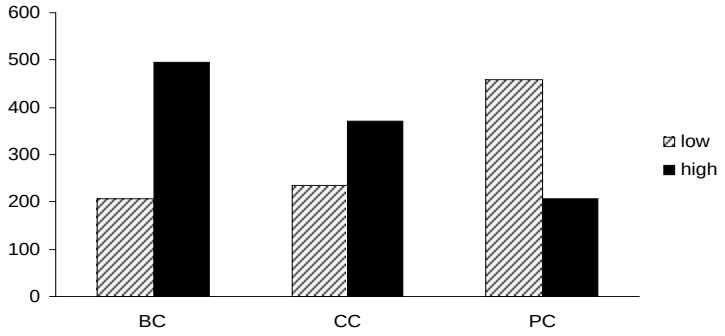


Figure 3. The number of propositions for web pages on breast cancer (BC), colorectal cancer (CC), and prostate cancer (PC) at low and high readability levels.

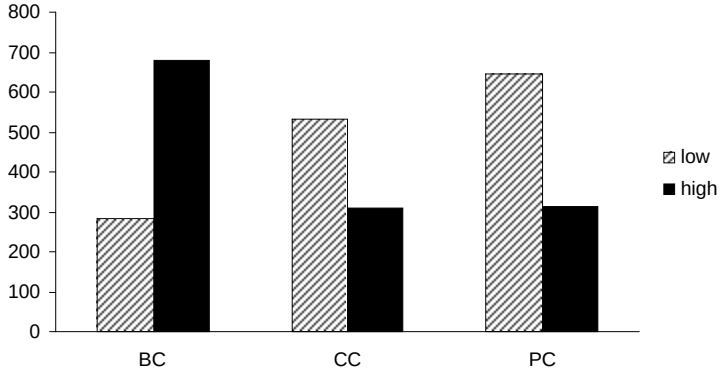


Figure 4. The number of concepts for web pages on breast cancer (BC), colorectal cancer (CC), and prostate cancer (PC) at low and high readability levels.

5.1.1.1.2 Discussion

It is interesting that for these six web pages, there were three different patterns of readability and word count. In one case, word count increased with reading difficulty, in another it decreased, and in the third there was no difference. Intuitively, one might think that a lower readability text would need to employ a greater number of words to convey a message.

Given that a proposition is composed of word concepts, it is reasonable to expect that the greater the number of propositions, the number of concepts would increase in turn. Recall that the number of propositions was greater in the high readability pages; this should mean that these pages should also contain more concepts. However, the results show the opposite, an overall greater number of concepts in the low than in the high readability pages. This means that any given low readability proposition contained more concepts, either total or unique, than a high readability proposition, despite the greater number of overall high readability propositions.

Looking at each cancer type individually, however, one notes that it is the colorectal cancer texts that break the sequence of the number of concepts increasing with the number of propositions. That is, this text produces an increase of propositions from low to high readability, but a decrease in the number of concepts for the same change in readability. This suggests that there may be something specific to the colorectal cancer text itself that is disrupting the relationship between the number of concepts and propositions. Indeed, about half of the low readability page was in bullet point format, whereas the high page contained none. Bullet points can disrupt the propositional content because they usually represent shortened sentences, abbreviated to the point where propositions can be incomplete (i.e., containing singular concepts without an obvious relational link). For instance, one bullet pointed section listed the problems whose occurrence should warrant a doctor consultation. Vomiting and constant tiredness were among those problems, which are concepts that contain two propositions at the most:

e.g., Vomiting
P1. (PROBLEM VOMITING)

e.g., Constant tiredness
P1. (PROBLEM TIREDNESS)
P2. (MOD TIREDNESS CONSTANT)

This results in a concept to proposition ratio of 1:1 and 1:2 respectively, which skews the concept and proposition relationship. The only other page to use bullet points was the high readability breast cancer one, but in that case, it was limited to about half the page, and the bullet points tended to be complete sentences rather than single concepts.

5.1.1.2 Propositional Density

5.1.1.2.1 Results

For the low readability web pages, the propositional density (P-D) scores varied from 0.39 to 0.44. Table 3 compares the P-D scores with the readability results of the SMOG, F-K, and FRE formulas. The SMOG scores indicate a similar pattern as the P-D scores, where the colorectal, breast, and prostate cancer pages respectively are ranked in increasing difficulty and density. However, since the SMOG is scored by hand and is applied only to select passages of the text, the scores may vary depending on which text excerpts are selected for analysis. For the F-K scores, only the prostate cancer page matched the P-D score as the most difficult text to read. The FRE scores show yet another pattern of difficulty, where the low score of 48.5 for the prostate cancer material indicates more difficult reading than the easier breast cancer text, with the score of 66.7.

For the high readability pages, the P-D scores varied from 0.35 to 0.38. Here, there appears to be no discernable pattern of agreement across the scores. For example, the colorectal cancer page is the most propositionally dense with a P-D score of 0.40. The SMOG score of 15 agrees that this page

has requires the highest grade level in order to be understood, however the F-K and FRE scores show disagreement with both the P-D and SMOG scores, and with each other.

Table 3. Propositional Density and Readability Scores for Breast Cancer, Colorectal Cancer, and Prostate Cancer at Low and High Readability Levels

	Low				High			
	P-D	SMOG	F-K	FRE	P-D	SMOG	F-K	FRE
BC	0.39 (2)	10 (2)	6.2 (1)	66.7 (1)	0.35 (1)	13 (1)	12 (2)	44.6 (1)
CC	0.36 (1)	9 (1)	8.7 (2)	51.5 (2)	0.40 (3)	15 (2)	12 (2)	31.2 (3)
PC	0.44 (3)	11 (3)	9.9 (3)	48.5 (3)	0.38 (2)	13 (1)	11.8 (1)	42.5 (2)

Note. The number in brackets represents the rank order of increasing difficulty, e.g., (1) is the easier while (3) is the most difficult.

5.1.1.2.2 Discussion

The variability of the P-D scores for both the low and high readability web pages is unfortunately too narrow to make informed statements about how P-D relates to standard readability formulas. Even amongst the readability formulas themselves, there appears to be inconsistency. For example, the high readability breast and colorectal pages both have F-K scores of 12, indicating the same grade level; however their related FRE scores are almost 15 points apart. This speaks to the limitations of computerized readability formulas, which in this case, were perhaps influenced by the formatting of the pages (e.g., interpreting bullet points as periods). Of the three formulas, the SMOG appears to be most in line with the P-D results. This is encouraging, as the SMOG grading formula has been used widely and has been adopted by the National Cancer Institute as the preferred method for assessing the readability of patient communications after a comprehensive review of advantages and disadvantages of alternative readability formulas (Romano, 1979). Still, the disparity between the SMOG and F-K scores suggest that more than one readability formula should be used when

determining text difficulty. Using multiple readability formulas is not done on a routine basis because most users access the formulas through document processing programs that often contain only one formula. For example, Microsoft Word calculates scores for only the FRE and F-K readability formulas. These results also support question 1A, which asked if there would be as association between F-K, FRE, and SMOG scores and P-D. It appears that P-D and the readability formulas are indeed measuring different factors and should be considered to provide distinct contributions to overall comprehension measurement.

However, the use of P-D as an indicator of reading difficulty is supported by its strong theoretical basis. P-D asserts that the greater the number of propositions per fewest words, the greater is the resulting reading difficulty. Several studies have shown that propositionally dense text is difficult to process and integrate for later recall and comprehension (Kintsch, 1994; Kintsch & Keenan, 1973; Kintsch & van Dijk, 1978). This finding may be related to limitations in working memory capacity that preclude holding large amounts of information simultaneously. If propositions are not well integrated into the working knowledge representation, then the information may not be available for later recall (Kintsch, 1994; Kintsch & van Dijk, 1978). Also, increased propositional density in text is related to increased processing and item difficulty (Embretson & Wetzel, 1987; Kintsch & Keenan, 1973). Furthermore, it is more difficult to answer comprehension and inference questions on text that is propositionally dense, especially if it includes many modifier propositions such as adjectives and adverbs (Embretson & Wetzel, 1987; Kintsch & van Dijk, 1978).

5.1.1.3 Coherence

5.1.1.3.1 Results

An assessment of coherence on the basis of a text's underlying meaning requires an examination of the propositions and how they relate to one another. As such, embedding and linking propositions provide a measure of coherence because they both in turn refer to other propositions, whether through direct reference or through a relational predicate. The different types of propositions by cancer type and readability level are given in Table 4.

The degree of connectedness ranged from a low of 0.40 for the low readability breast cancer page to a high of 0.58 for the low readability colorectal cancer page. For both the colorectal and prostate texts, the low readability page contained a greater proportion of embedding and linking propositions than the high pages. Only for the low readability breast cancer text was coherence worse than in its high readability counterpart.

Table 4. Web Page Categorization: Single (S), Embedding (E), and Linking (L) Propositions at High and Low Readability Levels

	Readability Level									
	Low					High				
	S	*E*	*L*	*E+L*	*Prop*$_{(E+L)}$	*S*	*E*	*L*	*E+L*	*Prop*$_{(E+L)}$
BC	124	44	38	82	0.40	260	123	103	226	0.47
CC	98	44	93	137	0.58	196	60	116	176	0.47
PC	246	97	116	213	0.46	118	27	61	88	0.43

Note. BC – Breast cancer; CC – Colorectal cancer; PC – Prostate cancer; Prop$_{(E+L)}$ – Proportion of embedding and linking propositions to total number of propositions.

5.1.1.3.2 Discussion

Propositions that have overlapping arguments, and thus are semantically related, create coherence. The literature that shows that coherence facilitates comprehension (Kintsch & Vipond, 1979; Miller & Kintsch, 1980), and although readability formulas capture a component of comprehension, coherence is not part of that component. The results support question 1B that low readability scores would not ensure high levels of coherence, and that high readability scores would not ensure low levels of coherence. Of the three cancer types, only two of them had greater coherence for the low than for the high readability pages; the difference for the prostate cancer pages was marginal at 0.03.

5.1.1.4 Semantic Network Representations

5.1.1.4.1 Results

Figures 5, 6, and 7 contain a selection of partial web page semantic network representations (SNRs). The complexity and size of the complete SNRs of the web pages makes it difficult to present, therefore it was decided to show only clusters of the web pages. These clusters are detailed enough to demonstrate the elements of local and global coherence on a smaller scale. Also included in the SNRs are the main and subordinate ideas within each cluster. This provides some insight as to how ideas that differ in terms of importance relate with each other and with all other elements. Subheadings within each web page were used as natural divisions for determining the clustering of information.

A cluster about breast cancer prevention strategies written in high readability language can be found in Figure 5. Main ideas are identified by rectangular boxes; subordinate ideas are enclosed in diamond boxes. The information within this cluster discusses prevention strategies that have been recommended to women with BRCA1 and BRCA2 mutations. The way to read through the network is to start with one node, follow the arrow, and use any relational information to see how the starting

node relates to the ending node. For example, one prevention option is breast removal through a preventative mastectomy. To read this, begin with the option node in grey on the upper right side of the cluster. Then follow the arrow that connects it to the breast node, then again to removal, and finally ending at preventative mastectomy. Note that two options were mentioned in the web page: (1) a standard approach that does not exist; and (2) the preventative mastectomy which is considered to be drastic. One can see than the main idea tamoxifen has good local coherence, as demonstrated by its connections with several other nodes. In fact, on a whole, this cluster has good local and global coherence. There no components that are not explicitly linked to the rest of the text, and there are connections among all nodes without loose ends.

A cluster about high prostate-specific antigen (PSA) written in high readability language is given in Figure 6. The information in this cluster discusses the two possible causes of high PSA, namely a benign noncancerous enlargement of the prostate called benign prostatic hyperplasia (BPH) and prostate cancer. These two causes are highlighted in boxes as main ideas. One can see that many of the resulting arrows arrive at the third main idea of prostate cancer. This cluster also has good local and global coherence. All the components are connected with one another and there are no noticeable loose ends.

In the final example, Figure 7 contains a cluster about prognosis and treatment of colorectal cancer written in low readability language. This cluster discusses certain factors that affect a person's chance of recovery and treatment options. One can see that a couple of the main factors (and hence the main ideas) are the stage of the cancer and whether the cancer has reoccurred. Once again, this cluster's local and global coherence are good. All ideas are interconnected and there are no loose ends.

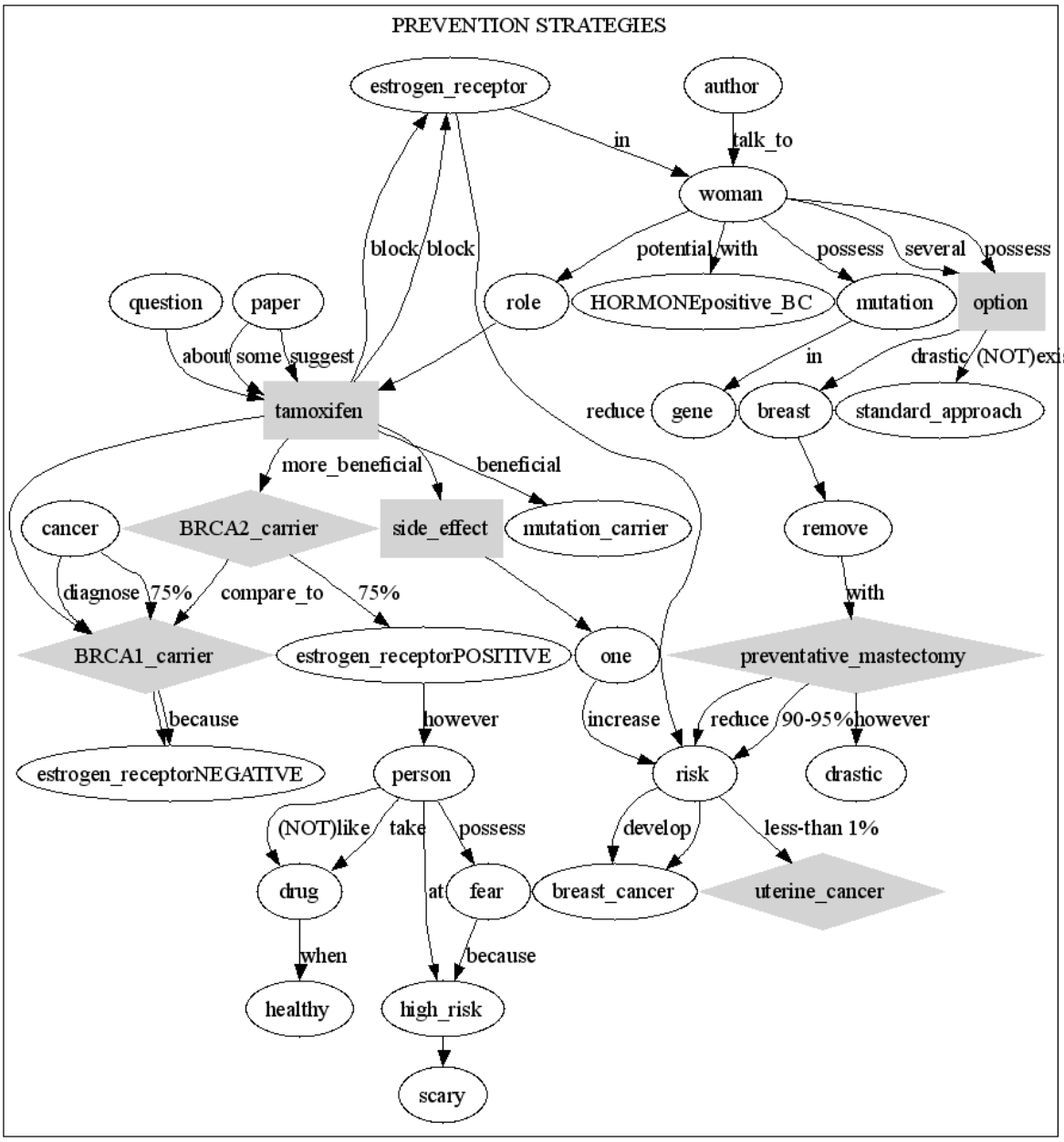


Figure 5. A SNR of a cluster about prevention strategies from the high readability breast cancer web page.

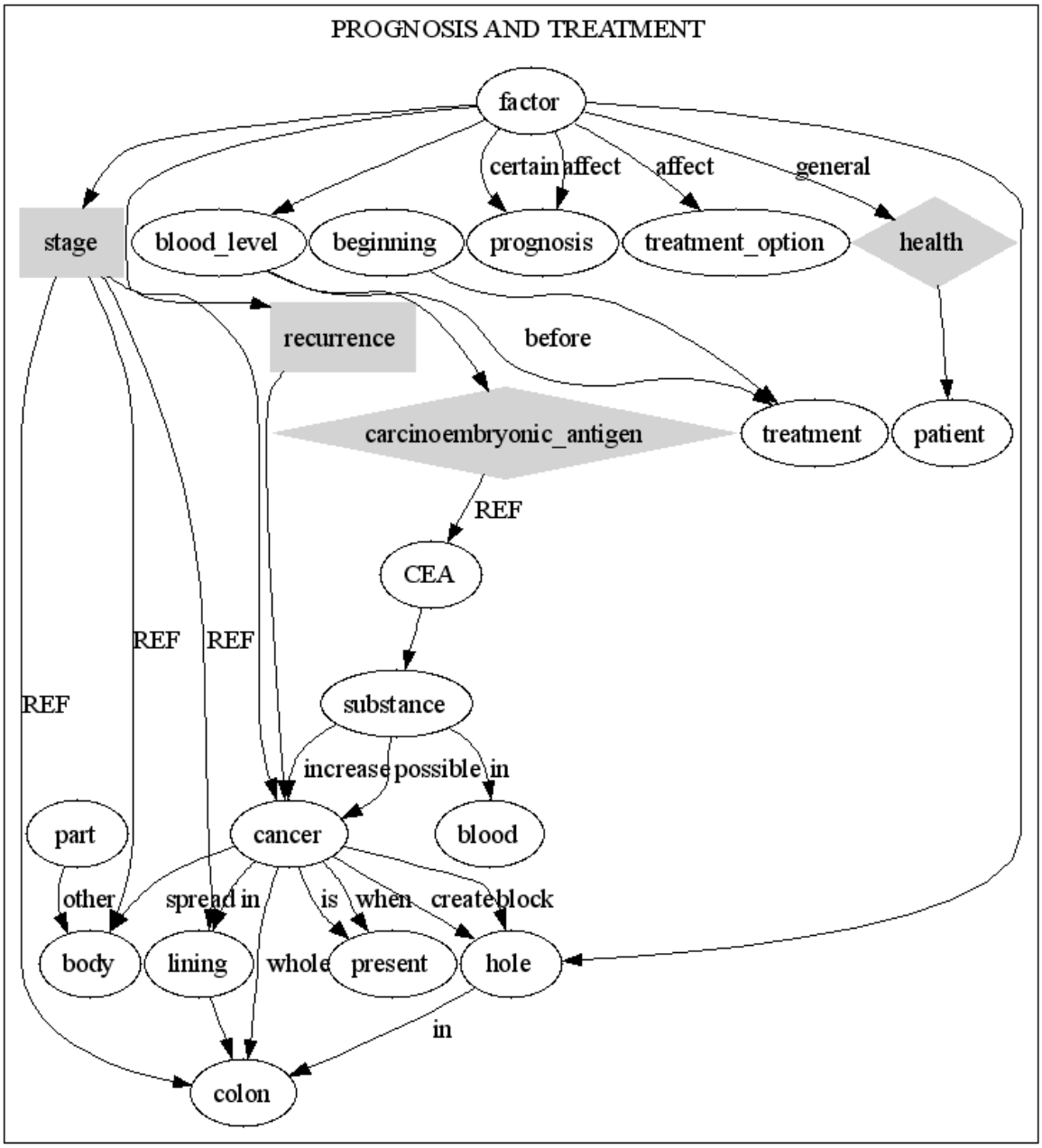


Figure 6. A SNR of a cluster about prognosis and treatment from the high readability prostate cancer web page.

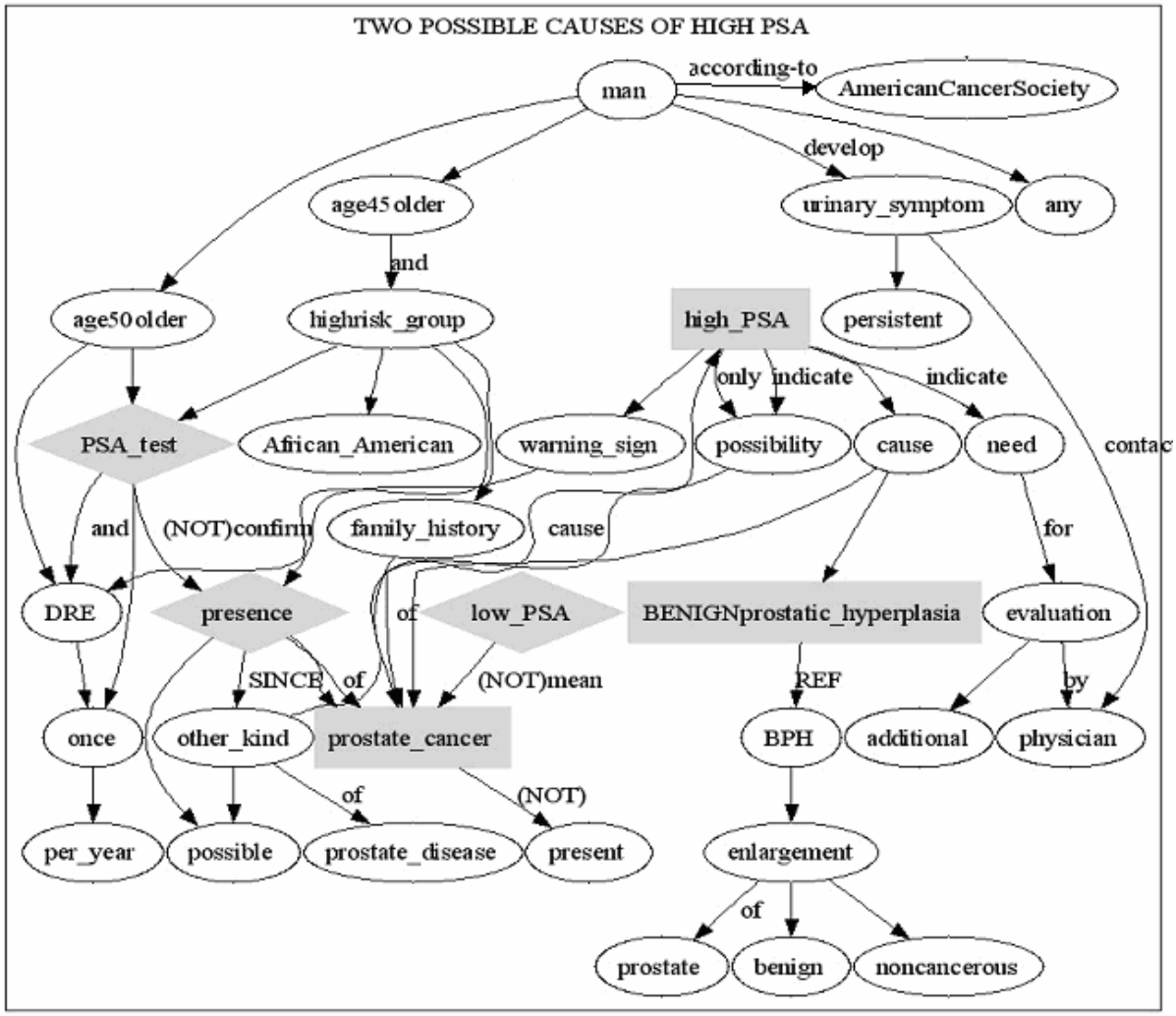


Figure 7. A SNR of a cluster about two possible causes of a high PSA from the low readability colorectal cancer web page.

5.1.1.4.2 Discussion

Although only three SNRs are illustrated, they are quite representative of the all the other SNRs about breast, prostate, and colorectal cancer, each varying in terms of readability level. There were no SNRs that had even minor breaks in either local or global coherence. Question 1B of this study projected that differences in coherence due to changes in readability level would be discernable through SNRs. Unfortunately, this was not the result. Perhaps it is because these representations are

not a sensitive enough measure. Another reason may be because of the repetition of concepts within each web page. For example, in the high readability prostate page, PSA was mentioned 11 times, prostate 12 times, biopsy 10 times, and physician seven times. In fact, only about half the total number concepts in the page were unique (i.e., mentioned only one time). Since so many of the concepts were repeated, this increases the likelihood that any given concept would be connected to another node-link-node triplet instead of remaining isolated. This, in turn, increases the coherence. Still, the SNRs serve as a good illustrative tool to demonstrate connections between concepts of a text. It is easy to see which concepts are mentioned multiple times; all one has to do is count the number of arrows originating and ending at each node.

5.2 Verbal Protocols

5.2.1 Propositional Analysis

One characteristic upon which the propositional analysis of the web pages and the protocols differ is that the protocols included more idiomatic expressions. Also, in contrast to writing on paper or typing on a keyboard, responses that are explained out loud tend to be organized more as a flow of ideas than as standard prose. For that reason, it was necessary to separate non-web page related verbalizations from the protocol content before beginning analysis. Below is a sample of a response to the general question from participant 1 talking about the high readability text; parts that are underlined represent the verbalizations that were used for analysis, while the parts that were inaudible on the tape are signified by blank spaces.

> "Well it can be prevented if you see to your general health and wellbeing, that's a protection, it can't protect you if it's a genetic problem, that may happen anyway even if you do all the things, but if you keep track of, if you get check-ups, if you know you have a genetic predisposition or a bowel predisposition you should get tested, and I guess get advice from a doctor of things to do, many of the things that _____ and it

didn't stress this in the article but fresh, clean water is very important _____ so I would stress water as well as vegetables and, and is this the article where, I don't think I read very much of that one so it must be this one where they said vitamin D (yeah) so you should have vitamins especially folic acid and you should have vitamin D and not too much sun because there's problems with skin cancer if you get too much sun so it's better to get it from a pill, fresh fruits and vegetables are good, whole grain cereals which I have to be careful with because I'm a little bit sensitive to them, bran and wheats I've got to stay away from but and there maybe other people that are sensitive like that too, got to be careful and I think for inflammation and stuff like that, I think maybe that what you're allergic to at some point you pay attention to that and that's not a general ___but I think it said that specifically in here…"

The propositional analysis of this protocol is as follows:

P1	IS	GENERAL-HEALTH	GOOD
P2	IS	WELLBEING	GOOD
P3	AND	P1	P2
P4	PREVENT	$	COLORECTAL-CANCER
P5	IF	P4	P3
P6	IS	P3	PROTECTION
P7	IS	COLORECTAL-CANCER	GENETIC-PROBLEM
P8	NEGATE	P4	
P9	IF	P8	P7
P10	GET	$	TEST
P11	POSSESS	$	GENETIC-PREDISPOSITION
P12	POSSESS	$	BOWEL-PREDISPOSITION
P13	OR	P11	P12
P14	IF	P13	P10
P15	POSSESS	$	VITAMIN
P16	EXAMPLE-OF	VITAMIN	FOLIC-ACID
P17	MOD	FOLIC-ACID	ESCECIALLY
P18	EXAMPLE-OF	VITAMIN	VITAMIN-D
P19	IS	FRUIT	GOOD
P20	MOD	FRUIT	FRESH
P21	IS	VEGETABLE	GOOD
P22	IS	CEREAL	GOOD
P23	MOD	CEREAL	WHOLE-GRAIN

The idiomatic expression “if you see to your general health and wellbeing” was propositionalized as general health and wellbeing being good. In their manual, Bovair and Keiras (1985) explain that it is often pointless to try and propositionalize an idiom as it stands and therefore the representation should normally be based on the meaning of the idiom. Proposition five is a linking proposition because it contains the logical relation, if. This proposition can loosely be understood as “colorectal cancer can be prevented if general health and wellbeing are good”. Proposition 13 is another linking proposition, which represents that if a person has a bowel or genetic predisposition that he or she should get tested.

Appendix E contains a propositional analysis of the six web page categories, one each from breast, prostate, and colorectal cancer at low and high readability levels. Each analysis includes the interview questions, the protocol response, and then the list of ordered propositions.

5.2.1.1 Descriptives

5.2.1.1.1 Results

The more simply written pages elicited a greater number of words from 12 of the 16 participants. Overall, females produced more verbose protocols for the colorectal cancer web pages than for the breast cancer ones (1177 vs. 780 words respectively). Their mean word count for the low readability colorectal pages was about 40% greater than for the high readability ones (735 vs. 442). In contrast, the protocols produced by males for the gender specific prostate cancer information were lengthier than those for colorectal cancer (894 vs. 513). This was primarily due to participants 15 and 16, whose prostate protocols were twice as long as the other male participants. These two males were the only ones for which readability level had a noticeable influence; both of their low readability

protocols were around three times wordier as their high ones. Readability level was not associated with colorectal protocol word count.

A similar pattern emerges for the number of propositions produced by the female and male participants, as shown in Figures 8 and 9. Again, 12 of the 16 participants' low readability protocols contained more propositions than their high ones. This was anticipated because as word count varies, so should the number of propositions. The proportion of propositions to word count remains relatively stable across readability level and cancer type, with a range of 0.28 to 0.37. Also, collapsing across gender, the low readability cancer information consistently produced a greater number of propositions.

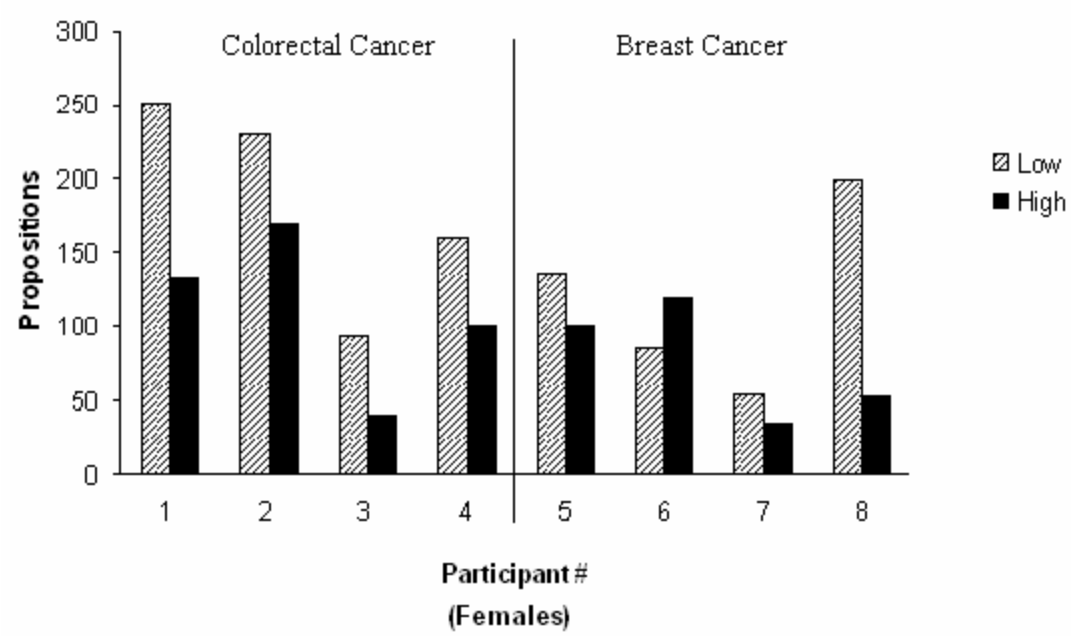


Figure 8. The number of propositions in female protocols for colorectal and breast cancer at low and high readability levels.

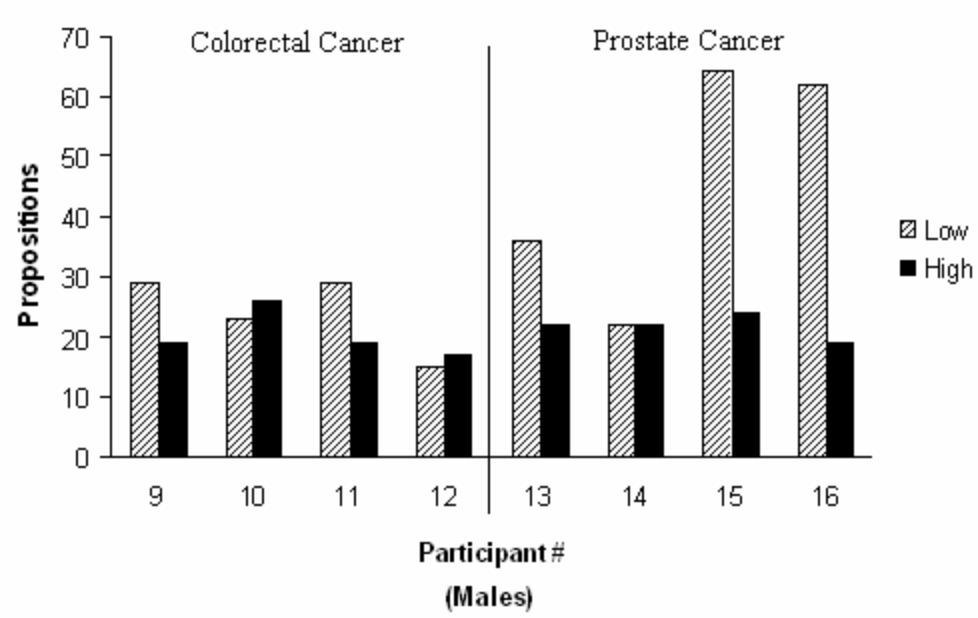


Figure 9. The number of propositions in male protocols for colorectal and prostate cancer at low and high readability levels.

As given in Figures 10 and 11, a greater number of unique concepts were produced by females for the low than for the high readability pages. The colorectal web pages elicited a greater number of unique concepts than the breast cancer ones. Generally, if a female participant generated a high number of concepts at one readability level, the number was also high for the other, when compared to the other participants. The exception to this pattern is participant 8, whose protocol contained 80 total concepts in the low readability breast cancer page, but only 21 for the high page, which ranks it among the lowest of all the protocols. Participant 8 is a 57 year old retiree, who at the time of testing, was taking a computer course and learning to use the Internet. She said that she preferred getting her health information from the library or simply by just talking to other people, and had previously read up about lymphoma. When asked what she thought about the web pages, she shared that she liked the low readability page because “it explained [the cancer] very simply, the basics of how it starts and what happens with the different tumours and how it can spread”. She also

said that she found the high readability page interesting, because she had been unaware about the two mutation types, BRCA1 and BRCA2. Going over her protocol question by question, one notes that while she was able to provide responses to both the general and follow-up questions, her high readability responses contained few actual concepts despite being wordy. For example, when asked what procedures are recommended for women with the mutations, she answered: "Well, they were certainly saying that the MRI was the way to go on it". That response only contains one concept, MRI. In contrast, her low readability responses contained a higher frequency of concepts, as illustrated by the following quote describing early signs of breast cancer:

> "Well, usually you feel a lump and sometimes people have a discharge from one or both nipples. Sometimes the nipple inverts itself a little. I believe they mentioned something about tenderness and soreness that was persistent, that didn't just go away in a week or two".

This characteristic of her protocol is likely the reason why there is such a discrepancy in the number of concepts generated in response to the different web pages.

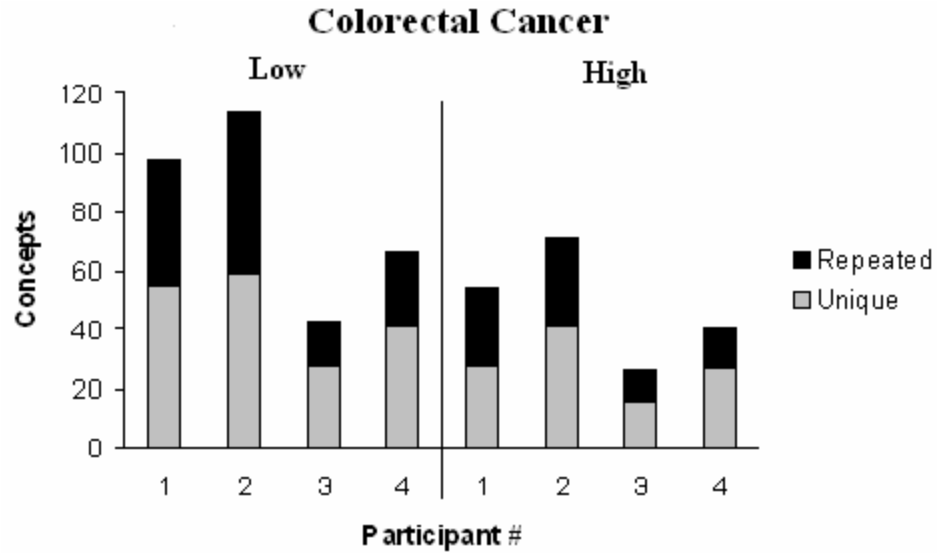


Figure 10. The number of repeated and unique concepts generated by females for the low and high readability colorectal cancer web pages.

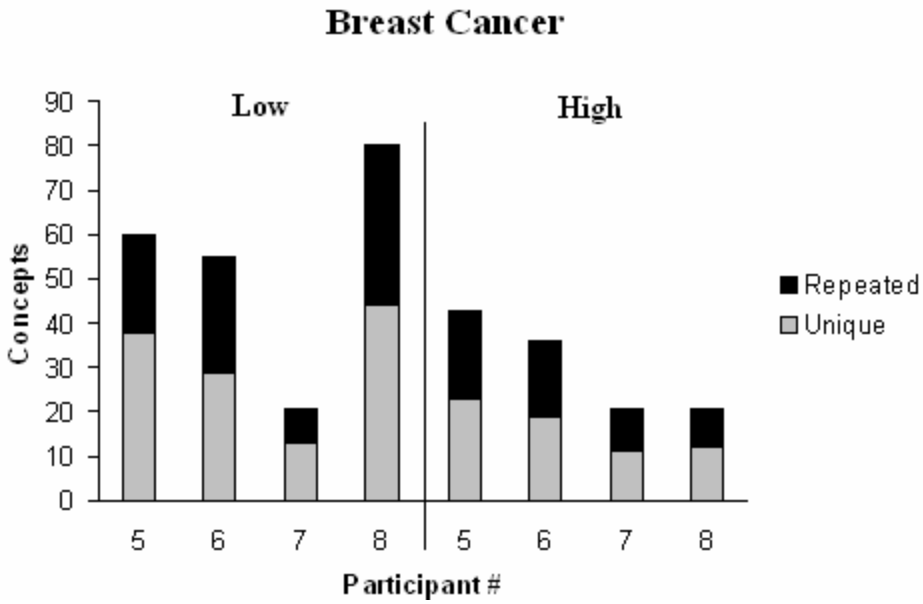


Figure 11**.** The number of repeated and unique concepts generated by females for the low and high readability breast cancer web pages.

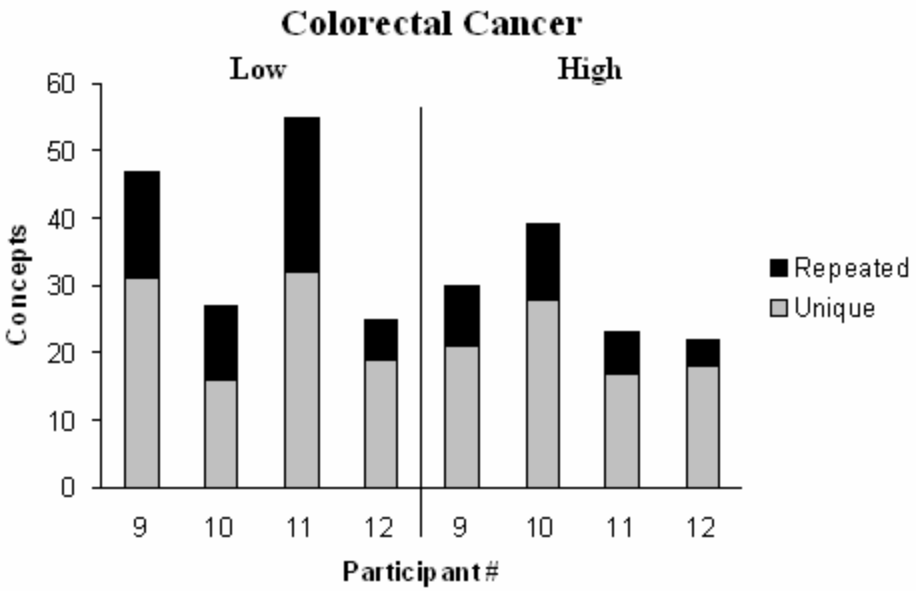


Figure 12. The number of repeated and unique concepts generated by males for the low and high readability colorectal cancer web pages.

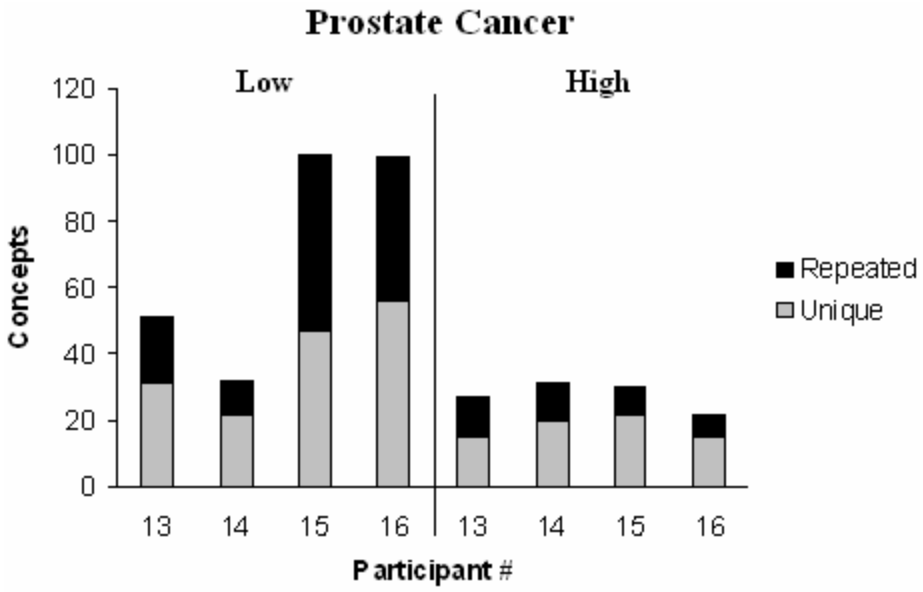


Figure 13. The number of repeated and unique concepts generated by males for the low and high readability prostate cancer web pages.

As given in Figures 12 and 13, overall, there were a greater number of concepts produced by males for the prostate cancer web pages than for the colorectal ones. However, this was greatly contributed to by participants 15 and 16, who both generated about twice as many concepts as the next highest participant. Also, the low readability pages elicited more concepts that the more difficulty written ones. Again, both these results follow the same pattern of results of the word and proposition counts. The low readability pages for both cancer types contained a greater proportion of repeating concepts, compared to the high readability pages. Some frequently mentioned concepts for breast cancer were mammogram, self-examination, and screening. For prostate cancer, protocols often included talk about antioxidants, balanced diet, and PSA-test. Finally, some of the repeated concepts in colorectal cancer protocols were digestive system, gastrointestinal problem, polyps, risk factors, and lifestyle.

Of the four different groups (i.e., colorectal low and high; prostate low and high), the high readability prostate text consistently elicited the fewest number of concepts from the participants. This web page described the digital rectal exam, the prostate-specific antigen test, the transrectal ultrasound and the biopsy, all tests used for the diagnosis of prostate cancer. Similar to participant 8, there was a large discrepancy between the number of concepts generated by participant 16 in his low and high readability protocols (99 vs. 22 respectively). This 65 year old male participant reported getting his health information from the Internet as well as from his person doctor. He shared that he had found both web pages to be interesting, especially the low readability one, because it dealt with how cancer rate is rising in people who move to the west from the east and adopt western lifestyles. His interest seemed to manifest itself for the low readability page by producing a protocol that contained a high number of concepts, half of which were repeated more that once. He spoke about food quite a lot, first in his response to the general question, then again to one of the follow-up questions. With regards to food, he said:

> "This one sort of gives you suggestions as far as diet is concerned, frying is not good, poaching or baking is much better… French fries and potato chips, and red meats, restrict your diet of that and you're better off with fish. This is an area that I found quite interesting, was the suggestion of fruits and the amount of portions that you should eat in a day, up to five combinations of fruits, one specifically was pineapple which I wasn't aware of…"

This interest in nutrition and diet is likely to have been one of the primary reasons for his conceptually dense protocol, as compared to his protocol for the high readability page. For that page, he was very general, even to the point of being vague. For example, answering the general question, he said depending on what his friend or family already know, he would either give them a print out of the web page, or direct them to the Canadian or American Cancer Society.

5.2.1.1.2 Discussion

Regardless of the proportion of the verbal protocols that was actually correct, females were more apt to talk about colorectal than about breast cancer information. Furthermore, the use of simpler language appeared to encourage their disposition to talk. Readability level had no noticeable influence on the length of male protocols, with the exception of participants 15 and 16, who spoke more about the low readability that high one. The structure of the protocols for these two participants was comparable to those of others; that is to say that they both responded to the general question and follow-up questions with no preference for any given topic. It is important to keep in mind, however, that for both males and females, the aptness to talk for low readability page did not necessarily translate into better comprehension. Participant 16's recall, for example, was comparable at both readability levels, despite having a much lengthier low readability protocol.

Examining the low readability colorectal cancer web page itself reveals a degree of repetition of 0.47. This number is not noticeably different from the repetition of the other web pages, so the reason why it would produce the greatest degree of concept repetition among the verbal protocols is unknown. Although it is not included as a property of readability formulas used in this study, the repetition of arguments does have an impact on readability and comprehension. By using a repeated concept to link the propositions underlying two sentences, a reader can often interrelate the propositions without ever considering larger unit of meaning (Smith & Swinney, 1992). If one viewed reading time as a function of readability, then a greater number of repeating concepts should translate into easier reading and faster reading times. Indeed, it has been found that people take less time to read a paragraph that contain a few often-repeated concepts than one that contained numerous rarely repeated concepts (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975).

The protocols of participants 8 and 16 were discussed in detailed for the difference in the number of concepts in their protocols for the low and high readability pages. Both participants

generated more concepts for the low readability than the high readability page. In terms of demographic factors, such as English competency and prior knowledge, the two participants had similar backgrounds. However, participant 8's verbal style seemed to contribute to the difference, while participant 16's interest level seemed to change in the same direction as his concept production. Whether these two factors had a real or hypothetical influence on the number of concepts produced, these examples do serve to illustrate the many considerations that exist.

5.2.1.2 Recall

5.2.1.2.1 Results

Recall was based on responses to both the general ("Pretend you are telling a friend or family member about the web page) and follow-up questions about specific content. The propositional analysis for each verbal protocol was directly compared to the propositional analysis of the appropriate web page, enabling a side-by-side analysis. Scoring for correct recall was strict, with some allowances made for small variations. For example, when asked for some possible signs of colorectal cancer, one partial response was "cramp" instead of "cramping". Another participant expressed that a person is more likely to get cancer if the "cancer is recurring" rather than "a recurrence of cancer" as stated in the web page.

Recall by female participants for the colorectal and breast cancer web pages is given in Table 5. The numbers represent the number of propositions recalled that exactly match propositions in the web pages. With regard to readability level, more propositions were remembered from the low readability pages for both cancer types.

Table 5. Female Propositional Correct Recall of Low and High Readability Colorectal and Breast Cancer Information by Non-prompted (NP) and Prompted (P) and Questions

	Colorectal Cancer					Breast Cancer			
	Low		High			Low		High	
Participant	NP	P	NP	P	Participant	NP	P	NP	P
1	6	3	4	5	5	0	4	0	0
2	13	10	15	3	6	1	0	4	0
3	0	7	1	7	7	3	1	1	0
4	2	13	0	7	8	0	3	2	0
Total	*21*	*33*	*20*	*22*	*Total*	*4*	*8*	*7*	*0*
% Overall	*0.25*		*0.33*		*% Overall*	*0.08*		*0.08*	
% Inference	*0.13*		*0.13*		*% Inference*	*0.12*		*0.15*	

Note. % Overall and % Inference denote the proportion of correctly recalled propositions and the proportion of inferences out of the total number of propositions in a protocol respectively.

Correct recall accounted for a larger proportion of overall propositions protocols for colorectal than for breast cancer. For example, for one-third of the propositions in high readability colorectal protocols correctly matched propositions in the web page, compared to only 8% for the breast cancer protocols. Neither cancer type nor readability level, however, appeared to influence inference production. Of the total number of propositions in any protocol only 12-15% of them were inference-based.

Recall scores were separated according to question type to see if prompted recall (i.e., follow-up question differed from non-prompted recall (i.e., the general question). The levels of recall for the breast cancer text were low, but participant five reading the low readability text benefited from the prompts, recalling four propositions after a zero correct recall response to the general question. Participants three and four reading the colorectal text at both readability levels responded in a similar

manner. The opposite was true for participant six reading the high readability breast cancer text, as the non-prompted question elicited four recalls while the follow-up questions failed to detect any more.

Looking at female propositional recall by individual, one notes that the recall of participants 2 and 4 was more than double for the low readability page than the other participants. Participant 2, in particular, stands out for possessing substantially greater recall for both colorectal web pages. She responded well to both the non-prompted and prompted questions, though her recall for the high readability page favoured the general question. This participant, a 64-year-old female retiree, reported having done previous cancer research, specifically on Hodgkin's lymphoma when her daughter developed the disease at age 17. The participant also shared that she and her husband had been learning about warning signs, tests, and new treatments of colon cancer since her mother-in-law was diagnosed with the disease. This previous exposure to colon cancer information likely facilitated her recall of the web pages. Indeed, looking at the content for the low readability web page, much of the propositions recalled dealt with anatomical references, possible signs and symptoms of colon cancer, and factors affecting prognosis. Furthermore, she not only described several symptoms, but also went on to clarify that some symptoms do not necessarily mean colon cancer and that some are better than others. This example of critical thinking and superior recall is likely related to her previous research on colon cancer and her vested interest in learning about preventative measures. To that effect, she commented: "I liked how there was a section on how to prevent…[it] had changes like change of bowel, diarrhea, stool content, colour. I thought it was good to also know stages because I think that's important if you're diagnosed". For the low readability page, much of the recall was about the drugs used to lower colorectal risk. She specifically recalled that "nonsteroidal drugs like Motrin or Advil [can have] a 20 to 50% lower risk of colorectal cancer", which no other participant verbalized with such accuracy. She went on to explain possible side effects of these drugs and how they may also help

improve other medical conditions. Although her recall was better for the low than the high readability page, a greater percentage of her protocol for the more difficulty written page consisted of correct recall (42 vs. 29%). Her recall is even more impressive when one considers that the majority of it was in response to the general question, without any prompts. Her protocol about the high readability page also contained a greater proportion of inferences (12 vs. 9%).

It is also informative to examine the relationship between protocol content and text structure. Figures 14 and 15 present protocol content plotted by text sequence of participant 2. This is based on the assumption that to successfully recall a text, propositions from each paragraph must be recalled. The content includes both correct recall and propositions that contain concepts that match the topic content of a certain paragraph. The numbers beside each data point signify the order in which propositions from a particular paragraph were recalled. For example, in the low readability page, participant 2 recalled five propositions from paragraph 3, five from paragraph 8, three from paragraph 12, then lastly six more from paragraph 8. It should be noted that for the low readability page the first three data points for both the low and high readability pages are composed of propositions recalled in response to the general question. The remaining data points capture propositions from the follow-up questions. The backwards movement through the paragraphs (e.g., moving from paragraph 12 back to paragraph 4) signifies that a follow-up question asked about information that the participant already touched upon in her general question response. Note that for the high readability page, participant 2 was only able to generate propositions relating to the first half of the text, which discussed diet and exercise, vitamins and calcium, as well as nonsteroidal anti-inflammatory drugs. The remaining paragraphs were about female hormones and hormone replacement theory, a topic that the participant failed to touch upon.

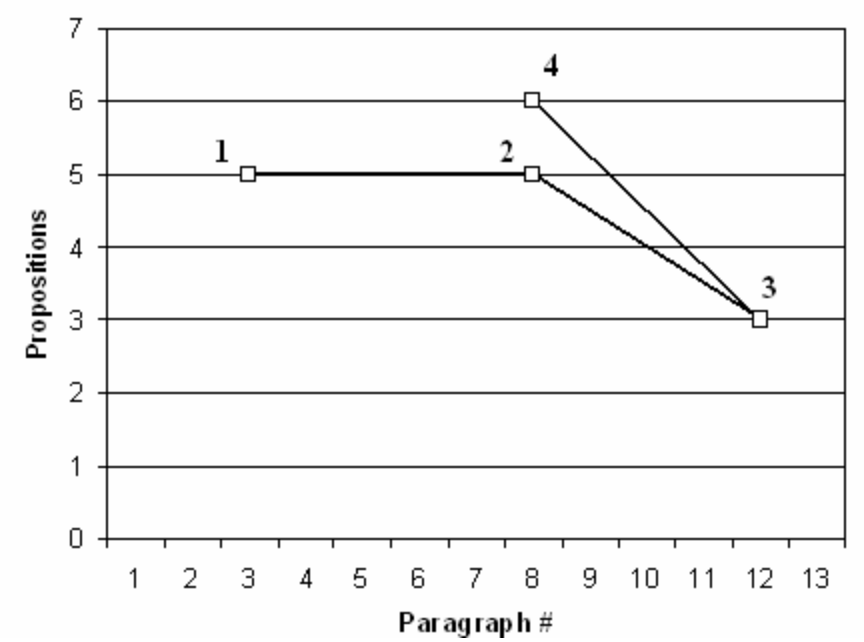


Figure 14. Paragraph sequence of participant 2's protocol for the low readability colorectal cancer web page.

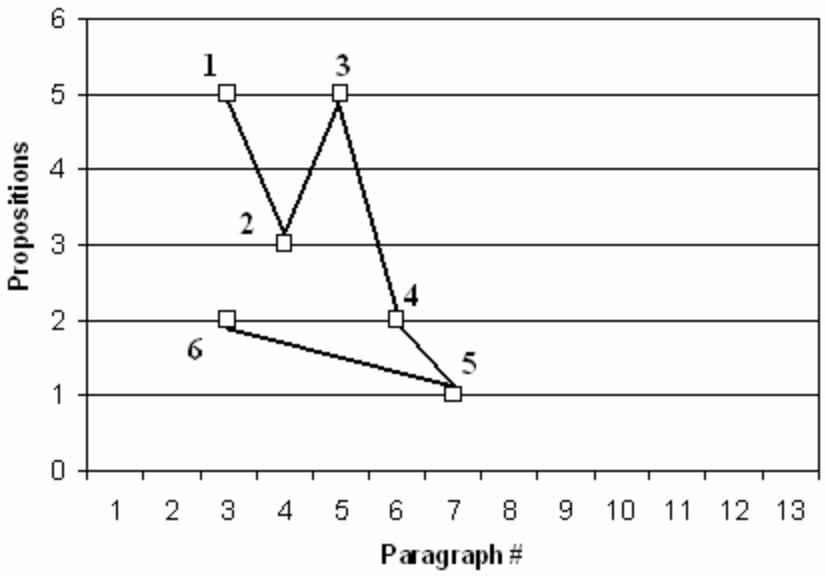


Figure 15. Paragraph sequence of participant 2's protocol for the high readability colorectal cancer web page.

Participant 12's recall shows a slight different pattern in his paragraph timeline, as shown in Figure 16 and 17. For the low readability web page, he began by recalling a proposition from the first paragraph and another proposition from the seventh, then back to the first paragraph for two more propositions, ending with a final one from the tenth paragraph. These first four data points were in response to the general question. This section of the participant's protocol touched upon superordinate points of the text, which is the reason for the paragraph jumping. He explained: "[This article] is a perfect catalogue of points that you need to consider, about symptoms, possible treatments, procedures, and risk factors". The remaining data points, numbers five and six, dealt with the follow-up questions. Similarly, the zigzag pattern of his recall for the high readability page was a product of general response containing general propositions about the main points of the text.

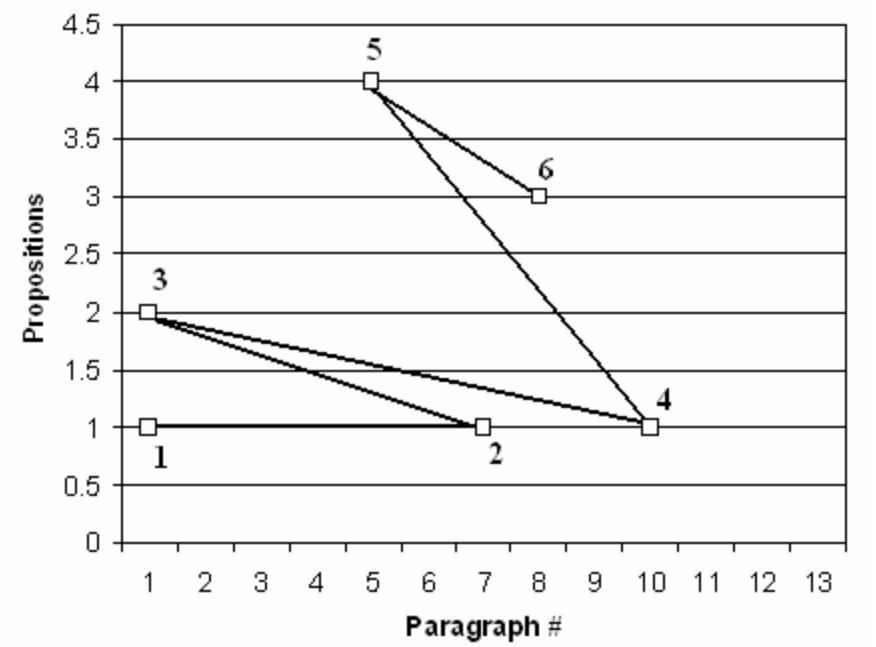


Figure 16. Paragraph sequence of participant 12's protocol for the low readability colorectal cancer web page.

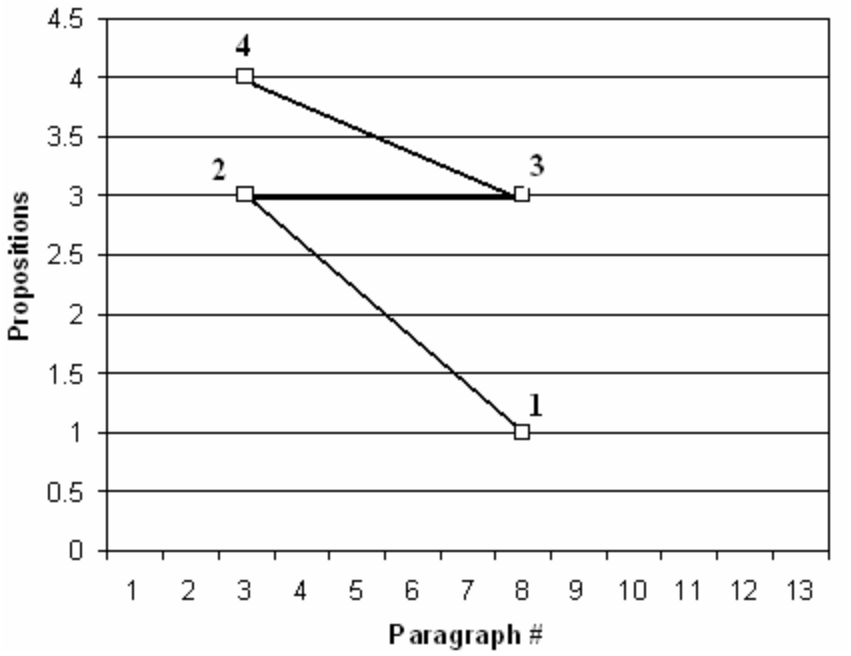


Figure 17. Paragraph sequence of participant 12's protocol for the high readability colorectal cancer web page.

A third example of the relationship between paragraph sequence and protocol recall is shown in Figures 18 and 19. Participant 5, a 70 year old female, was asked questions about the breast cancer. For the low readability page, her response to the general question was irrelevant to the text content, so the data points in Figure 18 represent responses to the follow-up questions. The protocol's range captured propositions from most of the web page, however it was not done in an ordinal manner. The answer of one question about the early signs of breast cancer was contained within the first three paragraphs of the text. The participant, however, included propositions from paragraphs seven, nine, and ten, which spoke more about the method in which many of the early signs are detected, namely breast self-examination. For the high readability page, the first five data points represent the general question response. She spoke mainly about the merits of the mammogram and ultrasound as screening methods which related to the paragraph structure in an orderly manner up to the forth data point.

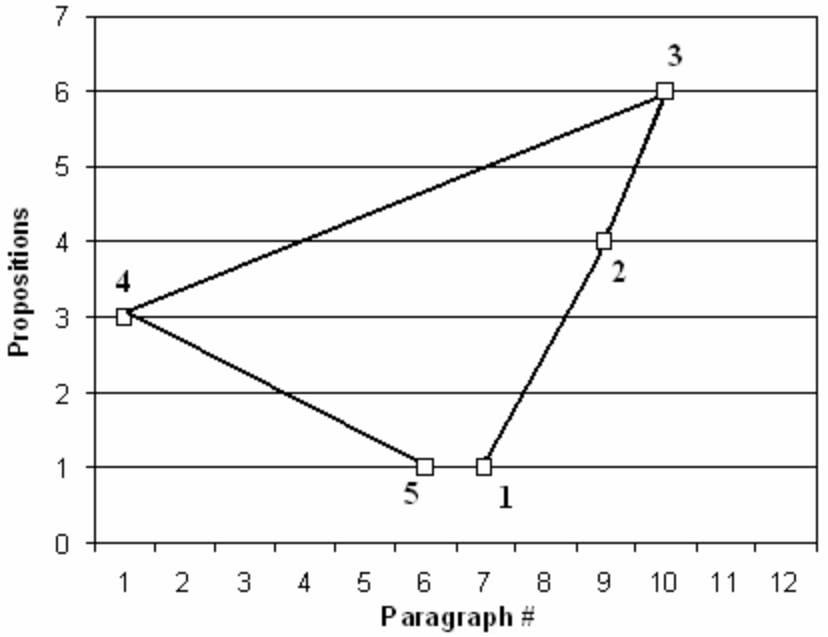


Figure 18. Paragraph sequence of participant 5's protocol for the low readability breast cancer web page.

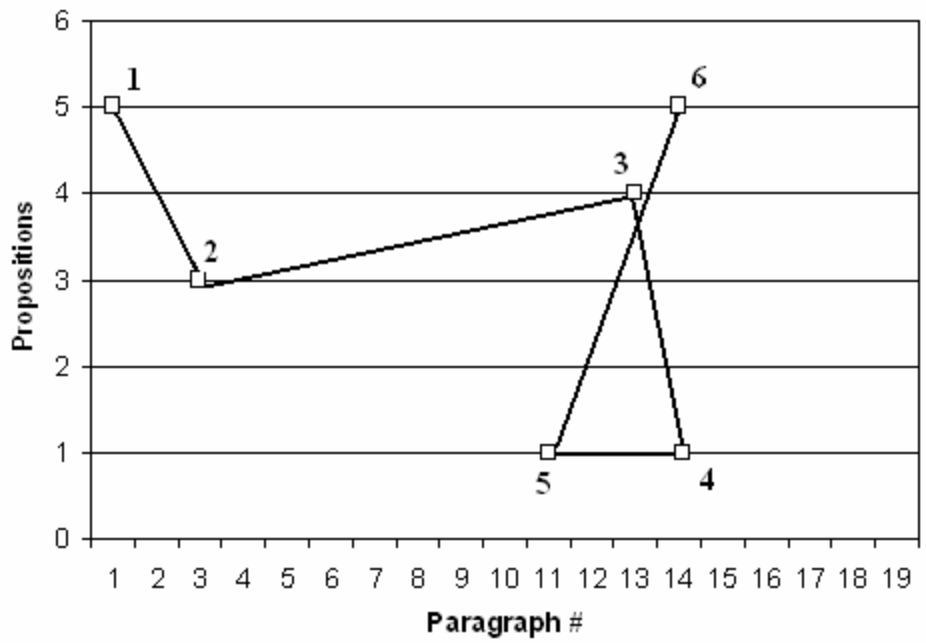


Figure 19. Paragraph sequence of participant 5's protocol for the high readability breast cancer web page.

Recall by male participants for the colorectal and breast cancer web pages is given in Table 6. One marked difference between the male and female recall is that male overall recall was much less than female recall. Unlike the females, the number of propositions recalled by males for colorectal cancer is about the same as for prostate cancer. Also, readability level seemed to have a different pattern of influence on male recall, with a greater number of correct propositions recalled for the more difficult colorectal cancer web page and no difference for the prostate cancer information. It should be noted that the high readability colorectal web page had more familiar content (e.g., diet and vitamins) than the low readability colorectal web page (e.g., risk factors, tests and procedures). This difference in familiarity may account for the increased recall for the more difficulty written colorectal cancer page.

Overall, correct recall accounted for a greater proportion of propositions in male colorectal prostate cancer protocols. The high readability colorectal cancer page, in particular, elicited the greatest proportion at 24%. This page also produced the greatest proportion of inferences in the protocols, though only by a small margin.

Table 6. Male Propositional Correct Recall of Low and High Readability Colorectal and Prostate Cancer Information by Prompted (P) and Non-prompted (NP) Questions

	Colorectal Cancer					Prostate Cancer				
	Low		High			Low		High		
Participant	NP	P	NP	P	Participant	NP	P	NP	P	
9	0	0	3	3	13	0	0	0	0	
10	0	0	0	6	14	4	3	3	1	
11	0	4	0	3	15	0	0	0	2	
12	0	4	3	2	16	3	4	0	5	
Total	*0*	*8*	*6*	*14*	*Total*	*7*	*7*	*3*	*8*	
% Overall	*0.08*		*0.24*		*% Overall*	*0.07*		*0.13*		
% Inference	*0.12*		*0.17*		*% Inference*	*0.08*		*0.14*		

Note. % Overall and % Inference denote the proportion of correctly recalled propositions and the proportion of inferences out of the total number of propositions in a protocol respectively.

Looking at prompted versus non-prompted recall, one notes that participants 10, 11, and16 relied heavily on prompts for certain web pages. Participant 10, for example, talking about the high readability colorectal page, generated six recalls to the follow-up questions after providing no correct propositions to the open ended question. Participant 16, a 65 yr old male, relied on prompts for the high readability prostate cancer page, but was able to provide a similar level of recall when responding to the low readability page for both types of questions. He was asked to talk about the high readability text first, so it is possible that the participant's improved recall to the general question for the low readability page was because he expected that specific questions would follow.

Similar to participant 2, participant 11 also reported having done previous cancer research. The 72-year-old male shared that he had endured prostate cancer and had consulted books, newspaper articles, magazines, and the Internet for health information. He also shared that he was most

interested in learning about diagnostics and screening procedures. Although his recall scores were not the highest (4 and 3 propositions for the low and high pages respectively), what he did recall correctly were the more general or superordinate propositions. For example, in the low readability page, he described how lifestyle and dietary considerations can influence the development of colon cancer. This is a valid statement, however none of those specific propositions were contained in the low readability web page. The text did contain a subheading of "diet and exercise", which listed what foods should be limited and those that are important to eat. The "exercise" portion of that section encapsulates what the participant meant by lifestyle considerations. Generally speaking, the amount of physical activity in which a person partakes could be viewed as a lifestyle choice. When the participant was asked specific question regarding risk factors, he confirmed his previous response: "One would be family history of colorectal cancer and the other would be dietary considerations and probably a third would be lifestyle in terms of exercising and keeping healthy". It is difficult to determine whether his comprehension was incomplete, or whether if prompted further, he would have gone into more detail. The specific questions could have served as prompts, yet his responses to these remained at a general level. Given that marking for correct recall was strict, no credit could be given for these types of propositions. Still, 14% of the low readability protocol was correct; recall for the protocol of the more difficult text was even higher at 16%. Inferences were also more frequent in the high than in the low readability protocols (21 vs. 9%).

Table 7 contains the propositional analysis and protocol scoring of the males reading the high readability prostate cancer web page. The successfully recalled propositions revolved around the prostate-specific antigen (PSA) level, and the ages in which men of different risk groups should take the PSA test. Participant 14 accurately recalled the ages in response to the general question, while participants 15 and 16 had to be specifically asked about the topic; participant 13's protocol failed to

match any web page propositions. Table 7 also serves to illustrate that relative to amount of information contained in the any of the web pages, the proportion of correct recall was extremely low.

Table 7. Commonly Recalled Propositions in Protocols about the High Readability Prostate Cancer Web Page

			Participant #			
ID		Proposition	13	14	15	16
P5	S1	P1 (OF PSA LEVEL)			1	
		P2 (MOD LEVEL HIGH)				1
		P3 (IN PSA BLOODSTREAM)			1	
		P4 (IS P1 WARNING-SIGN)				
		P5 (IS WARNING-SIGN P6)				
		P6 (PRESENT $ PROSTATE-CANCER)				
		P7 (MOD PRESENT POSSIBLE)				
P5	S2	P1 (OF KIND PROSTATE-DISEASE)				
		P2 (MOD KIND OTHER)				
		P3 (CAUSE P1 PSA-LEVEL)				
		P4 (MOD PSA-LEVEL HIGH)				
		P5 (SINCE P6 P1)				
		P6 (CONFIRM PSA-TEST PROSTATE-DISEASE)				
		P7 (OF PRESENCE PROSTATE-DISEASE)				
		P8 (MOD PSA-TEST ALONE)				
P5	S3	P1 (INDICATE PSA-LEVEL PROSTATE-CANCER)				
		P2 (MOD INDICATE ONLY)				
		P3 (OF POSSIBILITY PROSTATE-CANCER)				
		P4 (INDICATE PSA-LEVEL P5)				
		P5 (FOR NEED EVALUATION)				
		P6 (BY EVALUATION PHYSICIAN)				
		P7 (MOD EVALUATION ADDITIONAL)				
P5	S4	P1 (MEAN PSA-LEVEL P4)				
		P2 (MOD PSA-LEVEL LOW)				
		P3 (NEGATE P1)				
		P4 (IS PROSTATE-CANCER PRESENT)				
		P5 (NEGATE P4)				
		P6 (CONVERSELY P1)				
P6	S1	P1 (MOD AMERICAN-CANCER-SOCIETY ACCORDING-TO)				
		P2 (AGE MEN 50-AND-OLDER)		1		1
		P3 (AGE MEN AGE-45)		1		1
		P4 (IN AGE-45 HIGH-RISK-GROUP)		1		
		P5 (EXAMPLE-OF HIGH-RISK-GROUP AFRICAN-AMERICAN)				
		P6 (EXAMPLE-OF HIGH-RISK-GROUP P7)				
		P7 (WITH MEN FAMILY-HISTORY)				1
		P8 (OF FAMILY-HISTORY PROSTATE-CANCER)				
		P9 (AND P10 P11)				

		P10 (POSSESS MEN PSA-BLOOD-TEST)					1
		P11 (POSSESS MEN DIGITAL-RECTAL-EXAM)					
		P12 (DURATION-OF PSA-BLOOD-TEST ONCE-EVERY-YEAR)					1
		P13 (DURATION-OF DIGITAL-RECTAL-EXAM ONCE-EVERY-YEAR)					
			TOTAL:	0	4	2	5

5.2.1.2.2 Discussion

The second question asked if there would be greater correct recall for gender specific than for gender neutral cancer information. For females, the results showed the reverse in that there was greater recall for the colorectal than for the breast cancer web page. This was mostly due, however, to the recall of participants 2 and 4. When one compares the total score, males had comparable recall for the colorectal and breast cancer material. When one looks at recall as a proportion of overall propositions in a protocol, the results again show greater recall of the gender neutral information. One explanation is that when prior knowledge is low, there is a greater dependence on using the same words in a text because no schema exists. When prior knowledge about a particular subject is more established, however, a person is more likely to use synonyms (Pask, 1988). Another explanation for the non-preferential recall towards the gender specific information is that it was an assumption that gender would be a proxy for prior knowledge. While it intuitively makes sense that females would have a more established existing schema for breast cancer relative to colorectal cancer information and similarly for males regarding prostate cancer, this type of prior knowledge was never explicitly measured.

Although an individual's schemas (established through prior knowledge) is generally regarded as important in reading comprehension, it is not known what elicits the retrieval of the appropriate schemata during text processing and how it might be measured. Understanding occurs from the interaction of the new information presented by a text and an individual's schemas (Fincher-Kiefer, Post, Greene, & Voss, 1988). Both content and textual schemas provide information that aids

in the interpretation of meaning. That is, readers attempt to match content and textual schemas with text information. As the reader does this, he or she builds a mental concept for the meaning of the text. This concept is constructed partially out of information previously known and partially by the new information presented in the text. The processes of building and refining mental concepts of meaning allow comprehension to occur (Armbruster, 1986). As such, meaning is not a property of the text or of the individual, but arises out of the interaction of the two.

In contrast, readability is a product of the properties of a text. One would expect that if simpler language translated to correct recall that there would be a clear advantage for the low readability web pages. With the exception of the low readability page for colorectal cancer read by females, the results show a marginal difference due to readability. In fact, the low readability page for colorectal cancer read by males revealed an opposite pattern, with greater recall for the more difficult page. One must also note that recall alone does not imply understanding. Comprehension and learning depend on processing at the textbase and situation level (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; McNamara et al., 1996), whereas readability formulas are based on surface characteristics of the text. Measuring text elements that are primarily needed for surface processing does not adequately capture comprehension and learning, which is a concern for online health information providers. Although text characteristics can certainly predict aspects of readability, readability should be viewed as an interaction between a text and a reader's cognitive aptitudes (Kintsch, 1994; McNamara et al., 1996; Miller & Kintsch, 1980).

The protocols of participants 2 and 11 showed two different approaches to recall. The first participant was one who had done prior research on her cancer type, had a vested interest in learning about diagnostics, and recalled in detail propositions that included drug names and anatomical parts. Her recall scores of 23 and 18 for the low and high readability pages were much greater than the next highest scores by participant 4. The second participant, while having done prior research on cancer

information online, did not report having previously researched his cancer type. His protocols contained much fewer propositions overall than the first participant, and his recall consisted mainly of superordinate propositions. Though his recall scores were lower than hers, because his protocols contained fewer propositions overall, the proportion of recall propositions was slightly greater for him. Although recall is a popular indictor of comprehension, the information conveyed by a recalled proposition is also an important consideration. Clearly it is more substantive if a reader can remember what drugs can reduce risk and by how much than simply reciting that lifestyle and diet are important factors.

Another consideration is that text representations are built up sequentially. It is not possible to construct and integrate a text representation from an entire web page. The web page has to be processed word by word and sentence by sentence. As each text segment is processed, it is immediately integrated with the rest of the text that is currently being held in working memory. The construction-integration model (Kintsch, 1988) combines a construction process in which a textbase is constructed from textual input as well as from the comprehender's knowledge base, with an integration phase, in which this textbase is integrated into a coherent whole. Knowledge is represented as an associative net, the nodes of which are concepts or propositions. Comprehension is assumed to be organized in cycles and with each new cycle, a new net is constructed, including whatever is carried over in working memory from the previous cycle.

In this study, the protocols were composed of responses to two types of questions. The general question gave the participants the opportunity to talk about the web page without any constraints. It is reasonable to suppose then that the order in which propositions were generated in response to this question might be related to the order in which they were processed. For example, before answering the general question, participants had to filter through their respective associative nets and decide which propositions were relevant. The structure of this net, according to the

construction-integration theory, is based on the cycles in which the information was initially integrated. Participant 2's recall followed the paragraph structure of both colorectal cancer web pages. For the low readability page, her recall was consistently spaced out in terms of structure. She was able to recall propositions from the beginning, middle, and end of the web page. The structure of her high readability recall was narrower, capturing only the first half of the text. This participant serves as one illustration of how text structure and recall can be related. Participant 5 and 12's recall showed different relationship with text structure, demonstrating that the cycling of information can vary from person to person.

5.2.1.3 Inferences

5.2.1.3.1 Results

Inferences were extracted from verbal protocols by conducting a side-by-side comparison with the propositional analysis of the appropriate web page. By doing so, any propositions that were modifications of web page propositions could be highlighted. Inferences were separated in the following categories: (1) general to specific; (2) specific to general; and (3) same level. Table 8 contains examples of these three categories of inferences. General to specific inferences involved drawing a lower level conclusion from a higher level premise. For instance, from the premise "history of breast cancer", one can specify having cancer in the family is one example of having a history. Specific to general inferences involved drawing a higher level conclusion from a lower level premise. For instance, one of the breast cancer web pages explained that "three-quarters of the cancer that are diagnosed in BRCA1 carriers are estrogen-receptor negative". One participant inferred that statement to mean that most of the cancers are "generally negative". This participant could not remember the exact amount that was stated in the web page, but knew that it was high amount, and concluded it to be "generally negative". Lastly, same level inferences operate on the same level in that there is not

generalization or specification. For instance, "vomiting" is at the same level as "tendency to throw up" in that both concepts convey the same act. For the low readability web pages, females used general to specific inferences the most, followed by same level and then by specific to general ones. There was no difference in their inference types for the high readability pages. For males, same-level were preferred by males reading the low readability pages; for the more difficult pages, more deductive inferencing was employed.

Table 8. Examples of the General to Specific (G-S), Specific to General (S-G) and Same Level (SL) Inferences Produced in the Verbal Protocols

Category	Web Page	Verbal Protocol
General to Specific (G-S)	History of breast cancer	If you have cancer in the family…
	Breasts will not exactly match	Uneven breasts
	Avoid having red meat	Cut your beef back
Specific to General (S-G)	Three-quarters negative	Generally negative
	Whole grain foods	Grains
	Fecal occult blood test	Blood test
Same Level (SL)	Tendency to throw up	Vomiting
	Narrower stool	Skinnier stool
	High levels of dietary fat	Preponderance of fat

Relative to the amount of information in the web pages (i.e., number of propositions), the number of inferences made was small. The number of inferences did not noticeably differ across readability level or cancer type. Figures 20, 21, 22, and 23 present the number of inferences made by participants for the breast, prostate and colorectal cancer pages at low and high readability levels. Although it appears that female protocols contained more inferences than males, the difference becomes minute when taken as a proportion of the number of propositions in the web pages. For

example, the inferences only accounted for about 1% of all protocol propositions. This percentage is higher when looking at the protocols on an individual basis. Collapsing over readability, one remarks that cancer type had no effect on male inference production (26 vs. 25 for prostate and colorectal cancer respectively). For females, however, there were 44 colorectal versus only 32 breast cancer inferences. These results do not provide evidence for question 2 which predicted that there would be greater inferences for the gender consistent than gender neutral cancer information.

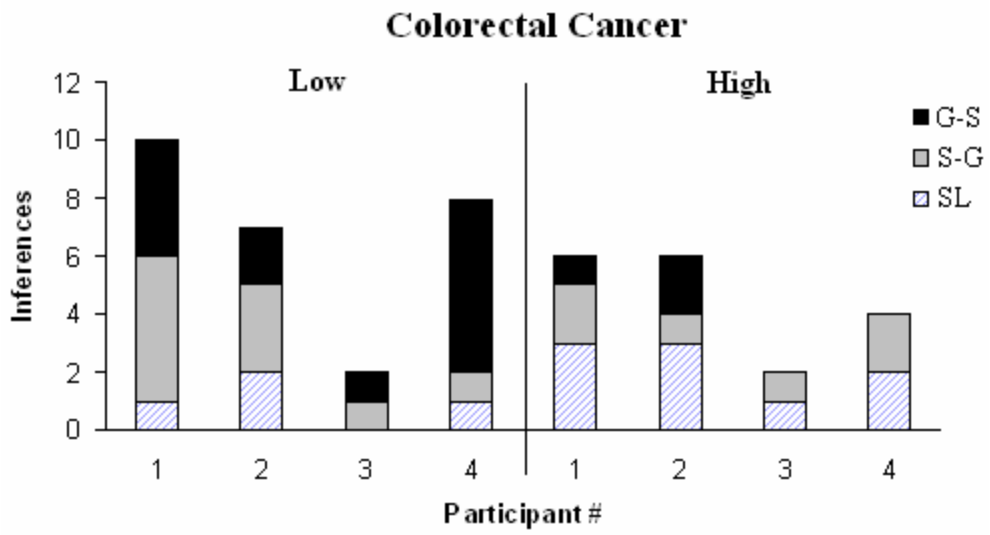


Figure 20. General-specific (G-S), specific-general (S-G) and same level (SL) inferences generated by females for low and high readability colorectal cancer web pages.

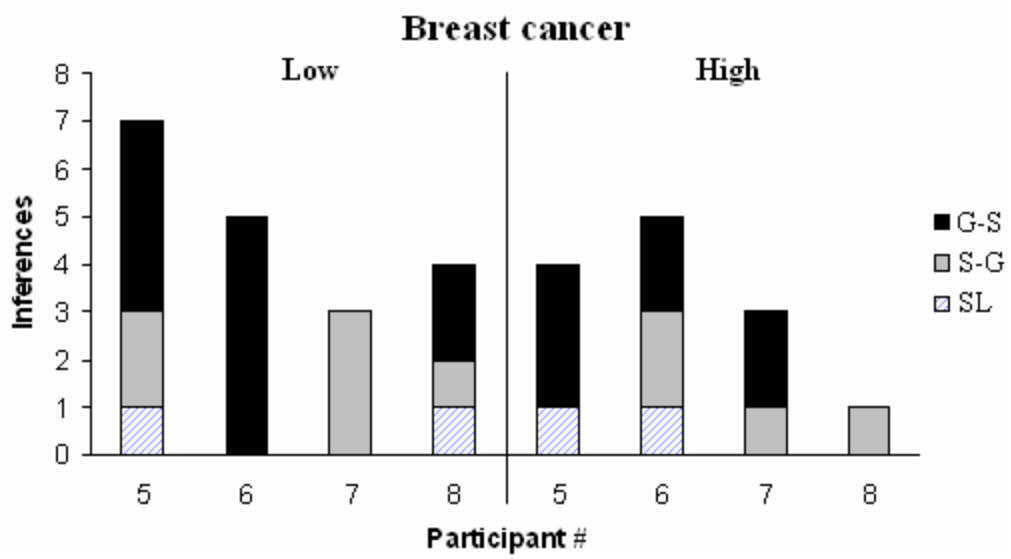


Figure 21. General-specific (G-S), specific-general (S-G) and same level (SL) inferences generated by females for low and high readability breast cancer web pages.

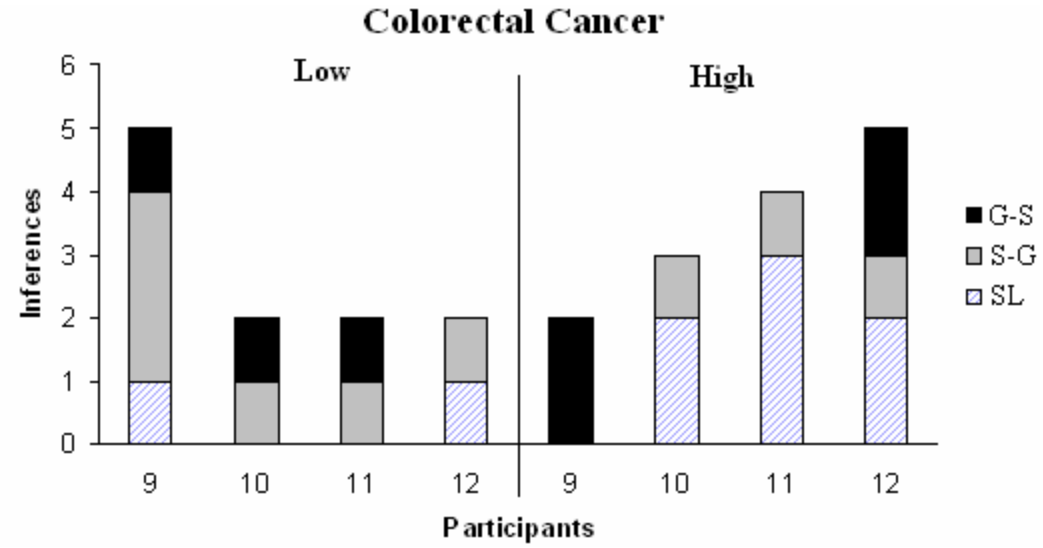


Figure 22. General-specific (G-S), specific-general (S-G) and same level (SL) inferences generated by males for low and high readability colorectal cancer web pages.

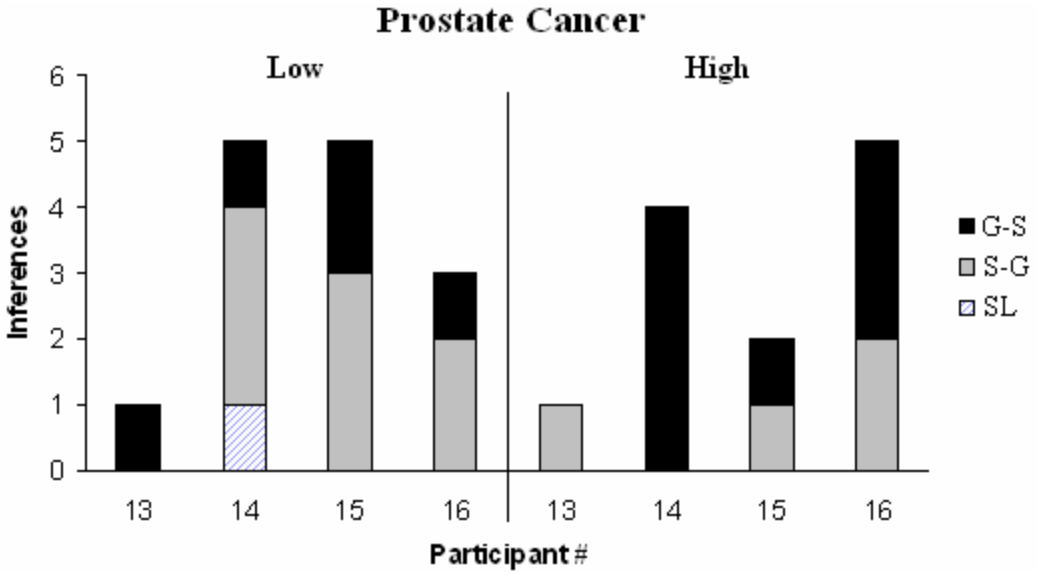


Figure 23. General-specific (G-S), specific-general (S-G) and same level (SL) inferences generated by males for low and high readability prostate cancer web pages.

Participant 6, a 65-year-old female, reported having a profound interest in breast and uterine cancer. She said that she was most interested learning about the symptoms, diagnosis, and treatment of those cancers. She also reported that she relied on the Internet as her primary source of health information, but also used reference books from the library. Recall for this participant was better for the high readability page than for the low readability page (4 vs. 1). As shown in Figure 20, she generated the same number of inferences for both levels of readability at five each. The compositional types, however, differed across readability. For low readability, all five inferences were general to specific; three of them were about symptoms involving the breast. The text in the web page stated "Look for any changes in contour of each breast, a swelling, a dimpling of skin, or changes in the nipple." The participant correctly recalled the symptom swelling, but contextualized it as "swelling from the nipple area". On the surface, this difference may appear to be minor, but on a semantic level, the difference is significant and represents an inference. The participant inferred that a

swelling on the breast meant that the swelling would involve the nipple area, which logically is a part of the breast. She also mentioned two other symptoms, lumps and discharge from the nipple, after reading the text “Squeeze nipple; check for discharge and lumps”. Although the text does not explicitly state that the discharge will come from the nipple, it is logical to infer that if one is checking for discharge after squeezing the nipple, that the source of the discharge would be the nipple. For the high readability page, the participant produced one specific to general inference, and two each of the same level and general to specific types. A section of the webpage discussed BRCA1 and BRCA2 mutations and how the effectiveness of certain drugs differed in women possessing the different mutations. One of the same level inferences was about whether the drug tamoxifen was more “favourable” in BRCA1 than in BRCA2 carriers. The web page used the term beneficial instead of favourable, yet the underlying meaning was similar in that the participant knew that tamoxifen was more effective in BRCA2 carriers. The other same level inference was about ultrasound and how the authors of the web page were against recommending using it because compared to the screening arsenal of the MRI and mammograms, “[the ultrasound] does not really add more”. The participant inferred this to mean that “ultrasounds were not as effective” in the sense that the currently recommended screening arsenal sufficiently detected abnormalities, and beyond that the benefits of the ultrasound are insignificant. Many of the participant’s inferences for both pages were related to her recall; that is, some of the propositions of the inferences matched propositions from the web page.

Compared to the female protocols which contained a good variety of the three inference types, as shown by Figure 23, there was only one same-level inference made by males reading the prostate cancer material. This was done by participant 14, a 56-year-old male, who reported having done research on prostate cancer in the past, but relied on brochures and pamphlets to do so. He correctly recalled seven and four propositions in the low and high readability pages respectively, which is impressive considering that both protocols only consisted of 22 propositions each and were

in response to the general question. His high readability protocol included four general to specific inferences, two of which were about symptoms and the other two dealt with the prostate-specific antigen (PSA) test. For example, the text described that prior to the testing process, physicians may ask about symptoms being experienced, in particular "problems with urination". The participant inferred this to mean "frequent or changes in urination", which although does not automatically translate into being a problem, could be viewed as deviating from normal, and thus be a potential concern. For the low readability page, his protocol contained one each of the specific to general and general to specific inference types, and three same level ones. One analogy was from the text that read "avoid having red meat more than 2-3 times a week." The participant described this as "cut your beef back to 2 to 3 times a week". The end result of the two messages is the same, although the web page seems to take a preventative position, whereas the participant's same level inference is more proactive. This example is also a good illustration of both an inference and correct recall (i.e., proposition that the amount of beef is 2-3 times a week). It is also interesting to note that of the seven points made in the overall message section of the web page, three of them were related to inferences made about fatty foods, red meat consumption, and daily recommendation portions of fruits and vegetables. This is notable because inferences are constructive processes that can influence global understanding. The fact that about half of the web page's overall message was related to the participant's inferences suggests that he derived and internalized a good portion of the text's main points.

5.2.1.3.2 Discussion

Kintsch's situation model is the mental representation of the people, setting, action, and events that are explicitly mentioned or inferentially suggested by a text. Most inferences generated during text comprehension are part of the constructed situation model. If a text is entirely explicit, in that every detail as well as the overall structure is made perfectly explicit, the textbase also is a good

situation model. This means that no further knowledge elaborations on the part of the comprehender are required.

At a general level, inference is a cognitive process used to construct meaning. Inference in reading comprehension is a constructive thinking process because the reader expands knowledge by proposing and evaluating competing hypotheses about the meaning of the text in an attempt to progressively refine understanding. If readers are unable to generate inferences that connect explicit information in a text to relevant world knowledge, they feel as though they do not comprehend the text and have difficulty remembering it. As such, increased inference complements comprehension and should aid in recall. A comparison of recall and inferences collapsed across readability is shown in Figure 24. With the exception of participant 15, the levels of recall and inference are very similar for participants 7-16. The inference levels here do not result in superior levels of recall, although that may be more a function of the type of recall task and the nature of probes used in this study. The recall score of participant 2 stands out as being almost twice as great the next highest score, while the inference generation level is comparable to those of other participants. Recall that this is the participant that had done prior research on her cancer type and had a close relative afflicted with the disease. Perhaps these factors contributed her recall and thus fewer inferences were needed for successful comprehension.

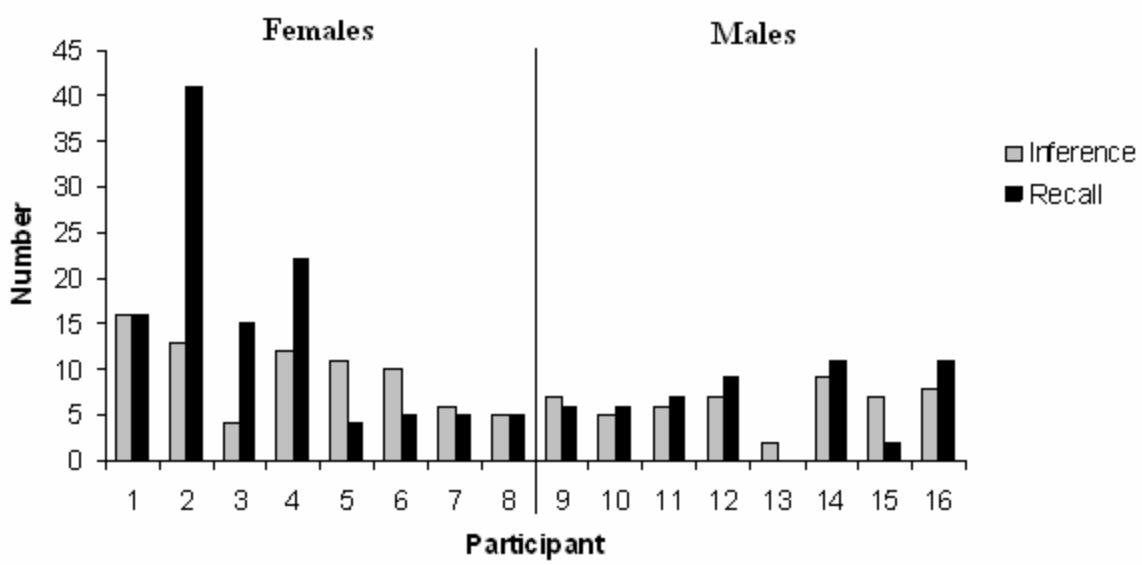


Figure 24. Relationship between total inference generation and total recall for males and females.

Although inference must work together with the textbase to provide complete picture of comprehension, there is a fine balance between inference, coherence, and recall. An excess of misinformed inferences can introduce variability in interpretation and possibly decrease the proportion of correct recall. Also, the lack of a clear influence of cancer type on male and female inference generation is not surprising given that gender was only an assumed proxy for prior knowledge.

5.2.1.4 Coherence

5.2.1.4.1 Results

In order to comprehend a text, a reader must create a well connected representation of the information in it. This connected representation is based on linking related pieces of textual information that occur throughout the text. The linking of information is a process of determining and maintaining coherence. One way of doing this is to examine propositions that capture information within paragraphs (i.e., embedding propositions) and between paragraphs (i.e., linking propositions).

In contrast to readability formulas which focus only on the word level, this method looks at the different levels of reading required for a coherent representation. This method also allows us to see what kind of information is captured by embedding and linking propositions. For example, do linking propositions represent general comments across paragraphs?

The different types of information linkages in the participant protocols are shown in Figures 25 and 26. It should be noted that comparisons across cancer type and gender are not the focus, but rather the individual compositions. The figures show that each protocol was overwhelmingly composed by propositions that contained information within paragraphs than across paragraphs. Although each web page contained different numbers of paragraphs, this did not appear to influence the proportion of linking to embedding propositions, as the number of linking propositions remained relatively similar across web pages. The large number of embedding propositions is mostly likely due to the questions posed to the participants. With the exception of the general question in which the participant's protocol was unconstrained, the follow-up questions targeted specific answers contained within one or two paragraphs. For example, participants asked to understand the low readability prostate cancer page were posed the question: "Why are people in eastern countries less likely than westerners to develop cancer?" The answer to this question, genetics and diet, was explained in two consecutive paragraphs, but unless a participant linked these two factors together, propositions were limited to the embedding type.

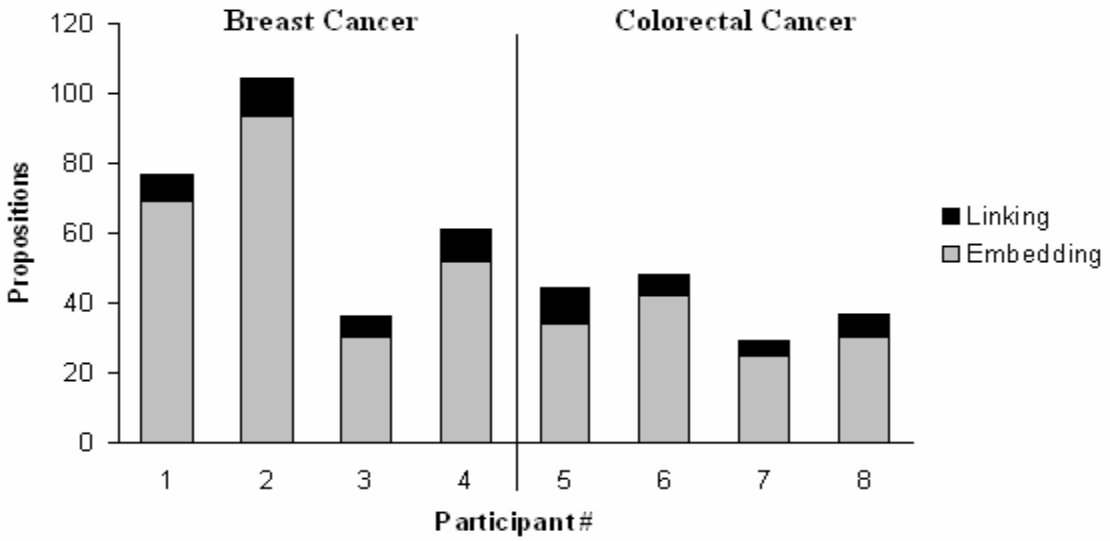


Figure 25. The number of linking and embedding propositions in female protocols for breast and colorectal cancer web pages.

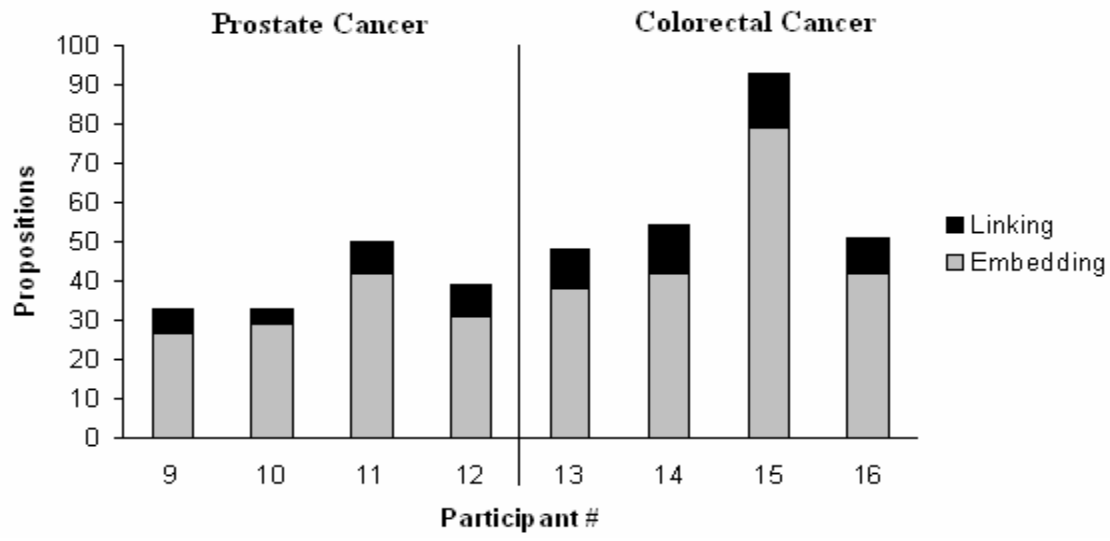


Figure 26. The number of linking and embedding propositions in male protocols for prostate and colorectal cancer web pages.

Recall participant 14, whose inference processes were described earlier. His protocols contained 42 embedding and 12 linking propositions. His response to the general question of the high

readability page was very broad: “Diagnosis and treatment and symptoms…of prostate cancer, I suppose”. On the surface, this response appears to be rudimentary in terms of comprehension. Yet his response in fact linked all the sections of the web page, which began by describing questions a physician may ask prior to the testing process, then went on to explain in detail the various tests that are used to diagnose prostate cancer, such as the digital rectal exam, prostate-specific antigen test (PSA), transrectal ultrasound, and biopsy. His other linking propositions were conditional in format; that is, if condition A exists, then the result is B. For example, he spoke about symptoms such as changes in urination and swelling, and then explained that those symptoms, along with age, can put a man in the high risk group. The participant concluded his thought by stating that this combination of factors indicates that a man should take one of the tests described in the web page. Another example was when he explained that if a man has a high PSA level, then it might mean cancer. Here, this linking proposition represents reading moving from a general to a more specific level, just as the web page moved from a section describing the PSA test, then to a subsection explaining the causes of a high PSA level.

Many other participants made similar linking propositions in their protocols. Table 9 contains excerpts which illustrate how participants linked the content of various paragraphs in a manner that summarized the main points of the text. Linking propositions are important indicators of comprehension because they contribute to global coherence. The fact that many of the linking propositions in the protocols represent general comments suggests that the content was actively processed and then filtered to produce a high-level summary of the main points of the text. This, in combination with embedding propositions, which were generally elicited by the follow-up questions in this study, can be thought of as the necessary contributors to coherence, and thereby overall comprehension.

Table 9. Examples of Linking Information that Represents General Comments across Paragraphs

Participant #	Excerpt
3	"…risk factors and how it can affect your life, the biopsy… [the article] explained certain factors, the prognosis, the chances of recovery are in here".
4	"[the article] gives you the symptoms of cancer, the colorectal cancer. It gives you the specifics about the type of people who are going to get the disease and their age and their previous physical history or medical history…"
6	"Basically [the article] just outlines approximately 10 different problems or symptoms that you might observe that would warrant going to your doctor for a further check".
8	"The early detection, [the article] listed some points of some of the symptoms a person might have or be able to look for, but spent most of the time on the exam, the self exam, and checking it out".
9	"[the article] gives a good overview as to what disease is…it describes some of the risk factors [and] possible signs you should look for…"
12	"[the article was] a perfect catalogue of points… about symptoms, about possible treatments, about procedures, about risk factors".
15	"Basically [the article] is a primer thing. It talks about… the tests… there are simple symptoms, you know, initial symptoms, the tests for it, the fact that the tests themselves are not 100% accurate, the treatments can vary".

5.2.1.4.2 Discussion

Regarding question 2 which asked if individuals who are familiar with cancer information (i.e., females reading breast cancer text, males reading prostate cancer text), relative to those who are unfamiliar (i.e., males and females reading colorectal text), would produce more coherent protocols, the results are inconclusive. As pointed out in the discussion section of the participant's recall, using gender as a proxy for prior knowledge and thereby increased familiarity, was an assumption, and one that was never measured. As such, the parameters of this study had no valid claim to suppose that a certain group of participants would produce more or less coherent protocols.

What the results did allow, however, was the in depth investigation into the types of information linkages made by each participant and what kind of information was conveyed by those linkages. This revealed that embedding propositions were dominant over linking propositions in all protocols, though this was likely a function of the follow-up questions since their answers could be found within a single paragraph. If participants' representations were entirely dependent on the information within paragraphs and were unable to relate the paragraphs themselves, then comprehension would be limited to the few propositions maintained in active memory. However, the information contained within the linking propositions show that participants were able to connect information at the paragraph level to information that was not currently activated.

Kintsch and van Dijk's theory of text comprehension (1978) emphasizes text-based sources of coherence. The term text-based refers to devices used by readers to communicate relations between text elements. These include the use of connectives (e.g., and, because), signals (e.g., headings), and other devices. If readers do no respond appropriately to these indicators of relations among propositions, their text representations will not be coherent. Many of the linking propositions in the protocols were obviously signaled by the headings and subheadings in the web pages. Risk factors, symptoms, and treatments are just a few of the headings that were mentioned by several of the participants. Of course, reader-based sources of coherence such as knowledge and strategies also play a role, as highlighted by the situation model. For example, readers identify causal relations among events to construct coherent representations of narratives, and readers draw on background knowledge to integrate text statements with each other and their prior knowledge. While the only gauge of knowledge was if the participant mentioned having previously researched their cancer type, the conditional type linking propositions made required that some prior world knowledge. For instance, the participant who linked a high PSA level with the diagnosis of prostate cancer had to possess the knowledge that a high level suggested the presence of a prostate condition and that it was

cancer. This required a causal relation because the participant needed to reactivate from memory the statement that “certain prostate conditions can cause larger amounts of PSA to leak into the blood” with “prostate cancer is a possible cause of a high PSA level”.

5.2.1.5 Semantic Network Representations

5.2.1.5.1 Results

Figures 27 and 28 contain two examples of protocol generated SNRs. The clustering of information was determined by the questions asked. The graphical representation of a protocol by participant 1 responding to questions about the low readability colorectal cancer web page is given in Figure 27. This protocol is an example of a SNR that has adequate local coherence but poor global coherence. The cluster labeled “Describing the article” illustrates the semantic connections of the propositionalized response to the general question “Please describe the information from the web page as you would to a friend or a family member”. The participant, 64 year old female, shared that while she was fortunate not to have a family history of cancer, she was still interested in preventative cancer information. She began by defining cancer as the growth of cells, and that this was more likely to occur if a person had a family history or if it was a reoccurrence. Looking at the SNR, one can follow this first set of propositions by beginning with the node cancer located midway down in the cluster, on the right side. She then continued to explain that it makes a difference where the problem is located, that it worsens if it is in the inner lining of the intestine, throughout the colon, or throughout the body. The participant went on to specify that there are two hereditary kinds of prostate cancer, one where the whole intestine had to be removed during a person’s twenties, and the other where nothing has to be removed. This second set of propositions begins in the upper right hand corner of the cluster with the node problem and continues vertically all the way down to the nodes first problem and second problem. She finished describing the article by outlining the different tests

used for diagnosis. The participant could not remember if it was called a colonoscopy or a sigmoidoscopy, but she described one test as a little tube with lights that retrieves cells or tissues to be looked at under a microscope. She also mentioned another one that consisted of a series of x-rays. This third set of propositions is contained in the bottom left hand corner of the cluster. Note that it is at the same level as the second set because the node colon is common to two node-link-node triplets within both sets. All other clusters were generated from the subsequent follow-up questions. One can immediately see that the cluster labeled “Risk factors for developing colorectal cancer” contributes to the poor global coherence. All though the nodes within that cluster are all interconnected, the cluster remains isolated from the other clusters.

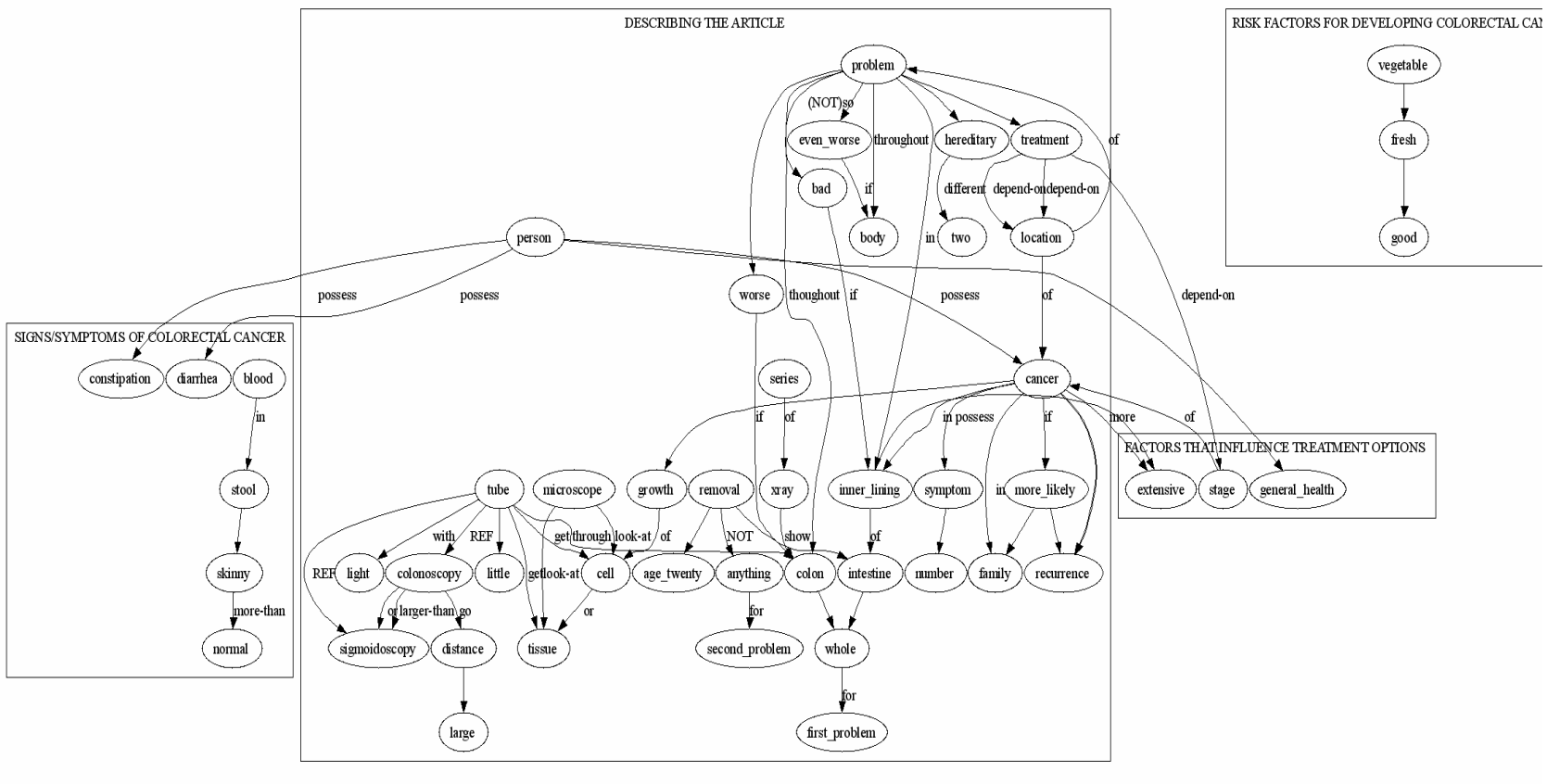


Figure 27. SN of participant 1's protocol for the low readability colorectal cancer web page.

In Figure 28, a more locally and globally coherent SNR is presented. This representation is by the same female, except this time responding to the high readability colorectal cancer web page. Here, all clusters are connected by a minimum of one node. For example, the “Describing the article” cluster connects through the linkage “person-exercise”. Then there are the clusters “Describing the article” and “Prevention or treatment”, which are connect by two linkages, “person-remove-colon” and “colon-genetic-predisposition”. The cluster “Foods that reduce risk” is even more interconnected through the node risk. Although this protocol could have been more coherent if nodes were connected to several other nodes instead of just one, it serves as a sufficient contrast of the SNR presented in Figure 27. The participant began by stating that good general health and well-being and being active can protect a person from developing colorectal cancer, but not if there is a genetic or bowel predisposition present. She then stressed the importance of fresh fruits, vegetables, and whole grains, as well as vitamins, especially vitamin D and folic acid. From the SNR, one can see that the central cluster is person-oriented. There are multiple arrows that originate from the person node and link to nutritional, lifestyle, and genetic components protocol. It is also through that node that three of the clusters are connected. Only the food cluster is indirectly connected to the person node through the concepts of vitamin D and folic acid. Although the participant left out many of the smaller details of the high readability web page, she managed to capture the primary message components, as illustrated by the good global coherence. While the follow-up questions were not specifically designed to measure coherence, they did touch upon important facts of the text, so it is significant that all the clusters were connected to the main cluster.

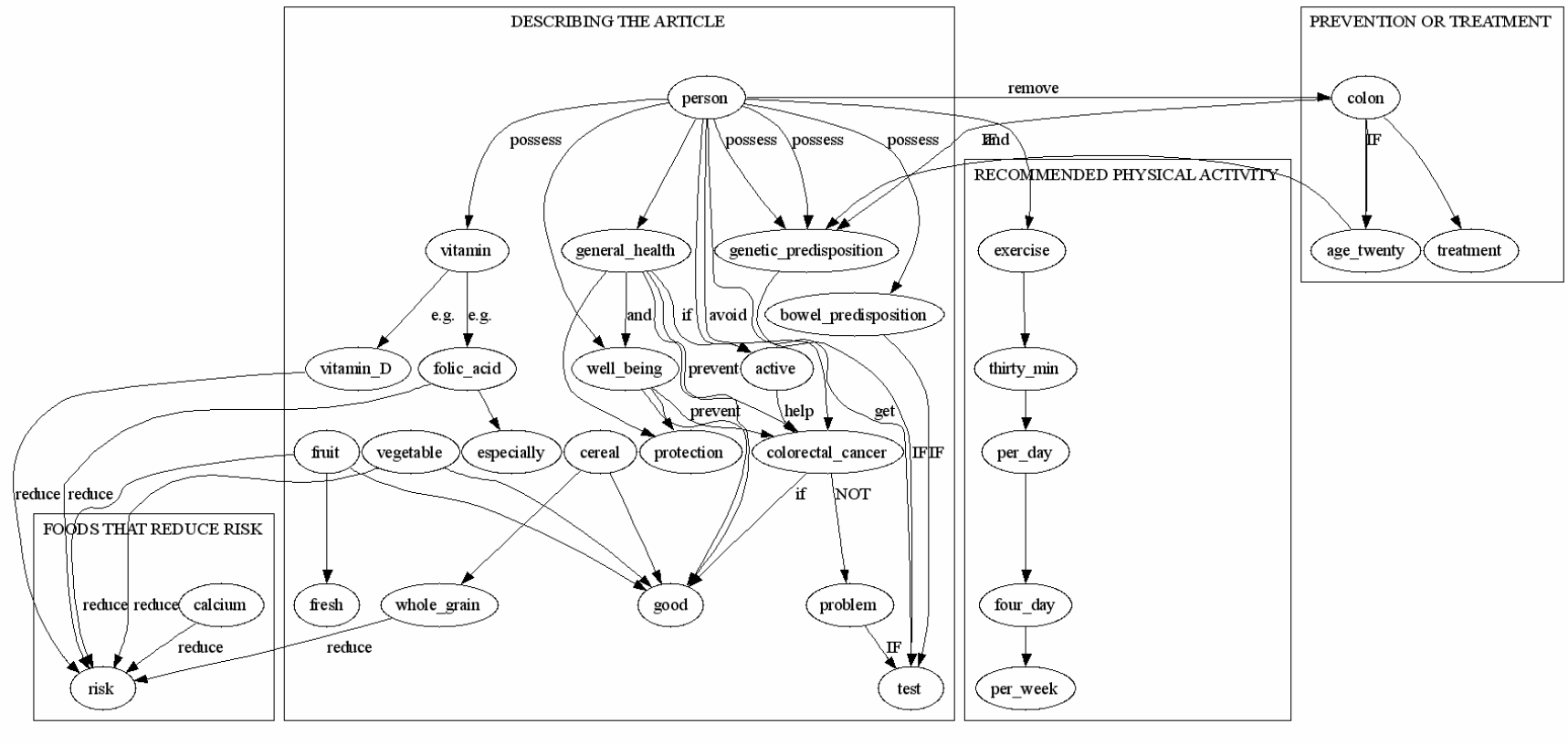


Figure 28. SNR of participant 1's protocol for the high readability colorectal cancer web page.

5.2.1.5.2 Discussion

As put forth in question 1B, the inconsistencies in the levels of coherence are more easily seen in the protocol generated SNRs than in the web page ones. This is primarily due to the range of questions that were posed to the participants. A participant who responded well to the "Describe the article" question was likely have a more interconnected representation simply because there are more nodes available to be linked to other clusters. The follow-up questions tended to focus on the main points of the web page content. Thus, not only can one assess coherence from the protocol SNRs, but one can also see the degree to which the participant addressed the main points of the web page in his or her initial description of the content. The SNRs also allowed the examination of specific links between clusters of information that contribute to local and global coherence. For the high readability example described earlier, it was evident that the person node served as the important link to other units of meaning.

5.2.1.6 Analysis of Sample Protocols

Given the limited number of participants in this study and the wide variability of their results, it is also important to focus on how the measures of coherence and recall relate to the individual protocols and their respective mental representations. In doing so, one can investigate individual comprehension processes such as how coherence manifests itself when different approaches are used.

5.2.1.6.1 Example 1: Female, High Readability, Colorectal Cancer

Example contains the SNR of participant 2's protocol. The participant is a 63 year old retired female who reported the Internet to be her primary source of cancer information. In Figure 31, one can see that the participant provided a rich response to the general question, but only answered a single follow-up question regarding recommended physical activity. She began by saying that the web page primarily dealt with things people can do to lower their risk and then went on to specify the

importance of fruits, vegetables, whole grains, and physical activity. She remarked that she found it interesting that taking a multivitamin containing folic acid could lower one's rate, as well as the nonsteroidal anti-inflammatory drugs (NSAIDs), such as Motrin or Advil, which can reduce the rate by 20 to 50%. But then she noted that these drugs could cause stomach irritation and hemorrhages, whereas female hormones, which also may slightly reduce one's risk, may also decrease the rate of heart disease, blood clots, and breast and uterine cancer. She ended her explanation by saying that there is genetic test for people at high risk, and gave the Ashkenazi Jews as an example of those at high risk.

This participant's representation had a good global coherence, as the general question was connected the follow-up question by a couple of linking nodes, as well as possessing good local coherence. The nodes in grey represent inferences made. For example, in the web page, it was explained that NSAIDs can cause serious or even life-threatening bleeding from stomach irritation. This participant inferred that "life-threatening bleeding" to mean "hemorrhage", which is profuse bleeding that can indeed be life-threatening. Another inference example is the substitution of "European Jewish" for "Ashkenazi Jews", as a group that has a slightly increased colorectal cancer risk. From this mental representation, one can also note that the concepts of female hormones, NSAIDs, and risk were among the concepts repeatedly mentioned. This is shown by the multiple links or arrows originating or ending at these nodes.

Her recall was described earlier in the participant recall section, but to summarize, it was superior in both number and level of detail of any other participant. Though she use more words to describe the low than high readability page (231 vs. 169), a greater proportion of her propositions were correct for the high than low page (42 vs. 29%). Her high readability protocol also contained a greater proportion of repeated concepts. It would appear that her comprehension of this page was more clear and complete for this page, regardless of its high readability level. It must be said that

reader-based factors such as previous exposure to her cancer type and having had a family member ill with the same disease likely played a part in her comprehension.

Across all the participants, there were definite commonalities in what was recalled. What made this participant stand out, however, was in what she recalled. Certain parts of the web page – diet and exercise, vitamins, and some general symptoms – were easily understood by all participants. She, on the other hand, recalled specific details, such as drug names and rates at which risk was lowered. The complex interaction of recalling higher level generalizations versus details units fits within van Dijk and Kintsch's (1983) framework of multiple representations. If a text is written well and provides appropriate signals to the reader as to what is important and what is not, readers are generally able to form an adequate textbase, even with minimal domain knowledge. Behaviours that are primarily based on the textbase include recognition and reproductive recall, as well as summarization. The situation model, on the contrary, represents the content of the text, the situation described by it, and what is already known by it. The formation of the situation model is influenced by the amount and the structure of the reader's domain knowledge. An appropriate situation model forms the basis for tasks that require one to apply the new knowledge in some way. However, although the distinction between the textbase and situation model is important for analytical purposes, the two representations are by no means independent. Means and Voss (1985) reported finding comprehension differences as a function of the quality of a readers' knowledge schemata. Better recall was observed for readers with more specific, elaborate knowledge schemata than for readers who could rely only on very general schemata. Readers that have no specific knowledge about a text are unable to form an adequate textbase because they cannot fall back on their general knowledge. In the case of this participant, it is believed that her background aided in the comprehension of the web page. Though prior knowledge was not screened for in this study, her representation suggests that she had an existing schema and general degree of familiarity with the text.

5.2.1.6.2 Example 2: Male, High Readability, Colorectal Cancer

Example 2 is about participant 10, a 68 year old retired male who reported using various sources such as books, magazines, and the Internet for cancer information. Figure 32 contains his SNR. In contrast to the first example, example 2 shows that the participant failed to produce any relevant response to the general question, but then went on to respond in moderate detail to the follow-up questions. The clusters about drugs and foods that reduce risk had good local coherence between themselves, as demonstrated by the links between and within these clusters. The links between these two clusters are more a function of the questions asked, as both dealt with aspects of risk, which was the concept that served as the common link in this representation. For example, the participant mentioned that food high in fibre, fruits, and vegetables, as well as aspirin can reduce one's risk of developing colorectal cancer. The remaining two clusters, however, contribute to the lack of overall global coherence, as they have no connections to other nodes in the other clusters. The paucity of the response to the general question is likely the primary reason the representation lacks of global coherence because without the initial presence of relevant nodes, subsequent propositions have fewer connection options.

His correct recall was impressive for the high readability page compared to the low one, which was non-existent (6 vs. 0 respectively), though he expressed about the same number of propositions for both pages (26 vs. 23 respectively). Three of the recalled propositions were in the food cluster; the remaining two were about physical activity. His inferences (i.e., the propositions in grey) were spread out over three clusters, one each in the drug, food, and physical activity boxes. His elaboration appeared to be most prominent when he talked about the recommended foods. Not only did he infer that "[everything except] the food that taste good like red meat and ice cream", but he also expanded to incorrectly state that high fibre foods also reduce risk. It would seem that this participant's framework included a general schema for good and bad food items, rather than one

specific for foods recommended to reduce colorectal risk. Also, his inference that "aspirin is like the cure all" again seems like a generalization for a drug that has multiple benefits, a message learned from various sources over a period of time, rather than a definitive statement from the web page. The text stated that "people who use aspirin… have a 20% to 50% lower risk of colorectal cancer and adenomatous polyps", a statement that discussed aspirin to have a specific reduction of risk for a particular health condition.

This representation seems to be characterized by succinctness in that the follow-up questions were required to act as probes to comprehension. If, for example, one had simply analysed the general question, the conclusion would have been that the participant either could not verbalize a response, or had failed to comprehend any of the web page text. Of course, the responses to the follow-up questions demonstrate that comprehension existed, though in a verbally concise form. Whereas the participant in example 1 was more disposed to elaborate about concepts, example 2's representation is more focused. This difference in characterization highlights perhaps one of the invariants in the representations of people's comprehension: elaborative vs. conciseness. With respect to van Dijk and Kintsch's (1983) framework, the second participant did not report having done previous research on his cancer type. He did, however, express interest in learning about the PSA test, false positives, and its associated costs, all topics not discussed in the high readability page. It is possible that these factors contributed to a situation model that was inadequate for internalizing the content of this particular web page. A situation model can incorporate previous experiences, and hence also previous textbases, regarding the same or similar situations. Whether it was a lack of interest or an impoverished background knowledge, it is appears that this participant's situation model failed to frame the high readability page in a manner that allowed a more coherent representation of its information.

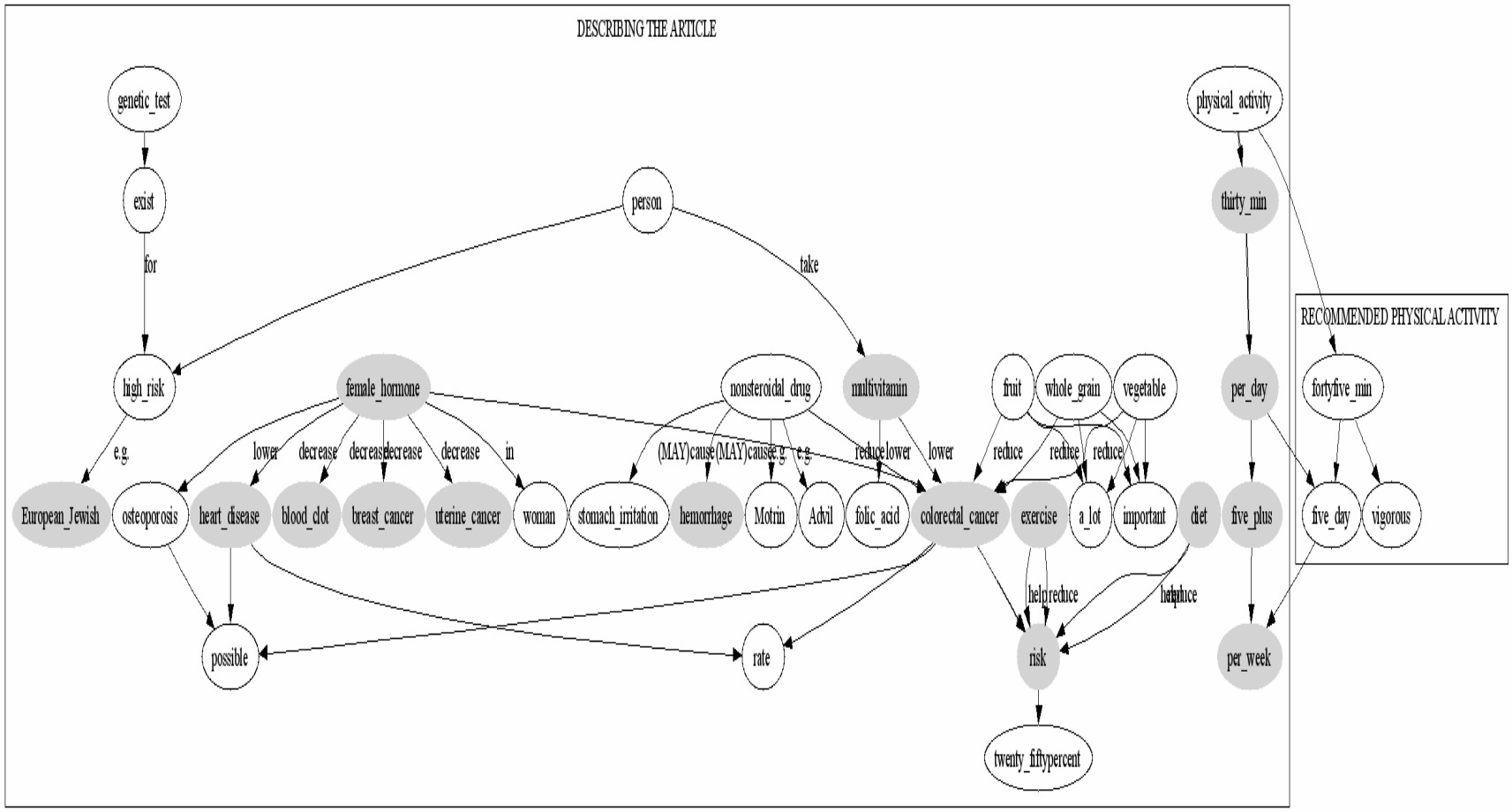


Figure 29. Mental representation of participant 2's protocol for the high readability colorectal cancer web page.

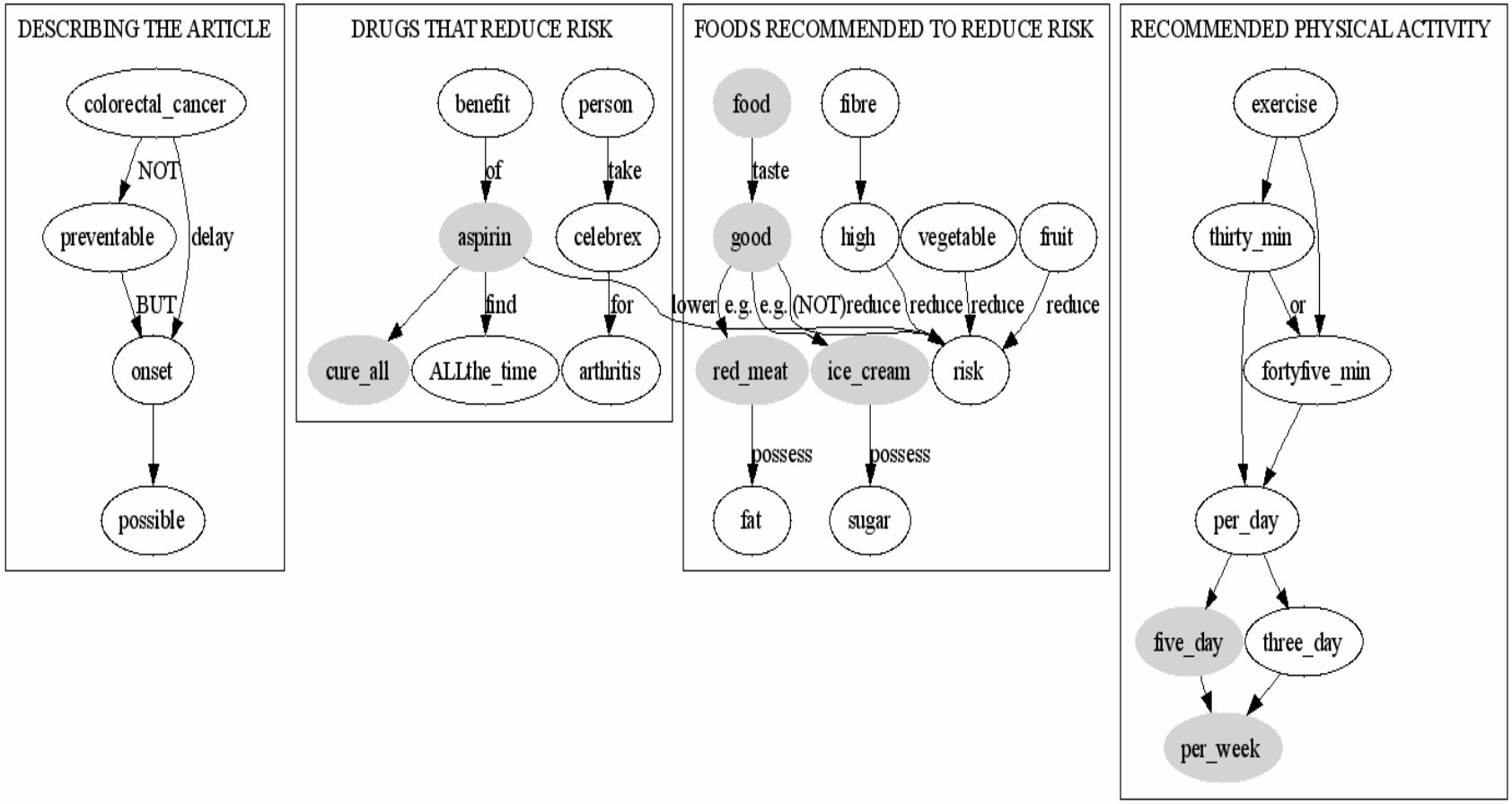


Figure 30. Mental representation of participant 10's protocol for the high readability colorectal cancer web page.

5.3 Summary of Results

Question 1A: Will there be an association between FRE, F-K, and SMOG scores and propositional density?

Results: Of the three readability formulas calculated, the SMOG scores were the most similar to propositional density scores. The FRE and F-K produced a different ranking of difficulty, and in some instances, were even inconsistent among themselves.

Question 1B: Will levels of local and global coherence, as discerned by the deconstruction of the text to its propositional textbase and then reconstructed into semantic network representations, have an association to FRE and F-K scores?

Results: All of the web page SNRs had comparable levels of local and global coherence, possibly due to the high proportion of concept repetition. The sensitivity of the SNR as a measure of coherence was inadequate in this case.

Coherence, as measured by embedding and linking propositions, showed that only the breast cancer text had greater coherence in the high than low readability page. The other cancer types, colorectal and prostate, had low readability pages that were more coherent than their high ones.

Question 2: Will participants produce more coherent and detailed protocols, and make a greater use of inferences for gender specific rather than for gender neutral cancer information?

Results: Overall, female recall and inferencing was greater for the gender neutral cancer information. For males, the gender specific cancer information was neither recalled nor inferenced more than the gender neutral type. It is important to note that since there was no screening for prior knowledge, that gender was not a valid measure for questioning whether there would be greater comprehension for one cancer type over another.

Coherence analysis by way of embedding and linking propositions revealed that participants generated more propositions about information within than between paragraphs. Of course, the nature of the interview questions contributed to the focus on within paragraph materials. The linking information was found to represent general comments that summarized the main points of the texts. This was done by the majority of the participants.

Individual protocols:

One variation in the mental representations of the web pages was elaborative versus conciseness. Some participants responded well to both the general and follow-up questions. Others needed the follow-up questions to act as probes to comprehension. Speaking style is also related to this invariant: Some participants used a lot of words to express very few relevant concepts, while others were highly focused.

Though not explicitly measured, interest seemed to influence the propositions in each participant's representation. Interest was indirectly gathered through comments about what specific subtopics they were interested in (e.g., diagnostics, treatment, etc.) and if they had personally affected by their cancer types (e.g., they themselves, or a close friend or family had the disease).

CHAPTER 6: Limitations

The data analyzed in this study were part of a larger research program (Friedman, Hoffman-Goetz, & Arocha, 2004, in press) designed to investigate health literacy and comprehensibility of health information naturally available on the Internet. The web pages used were selected so that they were representative of those most likely to be found by people using search engines, providing ecological validity to the data (Cicourel & Katz, 1996). The data were collected by another researcher for the purposes of the research program. Inconsistency in data collection was minimized in this study by having the same interviewer collect all the transcripts, adhere to a script, and counterbalance the order of the web pages. However, experimental control of the information presented was not possible.

The data themselves were limited to the responses elicited by the general and follow-up questions. For example, one participant answered the general question with only one or two sentences despite probes to encourage more verbal response. However, this same participant was able to provide correct response to the follow-up questions. This shows that the paucity of a response to the general question is not necessarily indicative of a participant's knowledge or comprehension of the web page. That being said, one can also argue that a short response to the general prompt is still valuable, as it may contain the most important propositions. Then, responses to any follow-up questions can be viewed simply as an extension of a participant's knowledge.

Runkel (1990) noted that if information generated from research is to be used locally, with a particular group or person, then generalization must go the opposite way. This is to say that research must go from describing what is true for many people to what is true to a few or even one person. Runkel also contrasted the traditional experimental method of research which he called "the method of relative frequencies" to that of "the method of possibilities", which focuses on what people <u>can</u> do rather than on what they <u>most likely</u> do. For example, in a study in which nine of 10 participants

produce a result, the remaining one participant is generally labelled as having failed to produce the desired behavior. Runkel would argue that the single case is still interesting because that participant is still doing something. This perspective is illuminating for this present study as it highlights that while the results here are limited to a small group of participants, each participant provides rich information about individual differences.

A second limitation is related to the demographics of the participants. Since the data source used a convenience sample of adults aged 50 and older, this study had little control over the selection of participants on most demographical information. However, since this study was focused on capturing aspects of the real world, it is likely that participants who signed up for the Internet search workshop and took part in the subsequent comprehension testing session had genuine interest in learning about cancer information. Indeed, those persons who participated may very well be within the targeted population for online cancer information providers because: a) they expressed interest in cancer information; and b) they acquired the skills to conduct searches for this information and discern it in terms of quality; and c) they may use these skills in the future to seek cancer information.

A third limitation is that the participants' level of prior knowledge was not assessed. Instead, gender-specific and gender neutral information were used as proxies for prior knowledge. The result is that this study could not properly infer the influence of prior knowledge on the comprehension of these web sites. Prior knowledge is known to influence comprehension (Mannes & St George, 1996).

A fourth limitation of this study was related to pragmatics. In any communicative act or speech, there exists a dynamic aspect of meaning in context. For this study, this could have included the goals and emotions of the participants. For example, if a participant read the web pages with the goal of learning about various cancer treatments, then he or she may have inadvertently read other sections at a superficial level. This in turn may have translated in fewer propositions being recalled for non-goal oriented information. Though the aim of this study did not focus on psychological

aspects that can produce a distinction between the literal meaning of a sentence (i.e., sentence meaning) and the concepts that a participant tries to convey (i.e., speaker meaning), one must acknowledge the role of person-based factors and their effects on verbal protocols.

CHAPTER 7: Conclusion & Implications

The growing popularity of the Internet has made it easier and faster to find health information. Much of the information is educational, which makes the Internet an invaluable dissemination resource, especially for consumers who may not otherwise have access to the material. It is equally, if not more important, for online health information providers to know that their readers understand the information. Unfortunately, efforts to make text more comprehensible often rely solely on readability formulas.

This study attempted to demonstrate the usefulness of propositional analysis as a method for the empirical investigation of the semantic structure for a select group of cancer web pages and for the verbal protocols of a select group of community dwelling older adults. Propositional analysis highlights the semantic complexity and relations of a person's verbalization and helps to characterize the process by which comprehension occurs. Several measures of comprehension are derived from this methodology, including concepts, inferences, and local and global coherence.

This research has also addressed some of the problems relating to the reliance on readability formulas in the evaluation of comprehension in the health domain and sought a complementary, or possibly alternative, tool in assessing comprehensibility. The tool was in the form of propositional density (P-D), or the amount of information contained within a given amount of text. The expectation was that a propositionally dense text would increase comprehension difficulty because there is more information to input, process, and synthesize. Another expectation was that P-D would differ from readability formula results because the two tools have different properties. Results showed that P-D are associated with the SMOG, but not with the FRE and F-K readability scores. However, despite this association, in principal, there is no reason to suppose that P-D and SMOG should be related. This indicates that P-D should still be considered as an independent comprehension marker.

Furthermore, the lack of agreement among the three readability formulas themselves suggests that more than one kind of readability formula should be used. A reliance on any one readability formula by a health information provider may result in the misleading conclusion about a text's reading difficulty (Meade & Smith, 1991). Of the readability formulas analysed in this study, the SMOG formula, which has a reputation for reading level accuracy, simple directions, and speed of use (Meade & Bryd, 1989; Romano, 1979), appeared to have the greatest agreement with P-D.

This study also addressed whether coherence, as shown through semantic network representations, would differ according to readability ratings. For the web pages, the representations did not adequately differentiate between low and high readability scores. It was difficult to identify clear breaks in both local and global coherence, although the SNRs were still useful for determining relations between main and subordinate ideas. The arrows originating and ending at a node also served as an indication of the frequency in which a concept appeared in a web page. When coherence was measured as the proportion of linking and embedding propositions in a web page, there was a mismatch between coherence and readability scores for the breast cancer pages. The low readability breast cancer page, denoting easier reading, actually had a lower coherence score than its high readability page**,** further supporting the notion that readability, as measured by standard formulas, is independent of coherence.

The constructivist theory in Kintsch's (1998) model of text comprehension proposes that meaning or knowledge is constructed in the mind of the reader, and this construction is an active process. The theory also proposes that the reader generally forms a representation in his or her mind of the meaning of the text that is partly based on information in the text and partly on the reader's background information, which is the situation model. The reader seeks coherence in his/her situational models; that is, the models need to make sense. For many of the participants in this study, it became evident that self-interest and goals appear to have influenced what they later spoke about in

their protocols. Glenberg and Mathew (1992) pointed out that reading for good comprehension is often goal directed and strategic. An example of goal-directed reading occurs when a reader is searching for the answer to a question; non goal-directed reading can involve reading a text without a specific purpose. Zwaan and Graesser (1993) extended this point by suggesting that the amount of and type of inferences readers generate are a function of the reader's goals. Most inferences made during comprehension are knowledge based which include generic packages of world knowledge, such as scripts, frames, stereotypes, or schemata. Recall is similarly influenced such factors.

For the third question of this study, it was asked if there would be greater coherence, recall, and inferencing for gender-specific cancer information than for gender-consistent cancer information. This question assumed that female participants would have a more established schemata for breast cancer than for colorectal cancer, and males have more established schemata for prostate cancer than for colorectal cancer.

The exploratory groundwork of this study suggests that further research is warranted to better characterize how people process online health information. There exist alternative methods of measuring text comprehension, including text coherence (Foltz, Kintsch, & Landauer, 1998), such as latent semantic analysis (LSA) which extracts and represents the similarity of meaning of words and passages by analysis of large bodies of text (Landauer, Foltz, & Laham, 1998). There also is the method of constraint satisfaction (Thagard & Verbeurgt, 1997), in which maximizing coherence is a matter of maximizing satisfaction of a set of positive and negative constraints between pieces of information. Using one of these alternative methods to assess coherence and its relationship with readability would extend this study's work and supply additional evidence for online health information providers to use coherence measures when writing or revising their text. LSA may be particularly useful in the study of health literacy in larger populations, as it is a computerized tool that produces measures about word-word, word-passage, and passage-passage relations. Though LSA

does not capture semantics in the manner of propositional analysis, it does provide additional information (e.g., coherence) to complement readability scores. So even if the tailoring of information so that it is fully coherent and contains the appropriate amount of propositions is not the ultimate goal for these providers, research on these factors could provide useful guidelines to assist human text writers in the advancement of overall consumer comprehension.

Appendix A: Web pages

Breast cancer – Low readability web page

Early Detection of Breast Cancer

Early Signs of Breast Cancer

- A lump is detected, which is usually single, firm, and most often painless.
- A portion of the skin on the breast or underarm swells and has an unusual appearance.
- Veins on the skin surface become more prominent on one breast.
- The breast nipple becomes inverted, develops a rash, changes in skin texture, or has a discharge other than breast milk.
- A depression is found in an area of the breast surface.

Women's breasts can develop some degree of lumpiness, but only a small percentage of lumps are malignant.

While a history of breast cancer in the family may lead to increased risk, most breast cancers are diagnosed in women with no family history. If you have a family history of breast cancer, this should be discussed with your doctor.

Facts

- Every two minutes a woman is diagnosed with breast cancer.
- This year more than 211,000 new cases of breast cancer are expected in the United States.
- One woman in eight who lives to age 85 will develop breast cancer during her lifetime.
- Breast cancer is the leading cause of death in women between the ages of 40 and 55.
- 1,600 men are expected to be diagnosed with breast cancer this year and 400 are predicted to die.
- Seventy percent of all breast cancers are found through breast self-exams. Not all lumps are detectable by touch. We recommend regular mammograms and monthly breast self-exams.
- Eight out of ten breast lumps are not cancerous. If you find a lump, don't panic-call your doctor for an appointment.
- Mammography is a low-dose X-ray examination that can detect breast cancer up to two years before it is large enough to be felt.
- When breast cancer is found early, the five-year survival rate is 96%. This is good news! Over 2 million breast cancer survivors are alive in America today.

Detection Plan

An Early Breast Cancer Detection Plan should include:

- Clinical breast examinations every three years from ages 20-39, then every year thereafter.
- Monthly breast self-examinations beginning at age 20. Look for any changes in your breasts.
- Baseline mammogram by the age of 40.

- Mammogram every one to two years for women 40-49, depending on previous findings.
- Mammogram every year for women 50 and older.
- A personal calendar to record your self-exams, mammograms, and doctor appointments.
- A low-fat diet, regular exercise, and no smoking or drinking.

How to do a Breast Self-Examination

IN THE SHOWER Fingers flat, move gently over every part of each breast. Use your right hand to examine left breast, left hand for right breast.

Check for any lump, hard knot or thickening. Carefully observe any changes in your breasts.

BEFORE A MIRROR Inspect your breasts with arms at your sides. Next, raise your arms high overhead.

Look for any changes in contour of each breast, a swelling, a dimpling of skin or changes in the nipple. Then rest palm on hips and press firmly to flex your chest muscles. Left and right breasts will not exactly match - few women's breasts do.

LYING DOWN Place pillow under right shoulder, right arm behind your head. With fingers of left hand flat, press right breast gently in small circular motions, moving vertically or in a circular pattern covering the entire breast. Use light, medium and firm pressure. Squeeze nipple; check for discharge and lumps. Repeat these steps for your left breast.

Breast cancer – High readability web page

More Than Mammograms: MRI for High-Risk Women

Written by:
Christine Haran -

Published on: October 5, 2004

Physicians and women have long awaited a better early detection tool for breast cancer than mammograms. And a recent study suggests that a new option, magnetic resonance imaging (MRI), can be added to the screening arsenal, though it's only recommended to women at high genetic risk for breast cancer.

Breast cancer screening is a top concern for women who have an inherited abnormality in one of two genes, known as the BRCA1 and BRCA2 genes. Yet few studies have examined what screening approach is best for these high-risk women. BRCA1 and BRCA2 mutation carriers are identified though blood tests that are given to certain women with a strong family history of breast cancer. Women of Ashkenazi Jewish descent are at particular risk: The mutations occur in about 2.5 percent of these women, compared to 1 percent of the general population.

A study published in the September 15th issue of the Journal of the American Medical Association looked at the usefulness of MRI, mammograms, ultrasound and clinical breast exams, which are breast exams performed by a healthcare professional, in screening healthy women with the BRCA1 and BRCA2 gene mutations for breast cancer. The researchers found that MRI, in which magnets and radiowaves create an image of a body part, was a helpful addition to the surveillance programs for these women.

Other risk factors for breast cancer include age, family history, use of hormone replacement therapy, radiation exposure, early onset of the menstrual period and late menopause. However, the JAMA study only examined women with one of the BRCA mutations. Breast cancers related to BRCA1 or 2 mutations make up about 5 percent of all breast cancers.

Below, study author Sandra Messner, MD, the medical coordinator of clinical breast services in preventive oncology at the Toronto Sunnybrook Regional Cancer Center, discusses the best screening options for these high-risk women to help make sure any breast cancer they develop is detected as early as possible.

What are the BRCA mutations?
There are two large genes that have been identified in all women called BRCA1 and BRCA2. They function, we think, as tumor-suppressor genes, so they keep cancers from developing. If they are abnormal or what we call mutated, then your risk of cancer is increased, and the risk of breast cancer and ovarian cancer particularly is affected. If a woman carries an abnormality in one of those two genes, she is thought to have up to an 85 percent lifetime risk of breast cancer. The average woman's lifetime risk of developing breast cancer is about 11 percent. The mutations also increase the risk of

ovarian cancer, more so with BRCA1 than with BRCA2. The BRCA2 mutations may also increase risk of other cancers, such as melanoma and pancreatic cancer.

It is possible to carry mutations in both genes, but it is very rare. There is no conclusive data on risk in this situation, but most experts suggest that the risk is not any higher with both genes than only one.

What prevention strategies have been recommended to women with these mutations?
There hasn't really been any kind of standard approach, but women who carry mutations in these genes have several options. The most drastic is to have their breast removed with a preventive mastectomy. This reduces the risk of developing breast cancer by about 90 to 95 percent. It's a pretty drastic thing, however.

We also talk to these women about the potential role of tamoxifen, which works by blocking estrogen receptors in women with hormone-positive breast cancer, to reduce their risk of breast cancer. There have been some papers that suggest it's also beneficial in mutation carriers. But there have been a lot of questions about whether it's as beneficial in BRCA1 carriers as it is in BRCA2 carriers, because tamoxifen works by blocking estrogen receptors, and three-quarters of the cancers that are diagnosed in BRCA1 carriers are estrogen-receptor negative. With BRCA2 mutations, about 75 percent of cancers are estrogen-receptor positive.

However, people don't like to take drugs when they are healthy. One of the side effects with tamoxifen can be an increased risk of uterine cancer. The uterine cancer risk is less than 1 percent. But if you've already got a fear of cancer because you're at such high risk, that's a scary thing.

What kind of screening is recommended to women with BRCA1 and 2?
The current screening recommendation for mutation carriers is annual mammography starting somewhere between 25 and 35 and clinical examination of the breasts every six to 12 months. Breast self-examination has been recommended by some people, although that the effectiveness of this screening tool is controversial. The only difference between these recommendations and those for the average woman is that you might do the mammograms more often; yearly instead of every two years and you certainly would start at a younger age.

Why did you decide to conduct your study?
There is a question as to whether mammography works as well in women with the BRCA mutations as it does in the general population because the kinds of cancers that they get, especially with BRCA1, seem to be somewhat different. They don't, for some reason, show up on mammography as well as some of the other kinds of breast cancers that other women get, possibly because they are less likely to form a mass initially. Because MRI was an up-and-coming technology and seemed to be useful in detecting other medical conditions, it was being considered for breast cancer. So, along with ultrasound, which was another technology that people were suggesting we could add to mammography to give us more information, we thought we'd see which of the techniques was best or whether a combination of the techniques was best.

What did your study find?
When our study was published, there were 22 breast cancers that had been identified in more than 200 women. What was interesting was that a lot of these cancers were found only on MRI. Some were found only on mammograms, and ultrasound generally didn't, overall, seem to contribute very much.

The MRI seemed to pick up the most cancers, and it seemed to pick them up at a relatively early stage.

Should a woman with BRCA1 or 2 be screened with MRI?
At this point, we're recommending that mutation carriers have MRI yearly and mammogram yearly starting between ages 25 and 30. Mammograms need to be done as well because some types of early breast cancer, specifically ductal carcinoma in situ, that show up on the mammogram as little specks of calcium, what we call microcalcifications, don't show up on MRI. We're not suggesting that women have the addition of ultrasound, because we think adding ultrasound doesn't really add much more. We also recommend a clinical breast examination twice a year.

Why isn't MRI recommended to the general public?
There are a lot of things that show up on MRI that aren't cancer, such as cysts and nodules, so you have to sort that out, hopefully, without having to do surgery to prove it. False-positive results are still a concern in women with the BRCA mutations, but the higher the risk of cancer, the more one is willing to tolerate the downside of false-positives.

But given the high number of false-positive results and knowing the technical problems with MRI, I don't think it will ever become a standard screening tool for the general population. It's expensive and time-consuming. It often takes almost an hour to do a screen. It involves an injection for the imaging. There is also a big issue regarding claustrophobia because the space inside the magnet is smaller in a breast MRI machine than it is in a regular MRI machine.

Do you recommend MRI to all high-risk women?
The results of our study are applicable only to mutation carriers. This does not imply that MRI is beneficial for all high-risk women because there are many other reasons for women to be at high risk, and we don't yet have evidence to say that all high-risk women should get MRI screening.

What research questions are still unanswered?
There's a whole question of: Do you alternate between MRIs and mammograms? Do you do something every six months? Also the question has been raised: Would it be better to do only MRI in women when they're very young since their breasts are very dense, making mammograms hard to read? These are all questions that we're playing with to figure out what would be ideal. The next step in our study is to see if we've had an impact on survival, because, obviously, that's the endpoint that we really want to reach.

What is your overall advice to a woman with one of the BRCA mutations?
Given the fact that the field is constantly changing and there are new recommendations and new studies coming out all the time, I think that it's important to hook yourself up with a high-risk program, if you possibly can. There are issues not just around breast cancer, there are issues around ovarian cancer. With BRCA2, there is increased risk of melanoma and pancreatic cancer. There are also psychological issues. This is a very scary thing for women, and most high-risk programs offer psychological counseling. So my advice is to try to get your screening through a high-risk program.

Colorectal cancer – Low readability web page

General Information About Colon Cancer

Key Points for This Section

- Colon cancer is a disease in which malignant (cancer) cells form in the tissues of the colon.
- Age and health history can affect the risk of developing colon cancer.
- Possible signs of colon cancer include a change in bowel habits or blood in the stool.
- Tests that examine the rectum, rectal tissue, and blood are used to detect (find) and diagnose colon cancer.
- Certain factors affect prognosis (chance of recovery) and treatment options.

Colon cancer is a disease in which malignant (cancer) cells form in the tissues of the colon.

The colon is part of the body's digestive system. The digestive system removes and processes nutrients (vitamins, minerals, carbohydrates, fats, proteins, and water) from foods and helps pass waste material out of the body. The digestive system is made up of the esophagus, stomach, and the small and large intestines. The first 6 feet of the large intestine are called the large bowel or colon. The last 6 inches are the rectum and the anal canal. The anal canal ends at the anus (the opening of the large intestine to the outside of the body).

Age and health history can affect the risk of developing colon cancer.

Risk factors include the following:

- Age 50 or older.
- A family history of cancer of the colon or rectum.
- A personal history of cancer of the colon, rectum, ovary, endometrium, or breast.
- A history of polyps in the colon.
- A history of ulcerative colitis (ulcers in the lining of the large intestine) or Crohn's disease.
- Certain hereditary conditions, such as familial adenomatous polyposis and hereditary nonpolyposis colon cancer (HNPCC; Lynch Syndrome).

Possible signs of colon cancer include a change in bowel habits or blood in the stool.

These and other symptoms may be caused by colon cancer or by other conditions. A doctor should be consulted if any of the following problems occur:

- A change in bowel habits.
- Blood (either bright red or very dark) in the stool.
- Diarrhea, constipation, or feeling that the bowel does not empty completely.
- Stools that are narrower than usual.
- General abdominal discomfort (frequent gas pains, bloating, fullness, or cramps).
- Weight loss with no known reason.
- Constant tiredness.
- Vomiting.

- Tests that examine the rectum, rectal tissue, and blood are used to detect (find) and diagnose colon cancer.

The following tests and procedures may be used:

- Physical exam and history: An exam of the body to check general signs of health, including checking for signs of disease, such as lumps or anything else that seems unusual. A history of the patient's health habits and past illnesses and treatments will also be taken.
- Fecal occult blood test: A test to check stool (solid waste) for blood that can only be seen with a microscope. Small samples of stool are placed on special cards and returned to the doctor or laboratory for testing.
- Digital rectal exam: An exam of the rectum. The doctor or nurse inserts a lubricated, gloved finger into the rectum to feel for lumps or abnormal areas.
- Barium enema: A series of x-rays of the lower gastrointestinal tract. A liquid that contains barium (a silver-white metallic compound) is put into the rectum. The barium coats the lower gastrointestinal tract and x-rays are taken. This procedure is also called a lower GI series.
- Sigmoidoscopy: A procedure to look inside the rectum and sigmoid (lower) colon for polyps, abnormal areas, or cancer. A sigmoidoscope (a thin, lighted tube) is inserted through the rectum into the sigmoid colon. Polyps or tissue samples may be taken for biopsy.
- Colonoscopy: A procedure to look inside the rectum and colon for polyps, abnormal areas, or cancer. A colonoscope (a thin, lighted tube) is inserted through the rectum into the colon. Polyps or tissue samples may be taken for biopsy.
- Biopsy: The removal of cells or tissues so they can be viewed under a microscope to check for signs of cancer.
- Virtual colonoscopy: A procedure that uses a series of x-rays called computed tomography to make a series of pictures of the colon. A computer puts the pictures together to create detailed images that may show polyps and anything else that seems unusual on the inside surface of the colon. This test is also called colonography or CT colonography.

Certain factors affect prognosis (chance of recovery) and treatment options.

The prognosis (chance of recovery) depends on the following:

- The stage of the cancer (whether the cancer is in the inner lining of the colon only, involves the whole colon, or has spread to other places in the body).
- Whether the cancer has blocked or created a hole in the colon.
- The blood levels of carcinoembryonic antigen (CEA; a substance in the blood that may be increased when cancer is present) before treatment begins.
- Whether the cancer has recurred.
- The patient's general health.

Treatment options depend on the following:

- The stage of the cancer.
- Whether the cancer has recurred.
- The patient's general health.

Colorectal cancer – High readability web page

Detailed Guide: Colon and Rectum Cancer

Can Colorectal Cancer Be Prevented?

Even though we do not know the exact cause of most colorectal cancer, it is possible to prevent many colorectal cancers. Following the American Cancer Society screening guidelines (see "Can Colorectal Polyps and Cancer Be Found Early?") can lower the number of cases of the disease by finding and removing polyps that could become cancerous, and can also lower the death rate from colorectal cancer by finding disease early when it is highly curable.

Prevention and early detection are possible because most colorectal cancers develop from adenomatous polyps. Polyps are noncancerous growths in the colon and rectum. Removing them can lower a person's cancer risk.

Diet and exercise:

People can lower their risk of developing colorectal cancer by managing the risk factors that they can control, such as diet and physical activity. It is important to eat plenty of fruits, vegetables, and whole grain foods and to limit intake of high-fat foods. Physical activity is another area that people can control. The American Cancer Society recommends at least 30 minutes of physical activity on 5 or more days of the week. If you participate in moderate or vigorous activity for 45 minutes on 5 or more days of the week you can lower your risk for breast and colorectal cancer even more. If you are overweight, you can ask your doctor about a weight loss plan that will work for you. For more information about diet and physical activity, refer to our document "American Cancer Society Guidelines for Nutrition and Cancer Prevention."

Vitamins and calcium:

Some studies suggest that taking a daily multivitamin containing folic acid, or folate, can lower colorectal cancer risk. Other studies suggest that increasing calcium intake via supplements or low-fat dairy products will lower risk. Some have suggested that vitamin D, which you can get from sun exposure or in a vitamin pill or in milk, can lower colorectal cancer risk. Indeed the rate of this cancer is lower in the Sunbelt states. Of course, excessive sun exposure can cause skin cancer and is not recommended as a way to lower colorectal cancer risk. Vitamin D may work better it you also take calcium.

Nonsteroidal anti-inflammatory drugs:

Many studies have found that people who regularly use aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen (Motrin, Advil) and naproxen (Aleve) have 20% to 50% lower risk of colorectal cancer and adenomatous polyps. Most of these studies, however, are based on observations of people who took these medications for reasons such as treatment of arthritis or prevention of heart attacks. Two recent studies have provided even stronger evidence regarding aspirin's ability to prevent the growth of polyps. The advantage of these recent studies is that people were randomly selected by the researchers to receive either aspirin or an inactive placebo. One study

included people who were previously treated for early stages of colorectal cancer, and the other study included people who previously had polyps removed.

But NSAIDs can cause serious or even life-threatening bleeding from stomach irritation, and currently available information suggests that the risks of serious bleeding outweigh the benefits of these medicines for the general public. For this reason, experts do not recommend NSAIDs as a cancer-prevention strategy for people at average risk of developing colorectal cancer. However, the value of these drugs for people at increased colorectal cancer risk is being actively studied. Celecoxib (Celebrex), has recently been approved by the FDA for reducing polyp formation in people with familial adenomatous polyposis. The advantage of this drug is it does not cause bleeding from the stomach.

Female hormones:

Hormone replacement therapy (HRT) in postmenopausal women may slightly reduce their risk of colorectal cancer. HRT also lowers the risk of developing osteoporosis, but it may increase the risk of heart disease, blood clots, and breast and uterine cancer. The decision to use HRT should be based on discussion of benefits and risks with your doctor.

There are other risk factors that can't be controlled, such as a strong family history of colorectal cancer. But even when people have a history of colorectal cancer in their family, they may be able to prevent the disease. For example, people with a family history of colorectal cancer may benefit from starting screening tests when they are younger and having them done more often than people without this risk factor.

Genetic tests can help determine which members of certain families have inherited a high risk for developing colorectal cancer. People with familial adenomatous polyposis (FAP) should start colonoscopy during their teens. Most doctors recommend they have their colon removed when they are in their twenties to prevent cancer from developing.

The risk for people with hereditary nonpolyposis colon cancer (HNPCC) is not as great as for those with FAP. Doctors recommend that people with HNPCC start colonoscopy screening during their twenties to remove any polyps and find any cancers at the earliest possible stage. But preventive removal of the colon is not usually suggested for people with HNPCC.

Ashkenazi Jews with the I1307K APC mutation have a slightly increased colorectal cancer risk, but do not develop these cancers when they are very young. For these reasons, most doctors recommend that they carefully follow the usual recommendations for colorectal cancer screening, but earlier or more frequent testing is usually not suggested.

Since some colorectal cancers can't be prevented, finding them early is the best way to improve the chance of a cure and reduce the number of deaths caused by this disease.

In addition to the screening recommendations for people at average colorectal cancer risk, the American Cancer Society has additional guidelines for people at moderate and high risk of colorectal cancer. These recommendations are described in the section. Can Colorectal Polyps and Cancer Be Found Early?" Ask your doctor how these guidelines might apply to you.

Prostate cancer – Low readability web page

The Role of Diet in the Prevention of Prostate Cancer

Can diet help prevent prostate cancer? There is a great deal of circumstantial evidence that suggests that diet is a factor in predisposing men to prostate cancer. However, we cannot be really sure that changing diet later in life can prevent prostate cancer developing. If it might make a difference, there are some easy changes one can make which may reduce the risk.

The information that follows reflects some of the current thinking about the benefits of a balanced diet. These benefits include building up resistance to prostate and other cancers. A balanced diet will also help to prevent other diseases such as heart disease.

Geography and Diet

People in Far Eastern countries such as China and Japan are less likely than Westerners to develop cancer. This difference may relate to the genetic differences between Easterners and Westerners. However, different risk levels may be related to other aspects of life in different parts of the world.

When people have migrated from Japan to the USA in the past, the rate of prostate cancer in their descendants has risen greatly. Since their genetic make-up is largely similar, as they have married within their own ethnic communities, the new risk of prostate cancer is likely to be related to environmental factors. Diet is an important one of these. Another pointer to reinforce this conclusion is that the levels of cancer in the East are rising. This coincides with changes in lifestyle - more people are living Western lifestyles and eating Western style foods.

The Role of Dietary Fat

Countries with a high level of heart disease such as the UK, often also have high levels of prostate cancer. These countries' cuisines often include high levels of fat, which has been shown to be linked to heart disease. It is thought that high levels of dietary fat may also create a higher risk of prostate cancer.

Most people in the UK would benefit from reducing their fat intake as part of a healthy diet, and in particular, reducing their saturated fat intake.

But how can you do this?

The best thing to do is avoid cooking methods that need fat such as frying and roasting. Instead, you could try baking, steaming, poaching, or grilling foods. You should also spread margarine or butter thinly on bread; and avoid adding fat to food when cooking.

Try to eat more fish and poultry. Avoid having red meat more than 2-3 times a week. Cut off any fat on the meat. Avoid high fat snacks such as crisps, biscuits and chocolate.

If you maintain a normal weight for your height then you are probably eating a healthy level of dietary fat.

Possible Protective Factors

Antioxidants are substances that may help in the prevention of prostate cancer. Antioxidants may help to prevent the damage caused by free radicals. Free radicals are harmful molecules that occur in the body. They can cause cell damage and can lead to diseases such as cancer and heart disease. Vitamin C is a well-known antioxidant. Others include vitamin E and selenium. Antioxidants are found in many fruits and vegetables.

Our knowledge on the role of these nutrients in prostate cancer is steadily growing – there has been some evidence showing a reduction in prostate cancer deaths when vitamin E was given to Finnish men who smoke, as a dietary supplement.

We are a long way from recommending routine Vitamin E supplementation for all men, as other studies have not shown the same benefits. There are also drawbacks. Some studies have shown more deaths from strokes in those groups taking vitamin supplements.

Selenium supplementation was also found to reduce prostate cancer in a small group of men, but again the evidence is not yet strong enough to recommend routine supplementation.
Rich sources of vitamin E are vegetable oils and spreads (but used sparingly to help maintain a low fat intake) and nuts. Selenium can be found in fish, lentils and Brazil nuts, all of which can be incorporated into a healthy diet.

A twelve-year study has just begun in the US to explore the links between Vitamin E and Selenium supplementation and a reduced risk of prostate cancer. In the mean time it is probably a good idea to ensure that your diet includes antioxidants.

How can I ensure that I get enough antioxidants in my diet?

The most important source for all antioxidants is fruit and vegetables – five or more portions a day is a good target. This could be fresh, frozen or tinned fruit and vegetables. A portion is:

- One piece of fruit, or a slice of very large fruit (e.g., pineapple);
- Two pieces of a small fruit,
- One cupful of grapes or raspberries,
- Two tablespoons of vegetables, raw, cooked or frozen.

Count up what you are eating now. If it is less than five portions a day add in more and gradually try to build it up. Sorry, but potatoes do not count!

Tomatoes have had a good press recently in relation to the prevention of prostate cancer. This is because they contain not only vitamins C and E, but also another antioxidant called lycopene. Lycopene is the compound that gives tomatoes their red colour. There is some evidence from the USA that men who eat the most processed tomatoes, such as found on pizza toppings and pasta sauces, are less likely to develop prostate cancer. It may therefore be a good idea, whilst awaiting further evidence, to try and increase your tomato consumption to one portion a day – this may take any form, either fresh ripe tomatoes or processed tomatoes e.g. tomato juice, soup, tinned chopped or plum tomatoes. Processed tomato products, such as ketchup, are particularly rich in a form of

lycopene that is easier for the body to metabolise. Tomatoes can also contribute to your five portions of fruit or vegetables a day!

Should you take supplements of lycopene? The research on lycopene is at very early stages and we may find out that there are other nutrients in tomatoes that are also beneficial. Therefore it is recommended that you try to increase your overall intake of fruit and vegetables including tomatoes rather than take supplements.

Soya

Soya beans have also been in the news recently due to their anti-cancer potential. Soya protein can be found in many foods including alternatives to meat such as tofu and non-dairy products such as soya yogurts and soya milk. The beans from which the protein is extracted contain compounds called isoflavanoids. The most well known of these is genistein. This has been shown to inhibit prostate cancer cell growth in the laboratory. Unfortunately, there is no evidence that this happens in humans, but overall inclusion of soya into the diet will not do any harm. It can easily be added to your diet as a glass of soya milk or soya yogurt. You could eat tofu as an alternative to red meat at an evening meal. Any increase from a minimal intake will help widen your food choices.

Overall the message is to:

- Maintain a normal weight for your height;
- Avoid fatty foods and try to decrease your fat intake;
- Eat red meat and processed meat in moderation;
- Include at least five portions of fruit and vegetables per day, including a regular intake of tomatoes;
- Perhaps include soya products within your diet as an occasional food product;
- Do not take high doses of vitamin supplements;
- Drink alcohol in moderation. A maximum of three units per day for men (one unit = a pub measure of spirits, a small glass of wine or half a pint of standard beer).

Prostate cancer – High readability web page

Diagnosing prostate cancer

Determining whether you have prostate cancer generally involves a series of tests and exams. Before starting the testing process, your physician may ask you questions about your medical history, your family history of cancer and any symptoms you may be having, particularly problems with urination. Then, your doctor will most likely proceed to one or more of the tests described below.

Digital Rectal Exam (DRE)
Because the prostate lies in front of the rectum, your physician can feel the prostate by inserting a gloved, lubricated finger into the rectum. This simple procedure is called a digital rectal examination (DRE). It allows your physician to determine whether the prostate is enlarged or has lumps or other types of abnormal texture.

Prostate-Specific Antigen (PSA) test
Used in addition to the DRE, a PSA test increases the likelihood of prostate cancer detection. PSA is the abbreviation for prostate-specific antigen, a substance produced by the prostate cells. A PSA test measures the level of PSA in the bloodstream and is reported as nanograms per milliliter, or ng/mL. Very little PSA escapes from a healthy prostate into the bloodstream, but certain prostate conditions can cause larger amounts of PSA to leak into the blood.

Two possible causes of a high PSA level are:
1. a benign noncancerous enlargement of the prostate called benign prostatic hyperplasia (BPH)
2. prostate cancer

A high level of PSA in the bloodstream is a warning sign that prostate cancer may be present. But since other kinds of prostate disease can also cause high PSA levels, PSA testing by itself cannot confirm the presence of prostate cancer. A high PSA level only indicates the possibility of prostate cancer and the need for additional evaluation by your physician. Conversely, a low PSA level does not always mean that prostate cancer is not present.

According to the American Cancer Society, men aged 50 and older, and those over the age of 45 who are in high-risk groups, such as African-American men and men with a family history of prostate cancer, should have a prostate-specific antigen (PSA) blood test and digital rectal exam (DRE) once every year. Any man who develops persistent urinary symptoms should contact his physician.

Transrectal Ultrasound (TRUS)
Transrectal Ultrasound (TRUS) is the use of soundwaves to create an image of the prostate. As the waves bounce off the prostate, they create a pattern that is converted into a picture by a computer. TRUS is used to detect abnormal prostate growth and to guide a biopsy of the abnormal prostate area.

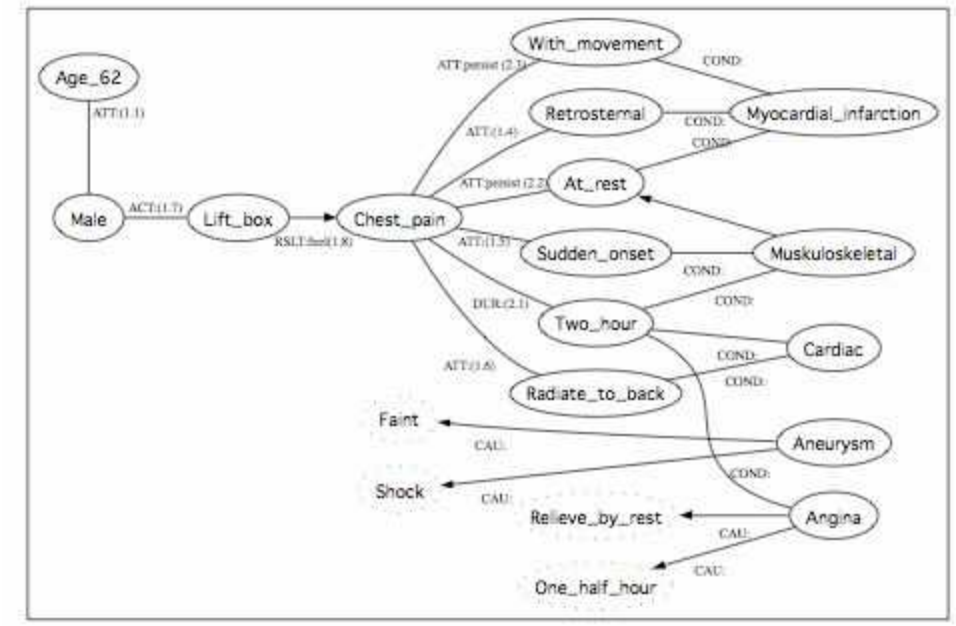
Biopsy
A biopsy is the removal of a sample of tissue, which is then examined under a microscope to check for cancerous changes. Only a biopsy can definitely confirm prostate cancer.

Typically, the physician takes multiple tissue samples for biopsy. Keep in mind that it is still possible to have cancer, even if the biopsy is negative. This is because, even though multiple samples are taken during a biopsy, it can still miss some cancers.

If the biopsy is taken and prostate cancer is found, the tumor is graded in the medical lab. The grade estimates how aggressive a prostate cancer is; that is, how fast it is growing and the likelihood of its spreading. Sometimes you will hear the grade referred to as the Gleason grade.

Once diagnosis is made, prostate cancer is categorized into stages based on the size and spread of the disease. Learn more about grading and staging of prostate cancer.

Appendix B: Example of Network Representation



Source:

Unpublished

Appendix C: Example of Propositional Analysis

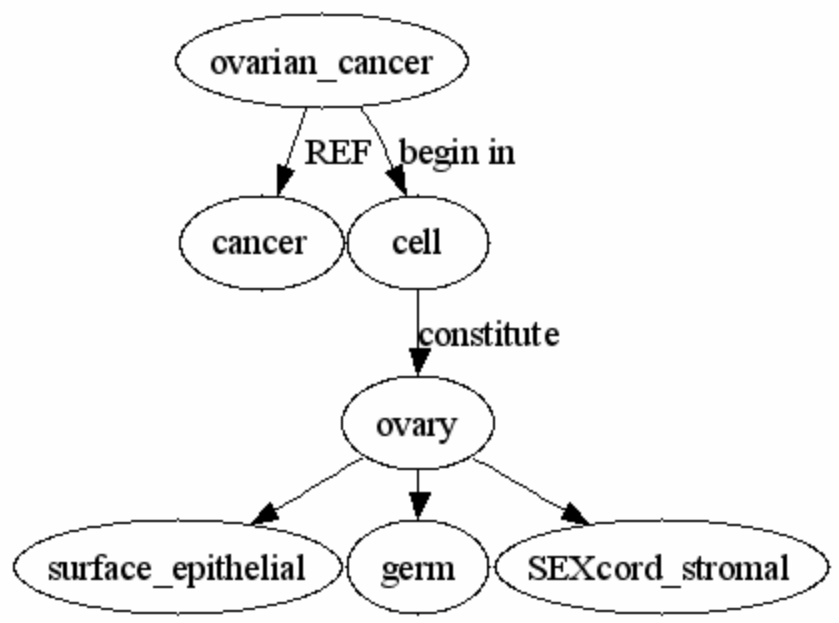
Text:

Ovarian cancer is cancer that begins in the cells that constitute the ovaries, including surface epithelial cells, germ cells, and the sex cord-stromal cells.

List of ordered propositions:

P1 REF OVARIAN-CANCER CANCER
P2 BEGINS OVARIAN-CANCER CELLS
P3 CONSTITUTE CELLS OVARIES
P4 INCLUDE P3 SURFACE-EPITHELIAL-CELLS
P5 INCLUDE P3 GERM-CELLS
P6 INCLUDE P3 SEX-CORD-STROMAL CELLS

Semantic representation:



Appendix D: Samples of Web Page Propositional Analysis

Web page 1: Breast cancer; Low readability

P1.S1. A lump is detected, which is usually single, firm, and most often painless.

P1 (DETECT $ LUMP)
P2 (AMOUNT-OF LUMP SINGLE)
P3 (MOD LUMP FIRM)
P4 (MOD LUMP PAINLESS)
P5 (MOD PAINLESS OFTEN)
P6 (AMOUNT-OF OFTEN MOST)
P7 (REF (P3,P4) P1)

P1.S2. A portion of the skin on the breast or underarm swells and has an unusual appearance.

P1 (ON SKIN BREAST)
P2 (ON SKIN UNDERARM)
P3 (AMOUNT-OF P1 PORTION)
P4 (AMOUNT-OF P2 PORTION)
P5 (OR P1 P2)
P6 (SWELLS P1)
P7 (SWELLS P2)
P8 (POSSESS P5 APPEARANCE)
P9 (MOD APPEARANCE UNUSUAL)
P10 (AND P6 P7)

P1.S3. Veins on the skin surface become more prominent on one breast.

P1 (BECOME VEINS PROMINENT)
P2 (ON VEINS SKIN-SURFACE)
P3 (DEGREE-OF PROMINENT MORE)
P4 (ON VEINS BREAST)
P5 (NUMBER-OF BREAST ONE)

P1.S4. The breast nipple becomes inverted, develops a rash, changes in skin texture, or has a discharge other than breast milk.

P1 (BECOME BREAST-NIPPLE INVERT)
P2 (DEVELOP BREAST-NIPPLE RASH)
P3 (CHANGE BREAST-NIPPLE SKIN-TEXTURE)
P4 (HAVE BREAST-NIPPLE DISCHARGE)
P5 (OTHER-THAN DISCHARGE BREAST-MILK)
P6 (OR BREAST-NIPPLE (P1,P2,P3,P4))

P1.S5. A depression is found in an area of the breast surface.

P1 (FOUND DEPRESSION BREAST-SURFACE)
P2 (IN AREA BREAST-SURFACE)

P2.S1. Women's breasts can develop some degree of lumpiness, but only a small percentage of lumps are malignant.

P1 (DEVELOP BREAST LUMPINESS)
P2 (POSSESS BREAST WOMAN)
P3 (DEGREE-OF LUMPINESS SOME)
P4 (IS LUMP MALIGNANT)
P5 (AMOUNT-OF P4 SMALL-PERCENTAGE)
P6 (MOD SMALL-PERCENTAGE ONLY)
P7 (BUT P1 P4)

P3.S1. While a history of breast cancer in the family may lead to increased risk, most breast cancers are diagnosed in women with no family history.

P1 (LEAD BREAST-CANCER RISK)
P2 (MOD P1 POSSIBLE)
P3 (IN BREAST-CANCER FAMILY)
P4 (AMOUNT-OF RISK INCREASE)
P5 (OF HISTORY BREAST-CANCER)
P6 (DIAGNOSE $ BREAST-CANCER)
P7 (WITH WOMEN FAMILY-HISTORY)
P8 (NEGATE P6)
P9 (AMOUNT-OF BREAST-CANCER MOST)

P3.S2. If you have a family history of breast cancer, this should be discussed with your doctor.

P4 (IF P2)
P1 (POSSESS PERSON FAMILY-HISTORY)
P2 (OF FAMILY-HISTORY BREAST-CANCER)
P3 (DISCUSS-WITH P1 DOCTOR)

P4.S1. Every two minutes a woman is diagnosed with breast cancer.

P1 (DIAGNOSE-WITH WOMAN BREAST-CANCER)
P2 (RATE-OF P1 MINUTES)
P3 (AMOUNT-OF MINUTES EVERY-TWO)

P4.S2. This year more than 211,000 new cases of breast cancer are expected in the United States.

P1 (EXPECT BREAST-CANCER-CASES)
P2 (MOD BREAST-CANCER-CASES NEW)
P3 (AMOUNT-OF P1 MORE-THAN-211,000)
P4 (IN P1 UNITED-STATED)
P5 (THIS-YEAR P1)

P4.S3. One woman in eight who lives to age 85 will develop breast cancer during her lifetime.

P1 (DEVELOP WOMAN BREAST-CANCER)
P2 (LIVES-TO WOMAN AGE-85)
P3 (NUMBER-OF P1 ONE-IN-EIGHT)
P4 (DURING WOMAN LIFETIME)

P4.S4. Breast cancer is the leading cause of death in women between the ages of 40 and 55.

P1 (CAUSE BREAST-CANCER DEATH)
P2 (MOD DEATH LEADING)
P3 (IN P1 WOMAN)
P4 (BETWEEN P1 AGE-40-55)

P4.S5. 1,600 men are expected to be diagnosed with breast cancer this year and 400 are predicted to die.

P1 (EXPECT-TO-BE-DIAGNOSED MAN BREAST-CANCER)
P2 (THIS-YEAR P1)
P3 (PREDICT P1 DIE)
P4 (NUMBER-OF DIE 400)
P5 (NUMBER-OF EXPECT-TO-BE-DIAGNOSED 1600)

P4.S6.1. Seventy percent of all breast cancers are found through breast self-exams.

P1 (FOUND-THROUGH BREAST-CANCER BREAST-SELF-EXAMS)
P2 (AMOUNT-OF BREAST-CANCER ALL)
P3 (NUMBER-OF BREAST-CANCER SEVENTY-PERCENT)

P4.S6.2. Not all lumps are detectable by touch.

P1 (DETECTABLE-BY LUMP TOUCH)
P2 (AMOUNT-OF LUMP ALL)
P3 (NEGATE P2)

P4.S6.3. We recommend regular mammograms and monthly breast self-exams.

P1 (RECOMMEND AUTHOR MAMMOGRAM)
P2 (MOD MAMMOGRAM REGULAR)
P3 (RECOMMEND AUTHOR BREAST-SELF-EXAM)
P4 (MOD BREAST-SELF-EXAM MONTHLY)
P5 (AND (P1,P3))

P4.S7.1. Eight out of ten breast lumps are not cancerous.

P1 (IS BREAST-LUMP CANCEROUS)
P2 (NUMBER-OF BREAST-LUMP EIGHT-OUT-OF-TEN)
P3 (NEGATE P1)

P4.S7.2. If you find a lump, don't panic-call your doctor for an appointment.

P1 (FIND PERSON LUMP)
P2 (IF P1 PANIC)
P3 (NEGATE PANIC)
P4 (CALL PERSON DOCTOR)
P5 (FOR P4 APPOINTMENT)

P4.S8. Mammography is a low-dose X-ray examination that can detect breast cancer up to two years before it is large enough to be felt.

P1 (IS-A MAMMOGRAPHY X-RAY-EXAMINATION)
P2 (DEGREE-OF X-RAY-EXAMINATION LOW-DOSE)
P3 (DETECT MAMMOGRAPHY BREAST-CANCER)
P4 (DURATION P3 UP-TO-TWO-YEARS)
P5 (TO-FEEL $ BREAST-CANCER)
P6 (DEGREE-OF P5 LARGE-ENOUGH)
P7 (BEFORE P3 P5)

P4.S9.1. When breast cancer is found early, the five-year survival rate is 96%.

P1 (WHEN P2)
P2 (FIND $ BREAST-CANCER)
P3 (MOD FIND EARLY)
P4 (OF RATE BREAST-CANCER)
P5 (MOD RATE SURVIVAL)
P6 (NUMBER-OF RATE 0.96)

P4.S9.2. This is good news!

P1 (IS $ NEWS)
P2 (MOD NEWS GOOD)

P4.S9.3. Over 2 million breast cancer survivors are alive in America today.

P1 (ALIVE SURVIVOR)
P2 (MOD SURVIVOR BREAST-CANCER)
P3 (NUMBER-OF P1 OVER-TWO-MILLION)
P4 (IN P1 AMERICA-TODAY)

P5.S1. An Early Breast Cancer Detection Plan should include:

P1 (SHOULD DETECTION-PLAN INCLUDE)
P2 (MOD DETECTION-PLAN EARLY-BREAST-CANCER)

P6.S1. Clinical breast examinations every three years from ages 20-39, then every year thereafter.

P1 (DURATION-OF CLINICAL-BREAST-EXAMINATIONS EVERY-THREE-YEARS)
P2 (FROM P1 AGES-20-39)
P3 (DURATION-OF CLINICAL-BREAST-EXAMINATIONS EVERY-YEAR-THEREAFTER)
P4 (THEN P1 P3)

P6.S2. Monthly breast self-examinations beginning at age 20.

P1 (BEGIN $ BREAST-SELF-EXAMINATIONS)
P2 (DURATION-OF P1 MONTHLY)
P3 (AT P1 AGE-20)

P6.S3. Look for any changes in your breasts.

P1 (LOOK $ CHANGE)
P2 (IN P1 BREAST)

P6.S4. Baseline mammogram by the age of 40.

P1 ($ BASELINE-MAMMOGRAM)
P2 (BY P1 AGE-40)

P6.S5. Mammogram every one to two years for women 40-49, depending on previous findings.

P1 ($ WOMAN-40-49 MAMMOGRAM)
P2 (DURATION-OF P1 EVERY-TWO-THREE-YEARS)
P3 (DEPENDING-ON P1 PREVIOUS-FINDING)

P6.S6. Mammogram every year for women 50 and older.

P1 ($ MAMMOGRAM)
P2 (DURATION-OF P1 EVERY-YEAR)
P3 (FOR P1 WOMAN-50-AND-OLDER)

P6.S7. A personal calendar to record your self-exams, mammograms, and doctor appointments.

P1 (RECORD PERSONAL-CALENDAR SELF-EXAMS)
P2 (RECORD PERSONAL-CALENDAR MAMMOGRAMS)

P3 (RECORD PERSONAL-CALENDAR DOCTOR-APPOINTMENTS)
P4 (AND (P1,P2,P3))

P6.S8. A low-fat diet, regular exercise, and no smoking or drinking.

P1 (MOD DIET LOW-FAT)
P2 (MOD EXERCISE REGULAR)
P3 (OR SMOKING DRINKING)
P4 (NEGATE P3)
P5 (AND (P1,P2,P4))

P7.S1. IN THE SHOWER Fingers flat, move gently over every part of each breast.

P1 (MOVE-OVER FINGER EACH-BREAST)
P2 (MOD EACH-BREAST EVERY-PART)
P3 (MOD FLAT FINGER)
P4 (IN P1 SHOWER)

P7.S2. Use your right hand to examine left breast, left hand for right breast.

P1 (EXAMINE RIGHT-HAND LEFT-BREAST)
P2 (EXAMINE LEFT-HAND RIGHT-BREAST)
P3 (USE WOMAN RIGHT-HAND)
P4 (USE WOMAN LEFT-HAND)

P8.S1. Check for any lump, hard knot or thickening.

P1 (FOR CHECK LUMP)
P2 (FOR CHECK HARD-KNOT)
P3 (FOR CHECK THICKENING)
P4 (MOD LUMP ANY)
P5 (MOD HARD-KNOT ANY)
P6 (MOD THICKENING ANY)
P7 (OR (P1,P2,P3))

P8.S2. Carefully observe any changes in your breasts.

P1 (OBSERVE $ CHANGE)
P2 (MOD CHANGES ANY)
P3 (IN P1 BREAST)
P4 (MOD OBSERVE CAREFULLY)

P9.S1. BEFORE A MIRROR Inspect your breasts with arms at your sides.

P1 (INSPECT $ BREAST)
P2 (WITH P1 ARM-AT-SIDE)
P3 (BEFORE-A-MIRROR P1)

P9.S2. Next, raise your arms high overhead.

P1 (RAISE $ ARM)
P2 (DEGREE-OF ARM HIGH-OVERHEAD)
P3 (NEXT P1)

P10.S1. Look for any changes in contour of each breast, a swelling, a dimpling of skin or changes in the nipple.

P1 (LOOK-FOR $ CHANGE)
P2 (OF CHANGE EACH-BREAST)
P3 (MOD CHANGE ANY)
P4 (LOOK-FOR $ SWELLING)
P5 (LOOK-FOR $ DIMPLING-OF-SKIN)
P6 (IN CHANGE NIPPLE)
P7 (OR (P1,P4,P5))

P10.S2. Then rest palm on hips and press firmly to flex your chest muscles.

P1 (REST $ PALM)
P2 (ON P1 HIP)
P3 (FLEX $ CHEST-MUSCLES)
P4 (PRESS-FIRMLY-TO P3)
P5 (AND (P1,P3))
P6 (THEN P5)

P10.S3. Left and right breasts will not exactly match - few women's breasts do.

P1 (WILL LEFT-BREAST MATCH)
P2 (WILL RIGHT-BREAST MATCH)
P3 (AND (P1,P2))
P4 (NEGATE P3)
P5 (MATCH $ WOMAN-BREAST)
P6 (AMOUNT-OF P5 FEW)

P11.S1. LYING DOWN Place pillow under right shoulder, right arm behind your head.

P1 (LYING-DOWN P2)
P2 (PLACE $ PILLOW)
P3 (UNDER P2 RIGHT-SHOULDER)
P4 (BEHIND RIGHT-ARM HEAD)

P12.S2. With fingers of left hand flat, press right breast gently in small circular motions, moving vertically or in a circular pattern covering the entire breast.

P1 (PRESS LEFT-HAND RIGHT-BREAST)
P2 (MOD FINGERS FLAT)
P3 (WITH-FINGERS P1)
P4 (MOD RIGHT-BREAST GENTLY)
P5 (IN P1 CIRCULAR-MOTIONS)
P6 (MOD CIRCULAR-MOTIONS SMALL)
P7 (MOVE LEFT-HAND VERTICALLY)
P8 (MOVE LEFT-HAND IN-CIRCULAR-PATTERN)
P9 (OR (P7,P8))
P10 (COVERING P9 ENTIRE-BREAST)

P12.S3. Use light, medium and firm pressure.

P1 (USE $ PRESSURE)
P2 (MOD PRESSURE LIGHT)
P3 (MOD PRESSURE MEDIUM)
P4 (MOD PRESSURE FIRM)
P5 (AND (P2,P3,P4))

P12.S4. Squeeze nipple; check for discharge and lumps.

P1 (SQUEEZE $ NIPPLE)
P2 (CHECK-FOR $ DISCHARGE)
P3 (CHECK-FOR $ LUMPS)
P4 (AND (P2,P3))

P12.S5. Repeat these steps for your left breast.

P1 (REPEAT $ STEP)
P2 (FOR P1 LEFT-BREAST)

Web page 2 : Breast cancer; High readability

P1.S1. Physicians and women have long awaited a better early detection tool for breast cancer than mammograms.

P1 (AWAIT PHYSICIAN TOOL)
P2 (AWAIT WOMAN TOOL)
P3 (AND (P1,P2))
P4 (IS DETECTION TOOL)
P5 (FOR DETECTION BREAST-CANCER)
P6 (IS DETECTION MAMMOGRAM)
P7 (BETTER-THAN P5 P3)
P8 (DURATION-OF AWAIT LONG)

P1.S2. And a recent study suggests that a new option, magnetic resonance imaging (MRI), can be added to the screening arsenal, though it's only recommended to women at high genetic risk for breast cancer.

P1 (AWAIT PHYSICIAN TOOL)
P2 (AWAIT WOMAN TOOL)
P3 (AND (P1,P2))
P4 (IS DETECTION TOOL)
P5 (FOR DETECTION BREAST-CANCER)
P6 (IS DETECTION MAMMOGRAM)
P7 (BETTER-THAN P5 P3)
P8 (DURATION-OF AWAIT LONG)
P9 (ALTHOUGH P5 P8)

P2.S1. Breast cancer screening is a top concern for women who have an inherited abnormality in one of two genes, known as the BRCA1 and BRCA2 genes.

P1 (IS-A BREAST-CANCER-SCREENING CONCERN)
P2 (MOD CONCERN TOP)
P3 (FOR P1 WOMAN)
P4 (INHERIT WOMAN ABNORMALITY)
P5 (INHERIT P4 GENE)
P6 (OF ONE TWO)
P7 (NUMBER-OF GENE TWO)
P8 (REF BRCA1 GENE1)
P9 (REF BRCA2 GENE2)
P10 (POSSESS WOMAN P4)

P2.S2. Yet few studies have examined what screening approach is best for these high-risk women.

P1 (EXAMINE STUDY SCREENING-APPROACH)
P2 (BEST SCREENING-APPROACH BEST)
P3 (FOR SCREENING-APPROACH WOMAN)
P4 (MOD WOMAN HIGH-RISK)
P5 (AMOUNT-OF STUDY FEW)
P6 (YET P6)

P2.S3. BRCA1 and BRCA2 mutation carriers are identified though blood tests that are given to certain women with a strong family history of breast cancer.

P1 (MOD MUTATION-CARRIER BRCA1)
P2 (MOD MUTATION-CARRIER BRCA2)
P3 (IDENTIFY P1 P2)
P4 (THOUGH P3 BLOOD-TEST)
P5 (GIVE-TO BLOOD-TEST WOMAN)
P6 (MOD WOMAN CERTAIN)
P7 (MOD FAMILY-HISTORY STRONG)
P8 (OF FAMILY-HISTORY BREAST-CANCER)
P9 (WITH WOMAN P8)

P2.S4. Women of Ashkenazi Jewish descent are at particular risk: The mutations occur in about 2.5 percent of these women, compared to 1 percent of the general population.

P1 (AT WOMAN RISK)
P2 (LABEL WOMAN ASHKENAZI-JEW)
P3 (MOD RISK PARTICULAR)
P4 (OCCUR-IN MUTATION P2)
P5 (NUMBER-OF WOMAN 0.025)
P6 (OCCUR-IN MUTATION GENERAL-POPULATION)
P7 (NUMBER-OF GENERAL-POPULATION 0.01)
P8 (COMPARE-TO P4 P6)

P3.S1. A study published in the September 15th issue of the Journal of the American Medical Association looked at the usefulness of MRI, mammograms, ultrasound and clinical breast exams, which are breast exams performed by a healthcare professional in screening healthy women with the BRCA1 and BRCA2 gene mutations for breast cancer.

P1 (LOOK STUDY USEFULNESS)
P2 (OF USEFULNESS MRI)
P3 (OF USEFULNESS MAMMOGRAM)
P4 (OF USEFULNESS ULTRASOUND)
P5 (OF USEFULNESS CLINICAL-BREAST-EXAM)

P6 (PERFORM HEALTHCARE-PROFESSIONAL CLINICAL-BREAST-EXAM)
P7 (LABEL ISSUE JAMA)
P8 (MOD ISSUE SEPT15)
P9 (PUBLISH-IN STUDY P8)
P10 (PURPOSE-OF P2,P3,P4,P5 SCREEN)
P11 (MOD WOMAN HEALTHY)
P12 (SCREEN P11)
P13 (POSSESS WOMAN GENE-MUTATION)
P14 (REF GENE-MUTATION BRCA1)
P15 (REF GENE-MUTATION BRCA2)
P16 (FOR P13 BREAST-CANCER)

P3.S2.The researchers found that MRI, in which magnets and radiowaves create an image of a body part, was a helpful addition to the surveillance programs for these women.

P1 (FIND RESEARCHER P2)
P2 (ADDITION-TO MRI SURVEILLANCE-PROGRAM)
P3 (MOD ADDITION-TO HELPFUL)
P4 (FOR MRI WOMAN)
P5 (CREATE MAGNET IMAGE)
P6 (CREATE RADIOWAVE IMAGE)
P7 (AND P5 P6)
P8 (OF P7 BODY-PART)
P9 (REF MRI P8)

P4.S1. Other risk factors for breast cancer include age, family history, use of hormone replacement therapy, radiation exposure, early onset of the menstrual period and late menopause.

P1 (FOR RISK-FACTOR BREAST-CANCER)
P2 (INCLUDE RISK-FACTOR AGE)
P3 (INCLUDE RISK-FACTOR FAMILY-HISTORY)
P4 (INCLUDE RISK-FACTOR HRT)
P5 (INCLUDE RISK-FACTOR RADIATION-EXPOSURE)
P6 (INCLUDE RISK-FACTOR MENSTRUAL-PERIOD)
P7 (OF EARLY-ONSET MENSTRUAL-PERIOD)
P8 (INCLUDE RISK-FACTOR LATE-MENOPAUSE)
P9 (AND (P2,P3,P4,P5,P6,P8))

P4.S2. However, the JAMA study only examined women with one of the BRCA mutations.

P1 (EXAMINE STUDY WOMAN)
P2 (WITH WOMAN MUTATION)

P3 (REF MUTATION BRCA)
P4 (NUMBER-OF MUTATION ONE)
P5 (REF STUDY JAMA)
P6 (ONLY P1)

P4.S3. Breast cancers related to BRCA1 or 2 mutations make up about 5 percent of all breast cancers.

P1 (RELATE-TO BREAST-CANCER BRCA1)
P2 (RELATE-TO BREAST-CANCER TWO-MUTATIONS)
P3 (OR P1 P2)
P4 (MAKE-UP P3 0.05)
P5 (OF 0.05 100%-BREAST-CANCER)

P5.S1. Below, study author Sandra Messner, MD, the medical coordinator of clinical breast services in preventive oncology at the Toronto Sunnybrook Regional Cancer Center, discusses the best screening options for these high-risk women to help make sure any breast cancer they develop is detected as early as possible.

P1 (DISCUSS AUTHOR SCREENING-OPTION)
P2 (FOR SCREENING-OPTION WOMAN)
P3 (MOD WOMAN HIGH-RISK)
P4 (HELP-MAKE-SURE P1 DETECT)
P5 (DETECT SCREENING-OPTION BREAST-CANCER)
P6 (MOD DETECT AS-EARLY-AS-POSSIBLE)
P7 (REF AUTHOR SANDRA-MESSNER-M.D.)
P8 (REF AUTHOR MEDICAL-COORDINATOR)
P9 (OF MEDICAL-COORDINATOR CLINICAL-BREAST-SERVICE)

P10 (PART-OF CLINICAL-BREAST-SERVICE PREVENTATIVE-ONCOLOGY)
P11 (LOCATE-IN CLINICAL-BREAST-SERVICE SUNNYBROOK-REGIONAL-CANCER-CENTRE)

P6.S1. There are two large genes that have been identified in all women called BRCA1 and BRCA2.

P1 (IDENTIFY $ GENE)
P2 (NUMBER-OF GENE TWO)
P3 (MOD GENE LARGE)
P4 (IN P1 WOMAN)
P5 (NUMBER-OF WOMAN ALL)
P6 (REF GENE BRCA1)
P7 (REF GENE BRCA2)

P6.S2. They function, we think, as tumor-suppressor genes, so they keep cancers from developing.

P1 (FUNCTION-AS GENE TUMOR-SUPRESSOR)
P2 (REF GENE TUMOR-SUPRESSOR)
P3 (KEEP GENE CANCER)
P4 (FROM P3 DEVELOPING)
P5 (SO P1 P4)

P6.S3. If they are abnormal or what we call mutated, then your risk of cancer is increased, and the risk of breast cancer and ovarian cancer particularly is affected.

P1 (IF P4)
P2 (IS GENE ABNORMAL)
P3 (IS GENE MUTATION)
P4 (OR P2 P3)
P5 (INCREASE P4 RISK)
P6 (OF RISK CANCER)
P7 (OF RISK BREAST-CANCER)
P8 (OF RISK OVARIAN-CANCER)
P9 (AND P7 P8)
P10 (AFFECT P4 P9)
P11 (DEGREE-OF P10 PARTICULARLY)

P6.S4. If a woman carries an abnormality in one of those two genes, she is thought to have up to an 85 percent lifetime risk of breast cancer.

P1 (IF P2 P5)
P2 (CARRY WOMAN ABNORMALITY)
P3 (IN P1 GENE)
P4 (NUMBER-OF GENE ONE)
P5 (POSSESS WOMAN RISK)
P6 (DURATION-OF RISK LIFETIME)
P7 (NUMBER-OF RISK UP-TO-85%)

P6.S5. The average woman's lifetime risk of developing breast cancer is about 11 percent.

P1 (DEVELOP RISK BREAST-CANCER)
P2 (DURATION-OF RISK LIFETIME)
P3 (POSSESS WOMAN P1)
P4 (MOD WOMAN AVERAGE)
P5 (AMOUNT-OF RISK ABOUT-11%)

P6.S6. The mutations also increase the risk of ovarian cancer, more so with BRCA1 than with BRCA2.

P1 (INCREASE MUTATION RISK)
P2 (OF RISK OVARIAN-CANCER)
P3 (REF MUTATION-1 BRCA1)
P4 (REF MUTATION-2 BRCA2)
P5 (THAN P3 P4)
P6 (MORE-SO-WITH P5)

P6.S7. The BRCA2 mutations may also increase risk of other cancers, such as melanoma and pancreatic cancer.

P1 (INCREASE MUTATION RISK)
P2 (MOD P1 POSSIBLE)
P3 (OF RISK CANCER)
P4 (OTHER-THAN CANCER BREAST-CANCER)
P5 (OTHER-THAN CANCER OVARIAN-CANCER)
P6 (REF MUTATION BRCA2)
P7 (EXAMPLE-OF OTHER-CANCER MELANOMA)
P8 (EXAMPLE-OF OTHER-CANCER PANCREATIC-CANCER)

P7.S1. It is possible to carry mutations in both genes, but it is very rare.

P1 (POSSESS $ MUTATION)
P2 (IN MUTATION GENE)
P3 (NUMBER-OF GENE BOTH)
P4 (MOD P1 RARE)
P5 (MOE RARE VERY)
P6 (MOD P1 POSSIBLE)
P7 (BUT P6 P4)

P7.S2. There is no conclusive data on risk in this situation, but most experts suggest that the risk is not any higher with both genes than only one.

P1 (IS CONCLUSIVE-DATA)
P2 (FOR CONCLUSIVE-DATA RISK)
P3 (NEGATE P1)
P4 (REF P1 CARRY-MUTATION-IN-BOTH-GENES)
P5 (SUGGEST EXPERT P6)
P6 (AMOUNT-OF EXPERT MOST)
P7 (IS RISK HIGHER)
P8 (NEGATE P7)
P9 (NUMBER-OF GENE BOTH)
P10 (NUMBER-OF GENE ONE)
P11 (THAN P9 P10)

P8.S1. There hasn't really been any kind of standard approach, but women who carry mutations in these genes have several options.

P1 (IS STANDARD-APPROACH)
P2 (NEGATE P1)
P3 (CARRY WOMAN MUTATION)
P4 (IN MUTATION GENE)
P5 (POSSESS P3 OPTION)
P6 (NUMBER-OF OPTION SEVERAL)
P7 (BUT P2 P5)

P8.S2. The most drastic is to have their breast removed with a preventive mastectomy.

P1 (REMOVE WOMAN BREAST)
P2 (IS OPTION P1)
P3 (IS P1 DRASTIC)
P4 (DEGREE-OF DRASTIC MOST)
P5 (HAVE-TO WOMAN P1)
P6 (WITH P1 PREVENTATIVE-MASTECTOMY)

P8.S3. This reduces the risk of developing breast cancer by about 90 to 95 percent.

P1 (REDUCE PREVENTATIVE-MASTECTOMY RISK)
P2 (DEVELOP RISK BREAST-CANCER)
P3 (OF P1 P2)
P4 (AMOUNT-OF REDUCE 90-95%)

P8.S4. It's a pretty drastic thing, however.

P1 (HOWEVER P2)
P2 (MOD PREVENTATIVE-MASTECTOMY DRASTIC)
P3 (MOD DRASTIC PRETTY)

P9.S1. We also talk to these women about the potential role of tamoxifen, which works by blocking estrogen receptors in women with hormone-positive breast cancer, to reduce their risk of breast cancer.

P1 (TALK-TO $ WOMAN)
P2 (ABOUT P1 TAMOXIFEN)
P3 (OF ROLE TAMOXIFEN)
P4 (MOD ROLE POTENTIAL)
P5 (BLOCK TAMOXIFEN ESTROGEN-RECEPTOR)
P6 (IN ESTROGEN-RECEPTOR WOMAN)
P7 (WITH WOMAN HORMONE-POSITIVE-BC)

P8 (WORKS-BY P5)
P9 (REDUCE P3 RISK)
P10 (OF RISK BREAST-CANCER)
P11 (POSSESS WOMAN P10)

P9.S2. There have been some papers that suggest it's also beneficial in mutation carriers.

P1 (SUGGEST PAPER TAMOXIFEN)
P2 (IS TAMOXIFEN BENEFICIAL)
P3 (IN BENEFICIAL MUTATION-CARRIER)
P4 (AMOUNT-OF PAPER SOME)

P9.S3. But there have been a lot of questions about whether it's as beneficial in BRCA1 carriers as it is in BRCA2 carriers, because tamoxifen works by blocking estrogen receptors, and three-quarters of the cancers that are diagnosed in BRCA1 carriers are estrogen-receptor negative.

P1 (BUT P2)
P2 (ABOUT QUESTION TAMOXIFEN)
P3 (AMOUNT-OF QUESTION A-LOT)
P4 (FOR BENEFICIAL BRCA1-CARRIER)
P5 (FOR BENEFICIAL BRCA2-CARRIER)
P6 (AS-BENEFICIAL-AS BRCA1 BRCA2)
P7 (WORKS-BY TAMOXIFEN BLOCK)
P8 (BLOCK TAMOXIFEN ESTROGEN-RECEPTOR)
P9 (DIAGNOSE $ CANCER)
P10 (IN P9 BRCA1-CARRIER)
P11 (AMOUNT-OF P9 THREE-QUARTER)
P12 (IS P9 ESTROGEN-RECEPTOR-NEGATIVE)
P13 (AND P6 P8)
P14 (BECAUSE P2 P13)

P9.S4. With BRCA2 mutations, about 75 percent of cancers are estrogen-receptor positive.

P1 (IS CANCER ESTROGEN-RECEPTOR-POSITIVE)
P2 (AMOUNT-OF CANCER 0.75)
P3 (MOD P2 ABOUT)
P4 (WITH P1 BRCA1-MUTATION)

P10.S1. However, people don't like to take drugs when they are healthy.

P1 (HOWEVER P2)
P2 (TAKE PEOPLE DRUG)

P2 (LIKE PEOPLE P2)
P3 (NEGATE P3)
P4 (MOD PEOPLE HEALTHY)
P5 (WHEN P1 P4)

P10.S2. One of the side effects with tamoxifen can be an increased risk of uterine cancer.

P1 (WITH SIDE-EFFECT TAMOXIFEN)
P2 (NUMBER-OF SIDE-EFFECT ONE)
P3 (OF RISK UTERINE-CANCER)
P4 (AMOUNT-OF RISK INCREASE)
P5 (CAN-BE P1 P3)

P10.S3. The uterine cancer risk is less than 1 percent.

P1 (IS P2 P1)
P2 (LESS-THAN RISK 0.01)
P3 (REF RISK UTERINE-CANCER)

P10.S4. But if you've already got a fear of cancer because you're at such high risk, that's a scary thing.

P1 (BECAUSE P2 P4)
P2 (POSSESS ALREADY FEAR)
P3 (OF FEAR CANCER)
P4 (POSSESS $ HIGH-RISK)
P5 (IS-A THING SCARY)
P6 (REF P6 P5)

P11.S1. The current screening recommendation for mutation carriers is annual mammography starting somewhere between 25 and 35 and clinical examination of the breasts every six to 12 months.

P1 (IS SCREENING-RECOMMENDATION MAMMOGRAPHY)
P2 (MOD SCREENING-RECOMMENDATION CURRENT)
P3 (MOD MAMMOGRAPHY ANNUAL)
P4 (FOR P2 MUTATION-CARRIER)
P5 (START-BETWEEN P1 25-35-YEARS-OLD)
P6 (AND P1 P7)
P7 (OF CLINICAL-EXAMINATION BREAST)
P8 (EVERY P6 6-12-MONTHS)

P11.S2. Breast self-examination has been recommended by some people, although that the effectiveness of this screening tool is controversial.

P1 (RECOMMEND PEOPLE BREAST-SELF-EXAMINATION)
P2 (AMOUNT-OF PEOPLE SOME)
P3 (OF EFFECTIVENESS SCREENING-TOOL)
P4 (IS EFFECTIVENESS CONTROVERSIAL)
P5 (IS EFFECTIVENESS CONTROVERSIAL)

P11.S3. The only difference between these recommendations and those for the average woman is that you might do the mammograms more often; yearly instead of every two years and you certainly would start at a younger age.

P1 (BETWEEN DIFFERENCE RECOMMENDATION)
P2 (MOD DIFFERENCE ONLY)
P3 (FOR RECOMMENDATION WOMAN)
P4 (MOD WOMAN AVERAGE)
P5 (DO P3 MAMMOGRAM)
P6 (MOD MAMMOGRAM OFTEN)
P7 (MOD OFTEN MORE)
P8 (MOD DO MIGHT)
P9 (INSTEAD-OF YEARLY EVERY-TWO-YEARS)
P10 (START P4 YOUNGER-AGE)
P11 (MOD START CERTAINLY)
P12 (AND P5 P10)

P12.S1. There is a question as to whether mammography works as well in women with the BRCA mutations as it does in the general population because the kinds of cancers that they get, especially with BRCA1, seem to be somewhat different.

P1 (AS-TO QUESTION MAMMOGRAPHY)
P2 (WITH WOMAN BRCA-MUTATION)
P3 (WORK MAMMOGRAPHY WELL)
P4 (WORK-IN MAMMOGRAPHY P2)
P5 (WORK-IN MAMMOGRAPHY GENERAL-POPULATION)
P6 (BECAUSE P7 P12)
P7 (AS-WELL-AS-IN P4 P5)
P8 (POSSESS P2 CANCER)
P9 (MOD CANCER KIND-OF)
P10 (WITH P2 BRCA1)
P11 (MOD BRCA1 ESPECIALLY)
P12 (SEEM-TO-BE P7 DIFFERENT)
P13 (MOD DIFFERENT SOMEWHAT)

P12.S2. They don't, for some reason, show up on mammography as well as some of the other kinds of breast cancers that other women get, possibly because they are less likely to form a mass initially.
P1 (SHOW-UP-ON BRCA-CANCERS MAMMOGRAPHY)
P2 (NEGATE P1)
P3 (MOD P2 FOR-SOME-REASON)

P4 (GET WOMAN KINDS-OF-BREAST-CANCER)
P5 (MOD WOMAN OTHER)
P6 (MOD KINDS-OF-BREAST-CANCER OTHER)
P7 (AMOUNT-OF P3 SOME)
P8 (AS-WELL-AS P1 P4)
P9 (FORM BRCA-CANCERS MASS)
P10 (LESS-LIKELY P8)
P11 (BECAUSE P1 P9)
P12 (MOD MASS INITIALLY)
P13 (MOD BECAUSE POSSIBLY)

P12.S3. Because MRI was an up-and-coming technology and seemed to be useful in detecting other medical conditions, it was being considered for breast cancer.

P1 (BECAUSE P2)
P2 (AND P4 P5)
P3 (CONSIDER-FOR MRI BREAST-CANCER)
P4 (REF MRI UP-AND-COMING-TECHNOLOGY)
P5 (SEEM-TO-DETECT MRI MEDICAL-CONDITION)
P6 (MOD MEDICAL-CONDITION OTHER)

P12.S4. So, along with ultrasound, which was another technology that people were suggesting we could add to mammography

P1 (REF ULTRASOUND OTHER-TECHOLOGY)
P2 (SUGGEST PEOPLE ULTRASOUND)
P3 (ADD-TO ULTRASOUND MAMMOGRAPHY)
P4 (GIVE P3 INFORMATION)
P5 (AMOUNT-OF INFORMATION MORE)
P6 (ALONG-WITH ULTRASOUND P7)
P7 (WHICH P7)
P8 (OR P8 P9)
P9 (IS TECHNIQUE BEST)
P10 (IS COMBINATION-TECHNIQUE BEST)

P13.S1. When our study was published, there were 22 breast cancers that had been identified in more than 200 women.

P1 (WHEN P2)
P2 (IS STUDY PUBLISH)
P3 (IDENTIFY STUDY BREAST-CANCER)
P4 (AMOUNT-OF BREAST-CANCER 22)
P5 (IN P1 WOMAN)
P6 (AMOUNT-OF WOMAN MORE-THAN-200)

P13.S2. What was interesting was that a lot of these cancers were found only on MRI.

P1 (FOUND CANCER MRI)
P2 (MOD MRI ONLY-ON)
P3 (AMOUNT-OF CANCER A-LOT)
P4 (MOD P1 INTERESTING)
P5 (WHAT P4)

P13.S3. Some were found only on mammograms, and ultrasound generally didn't, overall, seem to contribute very much.

P1 (FOUND CANCER MAMMOGRAM)
P2 (MOD MAMMOGRAM ONLY-ON)
P3 (CONTRIBUTE ULTRASOUND P1)
P4 (AMOUNT-OF CONTRIBUTE VERY-MUCH)
P5 (NEGATE P3)
P6 (MOD P3 GENERALLY)
P7 (MOD P3 OVERALL)

P13.S4. The MRI seemed to pick up the most cancers, and it seemed to pick them up at a relatively early stage.

P1 (SEEM-TO P2)
P2 (AND P3 P5)
P3 (PICK-UP MRI CANCER)
P4 (AMOUNT-OF CANCER MOST)
P5 (AT P3 EARLY-STAGE)
P6 (MOD EARLY-STAGE RELATIVELY)

P14.S1. At this point, we're recommending that mutation carriers have MRI yearly and mammogram yearly starting between ages 25 and 30.

P1 (RECOMMEND $ P2)
P2 (AND P3 P5)
P3 (HAVE MUTATION-CARRIER MRI)
P4 (MOD P1 YEARLY)
P5 (HAVE MUTATION-CARRIER MAMMOGRAM)
P6 (START-BETWEEN P5 AGE-25-30)

P14.S2. Mammograms need to be done as well because some types of early breast cancer, specifically ductal carcinoma in situ, that show up on the mammogram as little specks of calcium, what we call microcalcifications, don't show up on MRI.

P1 (SHOW-UP-ON BREAST-CANCER MAMMOGRAM)
P2 (MOD BREAST-CANCER EARLY)
P3 (AMOUNT-OF BREAST-CANCER SOME)

P4 (SHOW-UP-ON BREAST-CANCER MRI)
P5 (NEGATE P4)
P6 (AS P1 SPECK-OF-CALCIUM)
P7 (MOD SPECT-OF-CALCIUM)
P8 (MOD SPECT-OF-CALCIUM LITTLE)
P9 (REF P6 MICROCALCIFICATION)
P10 (REF BC DUCTAL-CARCINOMA-IN-SITU)
P11 (NEED-TO-BE MAMMOGRAM DONE)
P12 (MOD P10 AS-WELL)

P14.S3. We're not suggesting that women have the addition of ultrasound, because we think adding ultrasound doesn't really add much more.

P1 (SUGGEST $ ULTRASOUND)
P2 (OF ADDITION ULTRASOUND)
P3 (HAVE P2 WOMAN)
P4 (NEGATE P1)
P5 (BECAUSE P4 P8)
P6 (ADD ULTRASOUND MORE)
P7 (MOD MORE MUCH)
P8 (NEGATE P6)

P14.S4. We also recommend a clinical breast examination twice a year.

P1 (RECOMMEND $ CLINICAL-BREAST-EXAMINATION)
P2 (AMOUNT-OF CLINICAL-BREAST-EXAMINATION TWICE-A-YEAR)
P3 (ALSO P1)

P15.S1. There are a lot of things that show up on MRI that aren't cancer, such as cysts and nodules, so you have to sort that out, hopefully, without having to do surgery to prove it.

P1 (SHOW-UP-ON THING MRI)
P2 (AMOUNT-OF THING A-LOT)
P3 (IS THING CANCER)
P4 (NEGATE P3)
P5 (EXAMPLE-OF P3 CYST)
P6 (EXAMPLE-OF P3 NODULES)
P7 (SORT-OUT GENERAL-PUBLIC P1)
P8 (WITHOUT-HAVE P7 SURGERY)
P9 (TO-PROVE SURGERY P1)

P15.S2. False-positive results are still a concern in women with the BRCA mutations, but the higher the risk of cancer, the more one is willing to tolerate the downside of false-positives.

P1 (IS-A FALSE-POSITIVE-RESULT CONCERN)
P2 (IN FALSE-POSITIVE-RESULT WOMAN)
P3 (WITH WOMAN BRCA-MUTATION)
P4 (OF RISK CANCER)
P5 (AMOUNT-OF P4 HIGHER)
P6 (BUT P5)
P7 (TOLERATE GENERAL-PUBLIC FALSE-POSITIVE)
P8 (OF DOWNSIDE FALSE-POSITIVE)
P9 (WILLING-TO P7)
P10 (AMOUNT-OF P9 MORE)

P16.S1. But given the high number of false-positive results and knowing the technical problems with MRI, I don't think it will ever become a standard screening tool for the general population.

P1 (GIVEN P2)
P2 (OF NUMBER FALSE-POSITIVE-RESULT)
P3 (AND P2 P4)
P4 (AMOUNT-OF NUMBER HIGH)
P5 (WITH TECHNICAL-PROBLEM MRI)
P6 (KNOWING P5)
P7 (BECOME MRI STANDARD-SCREENING-TOOL)
P8 (FOR STANDARD-SCREENING-TOOL GENERAL-POPULATION)
P9 (NEGATE P8)

P16.S2. It's expensive and time-consuming.

P1 (AND P2 P3)
P2 (MOD MRI EXPENSIVE)
P3 (MOD MRI TIME-CONSUMING)

P16.S3. It often takes almost an hour to do a screen.

P1 (DO $ SCREEN)
P2 (TAKES P1 ONE-HOUR)
P3 (MOD ONE-HOUR ALMOST)
P4 (MOD P2 OFTEN)

P16.S4. It involves an injection for the imaging.

P1 (INVOLVE MRI INJECTION)
P2 (FOR P1 IMAGING)

P16.S3. There is also a big issue regarding claustrophobia because the space inside the magnet is smaller in a breast MRI machine than it is in a regular MRI machine.

P1 (REGARDING ISSUE CLAUSTROPHOBIA)
P2 (MOD ISSUE BIG)
P3 (BECAUSE P1 P7)
P4 (INSIDE SPACE MAGNET)
P5 (IN P3 BREAST-MRI-MACHINE)
P6 (IN P3 REGULAR-MRI-MACHINE)
P7 (SMALLER-THAN P4 P5)

P17.S1. The results of our study are applicable only to mutation carriers.

P1 (OF RESULT STUDY)
P2 (APPLICABLE-TO RESULT MUTATION-CARRIER)
P3 (MOD MUTATION-CARRIER ONLY)

P17.S2. This does not imply that MRI is beneficial for all high-risk women because there are many other reasons for women to be at high risk, and we don't yet have evidence to say that all high-risk women should get MRI screening.

P1 (IS MRI BENEFICIAL)
P2 (FOR P1 HIGH-RISK-WOMAN)
P3 (IMPLY P2)
P4 (NEGATE P3)
P5 (BECAUSE P5 P7)
P6 (FOR REASON WOMAN)
P7 (TO-BE-AT WOMAN HIGH-RISK)
P8 (MOD REASON OTHER)
P9 (AMOUNT-OF REASON MANY)
P10 (HAVE $ EVIDENCE)
P11 (NEGATE P10)
P12 (SHOULD-GET HIGH-RISK-WOMAN MRI-SCREENING)
P13 (TO-SAY P11 P12)

P18.S1. There's a whole question of: Do you alternate between MRIs and mammograms?

P1 (AND P1 P2)
P2 (ALTERNATE-BETWEEN $ MRI)
P3 (ALTERNATE-BETWEEN $ MAMMOGRAM)
P4 (QUESTION-OF P1)

P18.S2. Do you do something every six months?

P1 (DO $ SOMETHING)
P2 (DURATION-OF P1 EVERY-SIX-MONTHS)

P18.S3. Also the question has been raised: Would it be better to do only MRI in women when they're very young since their breasts are very dense, making mammograms hard to read?

P1 (RAISE $ QUESTION)
P2 (BETTER-TO-DO P3)
P3 (IN MRI WOMAN)
P4 (MOD MRI ONLY)
P5 (MOD WOMAN YOUNG)
P6 (MOD YOUNG VERY)
P7 (SINCE WOMAN BREAST)
P8 (MAKING P9 P11)
P9 (IS BREAST DENSE)
P10 (MOD DENSE VERY)
P11 (HARD-TO MAMMOGRAM READ)

P18.S4. These are all questions that we're playing with to figure out what would be ideal.

P1 (PLAY-WITH $ QUESTION)
P2 (AMOUNT-OF QUESTION ALL)
P3 (TO P1 P4)
P4 (FIGURE-OUT $ IDEAL)
P5 (IS WHAT IDEAL)

P18.S5. The next step in our study is to see if we've had an impact on survival, because, obviously, that's the endpoint that we really want to reach.

P1 (IN NEXT-STEP STUDY)
P2 (SEE-IF RESEARCH P3)
P3 (ON IMPACT SURVIVAL)
P4 (IS P3 ENDPOINT)
P5 (WANT-TO RESEARCH REACH)

P19.S1. Given the fact that the field is constantly changing and there are new recommendations and new studies coming out all the time, I think that it's important to hook yourself up with a high-risk program, if you possibly can.

P1 (GIVEN P4)
P2 (IS FIELD CHANGING)
P3 (MOD CHANGING CONSTANTLY)
P4 (AND P4 P5)

P5 (DURATION-OF P3 ALL-THE-TIME)
P6 (COME-OUT $ RECOMMENDATION)
P7 (COME-OUT $ STUDY)
P8 (MOD RECOMMENDATION NEW)
P9 (MOD STUDY NEW)
P10 (IMPORTANT P10)
P11 (HOOK-UP WOMAN HIGH-RISK-PROGRAM)
P12 (IF P10 POSSIBLE)

P19.S2. There are issues not just around breast cancer, there are issues around ovarian cancer.

P1 (AROUND ISSUE BREAST-CANCER)
P2 (MOD P1 JUST)
P3 (NEGATE P3)
P4 (AROUND ISSUE OVARIAN-CANCER)

P19.S3. With BRCA2, there is increased risk of melanoma and pancreatic cancer.

P1 (WITH-BRCA2 P2)
P2 (AND P3 P4)
P3 (OF RISK MELANOMA)
P4 (OF RISK PANCREATIC-CANCER)
P5 (AMOUNT-OF RISK INCREASE)

P19.S4. There are also psychological issues.

P1 (IS $ PSYCHOLOGICAL-ISSUE)
P2 (MOD P1 ALSO)

P19.S5. This is a very scary thing for women, and most high-risk programs offer psychological counseling.

P1 (AND P2 P5)
P2 (FOR PSYCHOLOGICAL-ISSUE WOMAN)
P3 (MOD PSYCHOLOGICAL-ISSUE SCARY)
P4 (MOD SCARY VERY)
P5 (OFFER HIGH-RISK-PROGRAM PSYCHOLOGICAL-COUNSELING)
P6 (MOD HIGH-RISK-PROGRAM MOST)

P19.S6. So my advice is to try to get your screening through a high-risk program.

P1 (OF ADVICE AUTHOR)

P2 (IS P1 P3)
P3 (GET WOMAN SCREENING)
P4 (THROUGH P1 HIGH-RISK-PROGRAM)

Web page 3: Colorectal cancer; Low readability

P1.S1. Colon cancer is a disease in which malignant (cancer) cells form in the tissues of the colon.

P1 (IS-A COLON-CANCER DISEASE)
P2 (FORM-IN MALIGNANT-CELL TISSUE)
P3 (OF TISSUE COLON)
P4 (REF P1 P2)

P1.S2. Age and health history can affect the risk of developing colon cancer.

P1 (AND P2 P3)
P2 (CAN-AFFECT AGE RISK)
P3 (CAN-AFFECT HEALTH-HISTORY RISK)
P4 (OF RISK COLON-CANCER)
P5 (DEVELOP $ COLON-CANCER)

P1.S3. Possible signs of colon cancer include a change in bowel habits or blood in the stool.

P1 (OF SIGN COLON-CANCER)
P2 (MOD SIGN POSSIBLE)
P3 (INCLUDE P1)
P4 (OR P4 P5)
P5 (IN CHANGE BOWEL-HABIT)
P6 (IN BLOOD STOOL)

P1.S4. Tests that examine the rectum, rectal tissue, and blood are used to detect (find) and diagnose colon cancer.

P1 (AND (P2,P3,P4))
P2 (EXAMINE TEST RECTUM)
P3 (EXAMINE TEST RECTAL-TISSUE)
P4 (EXAMINE TEST BLOOD)
P5 (USE-TO P1)
P6 (AND P2 P3)
P7 (DETECT(FIND) TEST COLON-CANCER)
P8 (DIAGNOSE TEST COLON-CANCER)

P1.S5. Certain factors affect prognosis (chance of recovery) and treatment options.

P1 (AND P2 P3)

P2 (AFFECT FACTOR PROGNOSIS)
P3 (AFFECT FACTOR TREATMENT-OPTION)
P4 (MOD FACTOR CERTAIN)

P3.S1. The colon is part of the body's digestive system.

P1 (PART-OF P2 P3)
P2 (IS COLON DIGESTIVE-SYSTEM)
P3 (POSSESS BODY DIGESTIVE-SYSTEM)

P3.S2. The digestive system removes and processes nutrients (vitamins, minerals, carbohydrates, fats, proteins, and water)

P1 (AND P2 P3)
P2 (REMOVE DIGESTIVE-SYSTEM NUTRIENT)
P3 (PROCESS DIGESTIVE-SYSTEM NUTRIENT)
P4 (FROM P2 FOOD)
P5 (FROM P3 FOOD)
P6 (AND (P7,P8,P9,P10,P11,P12))
P7 (EXAMPLE-OF NUTRIENT VITAMIN)
P8 (EXAMPLE-OF NUTRIENT MINERAL)
P9 (EXAMPLE-OF NUTRIENT CARBOHYDRATE)
P10 (EXAMPLE-OF NUTRIENT FAT)
P11 (EXAMPLE-OF NUTRIENT PROTEIN)
P12 (EXAMPLE-OF NUTRIENT WATER)
P13 (HELP-PASS DIGESTIVE-SYSTEM WASTE-MATERIAL)
P14 (OUT-OF P6 BODY)

P3.S3. The digestive system is made up of the esophagus, stomach, and the small and large intestines.

P1 (AND (P2,P3,P4,P5))
P2 (MADE-UP-OF DIGESTIVE-SYSTEM ESOPHAGUS)
P3 (MADE-UP-OF DIGESTIVE-SYSTEM STOMACH)
P4 (MADE-UP-OF DIGESTIVE-SYSTEM SMALL-INTESTINE)
P5 (MADE-UP-OF DIGESTIVE-SYSTEM LARGE-INTESTINE)

P3.S4. The first 6 feet of the large intestine are called the large bowel or colon.

P1 (OR P2 P3)
P2 (REF P4 LARGE-BOWEL)
P3 (REF P4 COLON)

P4 (OF FIRST-SIX-FEET LARGE-INTESTINE)

P3.S5. The last 6 inches are the rectum and the anal canal.

P1 (AND P2 P3)
P2 (REF LAST-SIX-INCH RECTUM)
P3 (REF LAST-SIX-INCH ANAL-CANAL)

P3.S4. The anal canal ends at the anus (the opening of the large intestine to the outside of the body).

P1 (END-AT ANAL-CANAL ANUS)
P2 (REF ANUS P3)
P3 (OF OPENING LARGE-INTESTINE)
P4 (TO LARGE-INTESTINE OUTSIDE)
P5 (OF OUTSIDE BODY)

P5.S1. Age 50 or older.

P1 (OR AGE-50 OLDER)

P5.S2. A family history of cancer of the colon or rectum.

P1 (OF FAMILY-HISTORY CANCER)
P2 (OR P3 P4)
P3 (OF P1 COLON)
P4 (OF P1 RECTUM)

P5.S3. A personal history of cancer of the colon, rectum, ovary, endometrium, or breast.

P1 (OF PERSONAL-HISTORY CANCER)
P2 (OR (P3,P4,P5,P6,P7))
P3 (OF P1 COLON)
P4 (OF P1 RECTUM)
P5 (OF P1 OVARY)
P6 (OF P1 ENDOMETRIUM)
P7 (OF P1 BREAST)

P5.S4. A history of polyps in the colon.

P1 (OF HISTORY POLYPS)
P2 (IN P1 COLON)

P5.S5. A history of ulcerative colitis (ulcers in the lining of the large intestine) or Crohn's

disease.

P1 (OR P2 P6)
P2 (OF HISTORY ULCERATIVE-COLITIS)
P3 (REF P4)
P4 (IN ULCER LINING)
P5 (OF P3 LARGE-INTESTINE)
P6 (OF HISTORY CROHN'S-DISEASE)

P5.S6. Certain hereditary conditions, such as familial adenomatous polyposis and hereditary nonpolyposis colon cancer (HNPCC; Lynch Syndrome).

P1 (AND P3 P5)
P2 (MOD HEREDITARY-CONDITION CERTAIN)
P3 (EXAMPLE-OF HEREDITARY-CONDITION FAMILIAL-ADENOMATOUS-POLYOSIS)
P4 (REF FAMILIAL-ADENOMATOUS-POLYOSIS HNPCC)
P5 (EXAMPLE-OF HEREDITARY-CONDITION HEREDITARY-NONPOLYOSIS-COLON-CANCER)

P6 (REF HEREDITARY-NONPOLYOSIS-COLON-CANCER LYNCH-SYNDROME)

P7.S1. These and other symptoms may be caused by colon cancer or by other conditions.

P1 (AND (P5.S1,P5.S2,P5.S3.P5.S4,P5.S5,P5.S6) SYMPTOM)
P2 (MOD SYMPTOM OTHER)
P3 (OR P4 P5)
P4 (CAUSE-BY P1 COLON-CANCER)
P5 (CAUSE-BY P1 OTHER-CONDITION)
P6 (MOD P3 POSSIBLE)

P8.S1. A change in bowel habits.

P1 (IN CHANGE BOWEL-HABIT)

P8.S2. Blood (either bright red or very dark) in the stool.

P1 (IN BLOOD STOOL)
P2 (EITHER P3 P4)
P3 (MOD BLOOD BRIGHT-RED)
P4 (MOD BLOOD VERY-DARK)

P8.S3. Diarrhea, constipation, or feeling that the bowel does not empty completely.

P1 (OR (P2,P3,P4))
P2 ($ DIARRHEA)
P3 ($ CONSTIPATION)
P4 (FEELING-OF P3)
P5 (EMPTY BOWEL COMPLETELY)
P6 (NEGATE P3)

P8.S4. Stools that are narrower than usual.

0

P1 (IS STOOL NARROWER)
P2 (MOD P1 THAN-USUAL)

P8.S5. General abdominal discomfort (frequent gas pains, bloating, fullness, or cramps).

P1 ($ GENERAL-ABDOMINAL-DISCOMFORT)
P2 (OR (P3,P5,P6,P7))
P3 (REF P1 GAS-PAIN)
P4 (MOD GAS-PAIN FREQUENT)
P5 (REF P1 BLOATING)
P6 (REF P1 FULLNESS)
P7 (REF P1 CRAMP)

P8.S6. Weight loss with no known reason.

P1 (WITH WEIGHT-LOSS KNOWN-REASON)
P2 (MOD KNOWN-REASON NO)

P8.S7. Constant tiredness.

P1 (MOD TIREDNESS CONSTANT)

P8.S8. Vomiting.

0

P1 ($ VOMITING)

P10.S1. Physical exam and history: An exam of the body to check general signs of health, including checking for signs of disease,

P1 (AND PHYSICAL-EXAM HISTORY)
P2 (REF P1 (P3,P7))
P3 (EXAM CHECK HEALTH)
P4 (OF GENERAL-SIGN HEALTH)

P5 (OF EXAM BODY)
P6 (INCLUDE P5 P7)
P7 (FOR CHECK SIGN-OF-DISEASE)
P8 (EXAMPLE-OF SIGN-OF-DISEASE LUMP)
P9 (EXAMPLE-OF SIGN-OF-DISEASE ANYTHING-UNUSUAL)

P10.S1.1. A history of the patient's health habits and past illnesses and treatments will also be taken.

P1 (AND (P3,P4,P5))
P2 (OF HISTORY P1)
P3 (POSSESS PATIENT HEALTH-HABIT)
P4 (POSSESS PATIENT PAST-ILLNESS)
P5 (POSSESS PATIENT TREATMENT)

P10.S2. Fecal occult blood test: A test to check stool (solid waste) for blood that can only be seen with a microscope.

P1 (REF FECAL-OCCULT-BLOOD-TEST P2)
P2 (CHECK TEST STOOL)
P3 (FOR P1 BLOOD)
P4 (SEE-WITH BLOOD MICROSCOPE)
P5 (MOD SEE-WITH ONLY)

P10.S2.1. Small samples of stool are placed on special cards and returned to the doctor or laboratory for testing.

P1 (AND P1 P6)
P2 (PLACE-ON SAMPLE SPECIAL-CARD)
P3 (OF SAMPLE STOOL)
P4 (MOD SAMPLE SMALL)
P5 (OR P7 P8)
P6 (FOR P5 TESTING)
P7 (RETURN-TO P2 DOCTOR)
P8 (RETURN-TO P2 LABORATORY)

P10.S3. Digital rectal exam: An exam of the rectum.

P1 (REF DIGITAL-RECTAL-EXAM P2)
P2 (OF EXAM RECTUM)

P10.S3.1. The doctor or nurse inserts a lubricated, gloved finger into the rectum to feel for lumps or abnormal areas.

P1 (OR P2 P3)
P2 (INSERT DOCTOR GLOVED-FINGER)
P3 (INSERT NURSE GLOVED-FINGER)
P4 (MOD GLOVED-FINGER LUBRICATED)
P5 (INTO P1 RECTUM)
P6 (OR P7 P8)
P7 (FEEL-FOR P5 LUMP)
P8 (FEEL-FOR P5 ABNORMAL-AREA)

P10.S4. Barium enema: A series of x-rays of the lower gastrointestinal tract.

P1 (OF SERIES X-RAY)
P2 (OF P1 LOWER-GASTROINTESTINAL-TRACT)
P3 (REF P2 BARIUM-ENEMA)

P10.S4.1. A liquid that contains barium (a silver-white metallic compound) is put into the rectum.

P1 (CONTAIN LIQUID BARIUM)
P2 (REF BARIUM SILVER-WHITE-METALLIC-COMPOUND)
P3 (PUT-INTO P1 RECTUM)

P10.S4.2. The barium coats the lower gastrointestinal tract and x-rays are taken.

P1 (AND P2 P3)
P2 (COAT BARIUM LOWER-GASTROINTESTINAL-TRACT)
P3 (TAKE $ X-RAY)

P10.S4.3. This procedure is also called a lower GI series.

P1 (REF PROCEDURE LOWER-GI-SERIES)

P10.S5. Sigmoidoscopy: A procedure to look inside the rectum and sigmoid (lower) colon for polyps, abnormal areas, or cancer.

P1 (REF SIGMOIDOSCOPY P2)
P2 (AND P3 P4)
P3 (LOOK-INSIDE PROCEDURE RECTUM)
P4 (LOOK-INSIDE PROCEDURE SIGMOID-COLON)
P5 (OR (P5,P6,P7))
P6 (FOR P2 POLYPS)
P7 (FOR P2 ABNORMAL-AREA)
P8 (FOR P2 CANCER)

P10.S5.1. A sigmoidoscope (a thin, lighted tube) is inserted through the rectum into the

sigmoid colon.

P1 (REF SIGMOIDOSCOPE THIN-LIGHTED-TUBE)
P2 (INSERT-THROUGH SIGMOIDOSCOPE RECTUM)
P3 (INTO P2 SIGMOID-COLON)

P10.S5.2. Polyps or tissue samples may be taken for biopsy.

P1 (OR P2 P3)
P2 (MOD P1 POSSIBLE)
P3 (TAKE $ POLYPS)
P4 (TAKE $ TISSUE-SAMPLE)
P5 (FOR P1 BIOPSY)

P10.S7. Biopsy: The removal of cells or tissues so they can be viewed under a microscope to check for signs of cancer.

P1 (OR P2 P3)
P2 (OF REMOVAL CELL)
P3 (OF REMOVAL TISSUE)
P4 (VIEW-UNDER P1 MICROSCOPE)
P5 (CHECK-FOR P4 CANCER)
P6 (OF SIGN CANCER)

P10.S8. Virtual colonoscopy: A procedure that uses a series of x-rays called computed tomography to make a series of pictures of the colon.

P1 (USE PROCEDURE X-RAY)
P2 (OF SERIES X-RAY)
P3 (REF P2 COMPUTED-TOMOGRAPHY)
P4 (MAKE COMPUTED-TOMOGRAPHY PICTURE)
P5 (OF SERIES PICTURE)
P6 (OF P5 COLON)

P10.S8.1. A computer puts the pictures together to create detailed images that may show polyps and anything else

P1 (TO P2 P3)
P2 (PUT COMPUTER PICTURE-TOGETHER)
P3 (CREATE COMPUTER IMAGE)
P4 (MOD IMAGE DETAIL)
P5 (AND P7 P8)
P6 (MOD P5 POSSIBLE)

P7 (SHOW IMAGE POLYPS)
P8 (SHOW IMAGE ANYTHING-UNUSUAL)
P9 (ON P8 COLON)
P10 (OF INSIDE-SURFACE COLON)

P10.S8.2. This test is also called colonography or CT colonography.

P1 (OR P2 P3)
P2 (REF TEST COLONOGRAPHY)
P3 (REF TEST CT-COLONOGRAPHY)

P12.S1. The stage of the cancer (whether the cancer is in the inner lining of the colon only, involves the whole colon,

P1 (OF CANCER STAGE)
P2 (REF P1 P3)
P3 (OR (P3,P6,P7))
P4 (IN CANCER INNER-LINING)
P5 (OF INNER-LINING COLON)
P6 (MOD P3 ONLY)
P7 (INVOLVE CANCER WHOLE-COLON)
P8 (SPREAD CANCER OTHER-PLACE)
P9 (IN OTHER-PLACE BODY)

P12.S2. Whether the cancer has blocked or created a hole in the colon.

P1 (OR P2 P3)
P2 (BLOCK CANCER HOLE)
P3 (CREATE CANCER HOLE)
P4 (IN P1 COLON)
P5 (WHETHER P4)

P12.S3. The blood levels of carcinoembryonic antigen (CEA; a substance in the blood that may be increased when cancer is present) before treatment begins.

P1 (OF BLOOD-LEVEL CARCINOEMBRYONIC-ANTIGEN)
P2 (REF CARCINOEMBRYONIC-ANTIGEN CEA)
P3 (REF P6)
P4 (BEFORE BLOOD-LEVEL TREATMENT)
P5 (BEGIN $ TREATMENT)
P6 (IN SUBSTANCE BLOOD)
P7 (IS CANCER PRESENT)
P8 (INCREASE-WHEN P6 P7)

P9 (MOD INCREASE-WHEN POSSIBLE)

P12.S4. Whether the cancer has recurred.

P1 (RECUR $ CANCER)
P2 (WHETHER P2)

P12.S5. The patient's general health.

P1 (POSSESS PATIENT GENERAL-HEALTH)

P13.S1. The stage of the cancer.

P1 (OF STAGE CANCER)

Web Page 4: Colorectal cancer; High Readability

P1.S1. Even though we do not know the exact cause of most colorectal cancer, it is possible to prevent many colorectal cancers.

P1 (KNOW $ P3)
P2 (NEGATE P1)
P3 (OF CAUSE COLORECTAL-CANCER)
P4 (AMOUNT-OF COLORECTAL-CANCER MOST)
P5 (MOD CAUSE EXACT)
P6 (PREVENT $ COLORECTAL-CANCER)
P7 (MOD PREVENT POSSIBLE)
P8 (AMOUNT-OF COLORECTAL-CANCER MANY)

P1.S2. Following the American Cancer Society screening guidelines (see "Can Colorectal Polyps and Cancer Be Found Early?") can lower the number of cases of the disease by finding and removing polyps that could become cancerous, and can also lower the death rate from colorectal cancer by finding disease early when it is highly curable.

P1 (FOLLOW $ SCREENING-GUIDELINE)
P2 (LABEL SCREENING-GUIDELINE AMERICAN-CANCER-SOCIETY)
P3 (LOWER P1 P4)
P4 (OF NUMBER CASE)
P5 (OF CASE DISEASE)
P6 (BY P3 P7)
P7 (AND P8 P9)
P8 (FIND $ POLYPS)
P9 (REMOVE $ POLYPS)
P10 (BECOME P6 CANCEROUS)
P11 (MOD BECOME POSSIBLE)
P12 (BY P13 P15)
P13 (LOWER $ DEATH-RATE)
P14 (FROM DEATH-RATE COLORECTAL-CANCER)
P15 (FIND DISEASE EARLY)
P16 (REF P15 HIGHLY-CURABLE)

P2.S1. Prevention and early detection are possible because most colorectal cancers develop from adenomatous polyps.

P1 (AND P2 P3)
P2 (IS PREVENTION POSSIBLE)
P3 (IS EARLY-DETECTION POSSIBLE)
P4 (BECAUSE P1 P6)
P5 (DEVELOP $ COLORECTAL-CANCER)

P6 (AMOUNT-OF COLORECTAL-CANCER MOST)
P7 (FROM P5 ADENOMATOUS-POLYPS)

P2.S2. Polyps are noncancerous growths in the colon and rectum.

P1 (IS POLYPS GROWTH)
P2 (MOD GROWTH NONCANCEROUS)
P3 (AND P4 P5)
P4 (IN GROWTH COLON)
P5 (IN GROWTH RECTUM)

P2.S3. Removing them can lower a person's cancer risk.

P1 (REMOVE $ POLYPS)
P2 (POSSESS PERSON CANCER-RISK)
P3 (LOWER P1 P2)

P3.S1. People can lower their risk of developing colorectal cancer by managing the risk factors that they can control, such as diet and physical activity.

P1 (LOWER PEOPLE RISK)
P2 (OF RISK COLORECTAL-CANCER)
P3 (DEVELOP $ COLORECTAL-CANCER)
P4 (BY P1 P5)
P5 (MANAGE $ RISK-FACTOR)
P6 (CONTROL PEOPLE RISK-FACTOR)
P7 (AND P8 P9)
P8 (EXAMPLE-OF P6 DIET)
P9 (EXAMPLE-OF P6 PHYSICAL ACTIVITY)

P3.S2. It is important to eat plenty of fruits, vegetables, and whole grain foods and to limit intake of high-fat foods.

P1 (AND (P2,P3,P4))
P2 (EAT $ FRUIT)
P3 (EAT $ VEGETABLE)
P4 (EAT $ WHOLE-GRAIN-FOOD)
P5 (MOD PLENTY FRUIT)
P6 (MOD PLENTY VEGETABLE)
P7 (MOD PLENTY WHOLE-GRAIN-FOOD)
P8 (AND P1 P9)
P9 (LIMIT $ INTAKE)
P10 (OF INTAKE FOOD)
P11 (MOD FOOD HIGH-FAT)
P12 (IS P8 IMPORTANT)

P3.S3. Physical activity is another area that people can control.

P1 (CONTROL PEOPLE PHYSICAL-ACTIVITY)
P2 (MOD CONTROL POSSIBLE)
P3 (IS PHYSICAL-ACTIVITY AREA)
P4 (MOD AREA ANOTHER)

P3.S4. The American Cancer Society recommends at least 30 minutes of physical activity on 5 or more days of the week.

P1 (RECOMMEND AMERICAN-CANCER-SOCIETY PHYSICAL-ACTIVITY)
P2 (AMOUNT-OF PHYSICAL-ACTIVITY AT-LEAST-30-MIN)
P3 (DURATION-OF PHYSICAL-ACTIVITY 5-OR-MORE-DAYS/WK)

P3.S5. If you participate in moderate or vigorous activity for 45 minutes on 5 or more days of the week you can lower your risk for breast and colorectal cancer even more.

P1 (PARTICIPATE-IN $ MODERATE-ACTIVITY)
P2 (PARTICIPATE-IN $ VIGOROUS-ACTIVITY)
P3 (OR P1 P2)
P4 (AMOUNT-OF P3 45-MINUTES)
P5 (DURATION-OF P3 5-OR-MORE-DAYS/WK)
P6 (LOWER P3 RISK)
P7 (AMOUNT-OF LOWER MORE)
P8 (OF RISK BREAST-CANCER)
P9 (OF RISK COLORECTAL-CANCER)

P3.S6. If you are overweight, you can ask your doctor about a weight loss plan that will work for you.

P1 (IS PERSON OVERWEIGHT)
P2 (ASK PERSON DOCTOR)
P3 (ABOUT P2 WEIGHT-LOSS-PLAN)
P4 (WORK $ WEIGHT-LOSS-PLAN)
P5 (FOR P4 PERSON)

P3.S7. For more information about diet and physical activity, refer to our document "American Cancer Society Guidelines for Nutrition and Cancer Prevention."

P1 (ABOUT INFORMATION DIET)
P2 (ABOUT INFORMATION PHYSICAL-ACTIVITY)
P3 (AMOUNT-OF INFORMATION MORE)
P4 (AND P1 P2)
P5 (FOR DOCUMENT P4)
P6 (REF DOCUMENT P7)

P7 (AND P8 P9)
P8 (FOR AMERICAN-CANCER-SOCIETY-GUIDELINE NUTRITION)
P9 (FOR AMERICAN-CANCER-SOCIETY-GUIDELINE CANCER-PREVENTION)

P4.S1. Some studies suggest that taking a daily multivitamin containing folic acid, or folate, can lower colorectal cancer risk.

P1 (SUGGEST STUDY MULITIVITAMIN)
P2 (AMOUNT-OF STUDY SOME)
P3 (DURATION-OF MULITIVITAMIN DAILY)
P4 (CONTAIN MULITIVITAMIN FOLIC-ACID)
P5 (REF FOLIC-ACID FOLATE)
P6 (LOWER MULITIVITAMIN COLORECTAL-CANCER-RISK)

P4.S2. Other studies suggest that increasing calcium intake via supplements or low-fat dairy products will lower risk.

P1 (SUGGEST STUDY CALCIUM-INTAKE)
P2 (MOD STUDY OTHER)
P3 (AMOUNT-OF CALCIUM-INTAKE INCREASE)
P4 (OR P5 P6)
P5 (THROUGH P2 SUPPLEMENT)
P6 (THROUGH P2 DAIRY-PRODUCT)
P7 (MOD DAIRY-PRODUCT LOW-FAT)
P8 (LOWER P3 RISK)

P4.S3. Some have suggested that vitamin D, which you can get from sun exposure or in a vitamin pill or in milk, can lower colorectal cancer risk.

P1 (SUGGEST SOME VITAMIN-D)
P2 (OR (P3,P4,P5))
P3 (FROM VITAMIN-D SUN-EXPOSURE)
P4 (FROM VITAMIN-D VITAMIN-PILL)
P5 (FROM VITAMIN-D MILK)
P6 (LOWER VITAMIN-D COLORECTAL-CANCER-RISK)

P4.S4. Indeed the rate of this cancer is lower in the Sunbelt states.

P1 (OF RATE CANCER)
P2 (IS RATE LOWER)
P3 (IN P2 SUNBELT-STATE)

P4.S5. Of course, excessive sun exposure can cause skin cancer and is not recommended as a way to lower colorectal cancer risk.

P1 (CAUSE SUN-EXPOSURE SKIN-CANCER)
P2 (AMOUNT-OF SUN-EXPOSURE EXCESSIVE)
P3 (RECOMMEND $ SUN-EXPOSURE)
P4 (NEGATE P3)
P5 (AS SUN-EXPOSURE WAY)
P6 (LOWER $ COLORECTAL-CANCER-RISK)

P4.S6. Vitamin D may work better if you also take calcium.

P1 (WORK $ VITAMIN-D)
P2 (MOD WORK BETTER)
P3 (IF P1 P4)
P4 (TAKE $ CALCIUM)

P5.S1. Many studies have found that people who regularly use aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen (Motrin, Advil) and naproxen (Aleve) have 20% to 50% lower risk of colorectal cancer and adenomatous polyps.

P1 (FIND STUDY P3)
P2 (AMOUNT-OF STUDY MANY)
P3 (AND P4 P5)
P4 (USE PEOPLE ASPIRIN)
P5 (USE PEOPLE NONSTEROIDAL-ANTI-INFLAMMATORY-DRUGS)
P6 (MOD USE REGULARLY)
P7 (REF NONSTEROIDAL-ANTI-INFLAMMATORY-DRUGS NSAIDS)
P8 (MOD NONSTEROIDAL-ANTI-INFLAMMATORY-DRUGS OTHER)
P9 (EXAMPLE-OF NONSTEROIDAL-ANTI-INFLAMMATORY-DRUGS IBUPROFEN)
P10 (EXAMPLE-OF IBUPROFEN MOTRIN)
P11 (EXAMPLE-OF IBUPROFEN ADVIL)
P12 (EXAMPLE-OF NONSTEROIDAL-ANTI-INFLAMMATORY-DRUGS NAPROXEN)
P13 (EXAMPLE-OF NAPROXEN ALEVE)
P14 (POSSESS PEOPLE LOWER-RISK)
P15 (AND P16 P17)
P16 (OF LOWER-RISK COLORECTAL-CANCER)
P17 (OF LOWER-RISK ADENOMATOUS-POLYPS)

P5.S2. Most of these studies, however, are based on observations of people who took these medications for reasons such as treatment of arthritis or prevention of heart attacks.

P1 (BASE-ON STUDY OBSERVATION)
P2 (AMOUNT-OF STUDY MOST)
P3 (OF OBSERVATION PEOPLE)
P4 (TAKE PEOPLE MEDICATION)
P5 (FOR MEDICATION REASON)
P6 (AND P7 P9)
P7 (EXAMPLE-OF REASON ARTHRITIS)
P8 (OF TREATMENT ARTHRITIS)
P9 (EXAMPLE-OF REASON PREVENTION)
P10 (OF PREVENTION HEART-ATTACK)
P11 (HOWEVER P1)

P5.S3. Two recent studies have provided even stronger evidence regarding aspirin's ability to prevent the growth of polyps.

P1 (PROVIDE STUDY EVIDENCE)
P2 (MOD STUDY RECENT)
P3 (NUMBER-OF STUDY TWO)
P4 (MOD EVIDENCE STRONGER)
P5 (MOD STRONGER EVEN)
P6 (REGARD EVIDENCE ASPIRIN)
P7 (OF ASPIRIN ABILITY)
P8 (TO ABILITY PREVENT)
P9 (PREVENT $ GROWTH)
P10 (OF GROWTH POLYPS)

P5S4. The advantage of these recent studies is that people were randomly selected by the researchers to receive either aspirin or an inactive placebo.

P1 (OF ADVANTAGE STUDY)
P2 (MOD STUDY RECENT)
P3 (IS ADVANTAGE P4)
P4 (SELECT RESEARCHER PEOPLE)
P5 (MOD SELECT RANDOMLY)
P6 (OR P7 P8)
P7 (RECEIVE PEOPLE ASPIRIN)
P8 (RECEIVE PEOPLE INACTIVE-PLACEBO)

P5.S5. One study included people who were previously treated for early stages of colorectal cancer, and the other study included people who previously had polyps removed.

P1 (INCLUDE STUDY PEOPLE)
P2 (TREAT $ PEOPLE)

P3 (MOD TREAT PREVIOUSLY)
P4 (FOR TREAT COLORECTAL-CANCER)
P5 (OF EARLY-STAGE COLORECTAL-CANCER)
P6 (AND P1 P7)
P7 (INCLUDE STUDY PEOPLE)
P8 (MOD STUDY OTHER)
P9 (FROM POLYPS PEOPLE)
P10 (REMOVE P9)

P6.S1. But NSAIDs can cause serious or even life-threatening bleeding from stomach irritation, and currently available information suggests that the risks of serious bleeding outweigh the benefits of these medicines for the general public.

P1 (CAUSE NSAIDS BLEEDING)
P2 (MOD CAUSE POSSIBLE)
P3 (OR P4 P5)
P4 (MOD BLEEDING SERIOUS)
P5 (MOD BLEEDING LIFE-THREATENING)
P6 (MOD LIFE-THREATENING EVEN)
P7 (FROM BLEEDING STOMACH-IRRITATION)
P8 (AND P1 P9)
P9 (SUGGEST INFORMATION P10)
P10 (OUTWEIGH P11 P12)
P11 (OF RISK SERIOUS-BLEEDING)
P12 (OF BENEFIT MEDICINE)
P13 (FOR MEDICINE GENERAL-PUBLIC)

P6.S2. For this reason, experts do not recommend NSAIDs as a cancer-prevention strategy for people at average risk of developing colorectal cancer.

P1 (RECOMMEND EXPERT NSAIDS)
P2 (NEGATE P1)
P3 (AS NSAIDS STRATEGY)
P4 (MOD STRATEGY CANCER-PREVENTION)
P5 (FOR STRATEGY PEOPLE)
P6 (AT PEOPLE AVERAGE-RISK)
P7 (OF RISK P8)
P8 (DEVELOP $ COLORECTAL-CANCER)

P6.S3. However, the value of these drugs for people at increased colorectal cancer risk is being actively studied.

P1 (OF VALUE DRUG)
P2 (STUDY $ VALUE)
P3 (MOD STUDY ACTIVELY)
P4 (FOR DRUG PEOPLE)

P5 (AT PEOPLE COLORECTAL-CANCER-RISK)
P6 (MOD COLORECTAL-CANCER-RISK INCREASE)
P7 (HOWEVER P1)

P6S4. Celecoxib (Celebrex), has recently been approved by the FDA for reducing polyp formation in people with familial adenomatous polyposis.

P1 (APPROVE $ CELECOXIB)
P2 (REF CELECOXIB CELEBREX)
P3 (MOD APPROVE RECENTLY)
P4 (BY APPROVE FDA)
P5 (FOR APPROVE P6)
P6 (REDUCE $ POLYPS-FORMATION)
P7 (IN P8)
P8 (WITH PEOPLE FAMILIAL-ADENOMATOUS-POLYPOSIS)

P6.S5. The advantage of this drug is it does not cause bleeding from the stomach.

P1 (OF ADVANTAGE DRUG)
P2 (IS ADVANTAGE P3)
P3 (CAUSE DRUG BLEEDING)
P4 (NEGATE P3)
P5 (FROM BLEEDING STOMACH)

P7.S1. Hormone replacement therapy (HRT) in postmenopausal women may slightly reduce their risk of colorectal cancer.

P1 (LABEL HORMONE-REPLACEMENT-THERAPY HRT)
P2 (IN P1 WOMAN)
P3 (MOD WOMAN POSTMENOPAUSAL)
P4 (REDUCE WOMAN RISK)
P5 (MOD REDUCE SLIGHTLY)
P6 (MOD REDUCE POSSIBLE)
P7 (OF RISK COLORECTAL-CANCER)

P7.S2. HRT also lowers the risk of developing osteoporosis, but it may increase the risk of heart disease, blood clots, and breast and uterine cancer.

P1 (LOWER HRT RISK)
P2 (MOD LOWER ALSO)
P3 (OF RISK P4)
P4 (DEVELOP $ OSTEOPOROSIS)
P5 (BUT P1 P6)
P6 (INCREASE HRT RISK)
P7 (MOD INCREASE POSSIBLE)

P8 (AND (P9,P10,P11,P12))
P9 (OF RISK HEART-DISEASE)
P10 (OF RISK BLOOD-CLOT)
P11 (OF RISK BREAST-CANCER)
P12 (OF RISK UTERINE-CANCER)

P7.S3. The decision to use HRT should be based on discussion of benefits and risks with your doctor.

P1 (USE $ HRT)
P2 (TO DECISION P1)
P3 (BASE-ON DECISION DISCUSSION)
P4 (AND P5 P6)
P5 (OF DISCUSSION BENEFIT)
P6 (OF DISCUSSION RISK)
P7 (WITH DISCUSSION DOCTOR)

P8.S1. There are other risk factors that can't be controlled, such as a strong family history of colorectal cancer.

P1 (CONTROL $ RISK-FACTOR)
P2 (MOD RISK-FACTOR OTHER)
P3 (MOD CONTROL POSSIBLE)
P4 (NEGATE P1)
P5 (EXAMPLE-OF RISK-FACTOR FAMILY-HISTORY)
P6 (MOD FAMILY-HISTORY STRONG)
P7 (OF FAMILY-HISTORY COLORECTAL-CANCER)

P8.S2. But even when people have a history of colorectal cancer in their family, they may be able to prevent the disease.

P1 (OF HISTORY COLORECTAL-CANCER)
P2 (IN COLORECTAL-CANCER FAMILY)
P3 (POSSESS PEOPLE P1)
P4 (ABLE-TO PEOPLE P5)
P5 (PREVENT $ DISEASE)
P6 (BUT P1)
P7 (MOD POSSESS EVEN)

P8.S3. For example, people with a family history of colorectal cancer may benefit from starting screening tests when they are younger and having them done more often than people without this risk factor.

P1 (WITH PERSON FAMILY-HISTORY)
P2 (OF FAMILY-HISTORY COLORECTAL-CANCER)
P3 (BENEFIT PEOPLE SCREENING-TEST)

P4 (MOD BENEFIT POSSIBLE)
P5 (FROM BENEFIT SCREENING-TEST)
P6 (MOD SCREENING-TEST START)
P7 (WHEN START P8)
P8 (IS PERSON YOUNG)
P9 (DO PERSON SCREENING-TEST)
P10 (OFTEN P1 P11)
P11 (MOD OFTEN MORE)
P12 (WITHOUT PERSON RISK-FACTOR)

P9.S1. Genetic tests can help determine which members of certain families have inherited a high risk for developing colorectal cancer.

P1 (DETERMINE $ GENETIC-TEST)
P2 (MOD DETERMINE HELP)
P3 (DETERMINE P4 INHERIT)
P4 (OF MEMBER FAMILY)
P5 (MOD MEMBER CERTAIN)
P6 (INHERIT FAMILY RISK)
P7 (MOD RISK HIGH)
P8 (FOR RISK P9)
P9 (DEVELOP $ COLORECTAL-CANCER)

P9.S2. People with familial adenomatous polyposis (FAP) should start colonoscopy during their teens.

P1 (WITH PERSON FAMILIAL-ADENOMATOUS-POLYPOSIS)
P2 (LABEL FAMILIAL-ADENOMATOUS-POLYPOSIS FAP)
P3 (START P1 COLONOSCOPY)
P4 (DURING P1 TEENS)

P9.S3. Most doctors recommend they have their colon removed when they are in their twenties to prevent cancer from developing.

P1 (RECOMMEND DOCTOR P2)
P2 (AMOUNT-OF DOCTOR MOST)
P3 (REMOVE $ COLON)
P4 (IN P3 TWENTIES)
P5 (PREVENT P3 P6)
P6 (FROM CANCER DEVELOP)

P10.S1. The risk for people with hereditary nonpolyposis colon cancer (HNPCC) is not as great as for those with FAP.

P1 (FOR RISK PEOPLE)
P2 (WITH PEOPLE HEREDITARY-NONPOLYPOSIS-COLON-CANCER)
P3 (LABEL HEREDITARY-NONPOLYPOSIS-COLON-CANCER HNPCC)

P4 (GREATER-THAN P2 P5)
P5 (WITH PEOPLE FAP)
P6 (NEGATE P4)

P10.S2. Doctors recommend that people with HNPCC start colonoscopy screening during their twenties to remove any polyps and find any cancers at the earliest possible stage.

and find any cancers at the earliest possible stage.

P1 (RECOMMEND DOCTOR P3)
P2 (WITH PEOPLE HNPCC)
P3 (START P2 COLONSCOPY-SCREENING)
P4 (DURING COLONSCOPY-SCREENING TWENTIES)
P5 (IN-ORDER-TO COLONSCOPY-SCREENING P6)
P6 (AND P7 P8)
P7 (REMOVE $ POLYPS)
P8 (FIND $ CANCER)
P9 (MOD CANCER ANY)
P10 (AT P8 STAGE)
P11 (MOD STAGE EARLIEST)
P12 (MOD STAGE POSSIBLE)

P10.S3. But preventive removal of the colon is not usually suggested for people with HNPCC.

P1 (OF REMOVAL COLON)
P2 (MOD REMOVAL PREVENTATIVE)
P3 (SUGGEST $ P1)
P4 (FOR P1 P5)
P5 (WITH PEOPLE HNPCC)
P6 (BUT P1)

P11.S1. Ashkenazi Jews with the I1307K APC mutation have a slightly increased colorectal cancer risk, but do not develop these cancers when they are very young.

these cancers when they are very young.

P1 (WITH ASHKENAZI-JEW MUTATION)
P2 (LABEL MUTATION I1307K-APC)
P3 (POSSESS P1 COLORECTAL-CANCER-RISK)
P4 (AMOUNT-OF COLORECTAL-CANCER-RISK INCREASE)
P5 (MOD INCREASE SLIGHTLY)
P6 (BUT P1 P7)
P7 (NEGATE P8)
P8 (DEVELOP $ COLORECTAL-CANCER)
P9 (WHEN P8 YOUNG)
P10 (MOD YOUNG VERY)

P11.S2. For these reasons, most doctors recommend that they carefully follow the usual recommendations for colorectal cancer screening, but earlier or more frequent testing is usually not suggested.

P1 (RECOMMEND DOCTOR P4)
P2 (AMOUNT-OF DOCTOR MOST)
P3 (WITH ASHKENAZI-JEW MUTATION)
P4 (FOLLOW P3 RECOMMENDATION)
P5 (MOD RECOMMENDATION USUAL)
P6 (FOR RECOMMENDATION COLORECTAL-CANCER-SCREENING)
P7 (OR P10 P11)
P8 (SUGGEST $ TESTING)
P9 (BUT P10)
P10 (NEGATE P8)
P11 (MOD TESTING EARLIER)
P12 (MOD TESTING FREQUENT)

P12.S1. Since some colorectal cancers can't be prevented, finding them early is the best way to improve the chance of a cure and reduce the number of deaths caused by this disease.

P1 (PREVENT $ COLORECTAL-CANCER)
P2 (AMOUNT-OF COLORECTAL-CANCER SOME)
P3 (NEGATE P1)
P4 (FIND COLORECTAL-CANCER EARLY)
P5 (IS P4 BEST-WAY)
P6 (IN-ORDER-TO BEST-WAY P7)
P7 (AND P8 P10)
P8 (IMPROVE $ CURE)
P9 (OF CHANCE OF)
P10 (REDUCE $ DEATH)
P11 (OF NUMBER DEATH)
P12 (CAUSE-BY P7 DISEASE)

P13.S1. In addition to the screening recommendations for people at average colorectal cancer risk, the American Cancer Society has additional guidelines for people at moderate and high risk of colorectal cancer.

P1 (FOR SCREENING-RECOMMENDATION PEOPLE)
P2 (AT PEOPLE COLORECTAL-CANCER-RISK)
P3 (MOD COLORECTAL-CANCER-RISK AVERAGE)
P4 (POSSESS AMERICAN-CANCER-SOCIETY GUIDELINE)
P5 (MOD GUIDELINE ADDITIONAL)
P6 (FOR GUIDELINE PEOPLE)
P7 (AND P9 P10)
P8 (OF RISK COLORECTAL-CANCER)

P9 (MOD RISK MODERATE)
P10 (MOD RISK HIGH)
P11 (IN-ADDITION-TO P1 P4)

P13.S2. These recommendations are described in the section.

P1 (DESCRIBE $ RECOMMENDATION)
P2 (IN RECOMMENDATION SECTION)

P13.S3. Can Colorectal Polyps and Cancer Be Found Early?"

P1 (AND P2 P3)
P2 (FIND $ COLORECTAL-POLYPS)
P3 (FIND $ CANCER)
P4 (MOD FIND EARLY)
P5 (CAN P1)

P13.S4. Ask your doctor how these guidelines might apply to you.

P1 (ASK $ DOCTOR)
P2 (APPLY $ GUILDLINE)
P3 (MOD APPLY POSSIBLE)
P4 (TO APPLY PERSON)
P5 (HOW P4)

Web Page 5: Prostate cancer; Low Readability

P1.S1. Can diet help prevent prostate cancer?

P1 (PREVENT DIET PROSTATE-CANCER)
P2 (MOD PREVENT HELP)
P3 (CAN P1)

P1.S2. There is a great deal of circumstantial evidence that suggests that diet is a factor in predisposing men to prostate cancer.

P1 (SUGGEST P2 P4)
P2 (AMOUNT-OF EVIDENCE GREAT-DEAL)
P3 (MOD EVIDENCE CIRCUMSTANTIAL)
P4 (IS-A DIET FACTOR)
P5 (PREDISPOSE P3 MAN)
P6 (TO PREDISPOSE PROSTATE-CANCER)

P1.S3. However, we cannot be really sure that changing diet later in life can prevent prostate cancer developing.

P1 (SURE-THAT P3 P5)
P2 (NEGATE P1)
P3 (CHANGE $ DIET)
P4 (LATE-IN-LIFE P3)
P5 (PREVENT P3 PROSTATE-CANCER)
P6 (DEVELOPING $ PROSTATE-CANCER)

P1.S4. If it might make a difference, there are some easy changes one can make which may reduce the risk.

P1 (MAKE PERSON CHANGE)
P2 (REDUCE CHANGE RISK)
P3 (MOD REDUCE MAY)
P4 (MOD CHANGE EASY)
P5 (AMOUNT-OF CHANGE SOME)
P6 (MAKE P4 DIFFERENCE)
P7 (MOD P6 MIGHT)

P2.S1. The information that follows reflects some of the current thinking about the benefits of a balanced diet.

P1 (REFLECT INFORMATION CURRENT-THINKING)
P2 (AMOUNT-OF CURRENT-THINKING SOME)

P3 (ABOUT CURRENT-THINKING BALANCED-DIET)
P4 (OF BENEFIT BALANCED-DIET)
P5 (MOD INFORMATION FOLLOWING)

P2.S2. These benefits include building up resistance to prostate and other cancers.

P1 (INCLUDE BENEFIT P3)
P2 (BUILD-UP $ RESISTANCE)
P3 (AND P4 P5)
P4 (TO P2 PROSTATE-CANCER)
P5 (TO P2 OTHER-CANCER)

P2.S3. A balanced diet will also help to prevent other diseases such as heart disease.

P1 (TO HELP PREVENT)
P2 (PREVENT BALANCE-DIET OTHER-DISEASE)
P3 (EXAMPLE-OF OTHER-DISEASE HEART-DISEASE)

P3.S1. People in Far Eastern countries such as China and Japan are less likely than Westerners to develop cancer.

P1 (LESS-LIKELY P2 P5)
P2 (IN PERSON FAR-EASTERN-COUNTRY)
P3 (EXAMPLE-OF FAR-EASTERN-COUNTRY CHINA)
P4 (EXAMPLE-OF FAR-EASTERN-COUNTRY JAPAN)
P5 (IN PERSON WESTERN-COUNTRY)
P6 (DEVELOP P5 CANCER)

P3.S2. This difference may relate to the genetic differences between Easterners and Westerners.

P1 (BETWEEN GENETIC-DIFFERENCE EASTERNER WESTERNER)
P2 (RELATE DIFFERENCE P1)
P3 (MOD P2 POSSIBLE)

P3.S3. However, different risk levels may be related to other aspects of life in different parts of the world.

P1 (RELATE-TO RISK-LEVEL LIFE)
P2 (MOD RISK-LEVEL DIFFERENT)
P3 (MOD P1 POSSIBLE)
P4 (IN P5 P7)
P5 (OF ASPECT LIFE)
P6 (MOD ASPECT OTHER)

P7 (OF PART WORLD)
P8 (MOD PART DIFFERENT)
P9 (HOWEVER P1)

P4.S1. When people have migrated from Japan to the USA in the past, the rate of prostate cancer in their descendants has risen greatly.

P1 (MIGRATE $ PERSON)
P2 (FROM P1 JAPAN)
P3 (TO P2 USA)
P4 (MOD P1 IN-THE-PAST)
P5 (RISE RATE PROSTATE-CANCER)
P6 (AMOUNT-OF RISE GREATLY)
P7 (IN P5 PERSON-DESCENDANT)

P4.S2. Since their genetic make-up is largely similar, as they have married within their own ethnic communities, the new risk of prostate cancer is likely to be related to environmental factors.

P1 (SINCE P8 P2)
P2 (BECAUSE P3 P2)
P3 (MARRY-WITHIN PERSON-DESCENDANT ETHNIC-COMMUNITY)
P4 (POSSESS PERSON-DESCENDANT GENETIC-MAKE-UP)
P5 (MOD GENETIC-MAKE-UP SIMILAR)
P6 (OF RISK PROSTATE-CANCER)
P7 (MOD RISK NEW)
P8 (RELATE P5 ENVIRONMENTAL-FACTOR)
P9 (MOD P5 LIKELY)

P4.S3. Diet is an important one of these.

P1 (IS DIET ENVIRONMENTAL-FACTOR)
P2 (MOD ENVIRONMENTAL-FACTOR IMPORTANT)

P4.S4. Another pointer to reinforce this conclusion is that the levels of cancer in the East are rising.

P1 (IS P4 P2)
P2 (OF LEVEL CANCER)
P3 (RISE P1 IN-THE-EAST)
P4 (REINFORCE POINTER CONCLUSION)
P5 (MOD POINTER ANOTHER)

P4.S5. This coincides with changes in lifestyle - more people are living Western lifestyles and eating Western style foods.

P1 (IN LIFESTYLE CHANGE)
P2 (AND P3 P5)
P3 (LIVE PERSON WESTERN-LIFESTYLE)
P4 (AMOUNT-OF PERSON MORE)
P5 (EAT PERSON WESTERN-STYLE-FOOD)
P6 (COINCIDE $ P1)

P5.S1. Countries with a high level of heart disease such as the UK, often also have high levels of prostate cancer.

P1 (WITH COUNTRY HEART-DISEASE)
P2 (OF HIGH-LEVEL HEART-DISEASE)
P3 (EXAMPLE-OF COUNTRY U.K.)
P4 (OFTEN P1 P5)
P5 (POSSESS COUNTRY P5)
P6 (OF HIGH-LEVEL PROSTATE-CANCER)

P5.S2. These countries' cuisines often include high levels of fat, which has been shown to be linked to heart disease.

P1 (POSSESS COUNTRY CUISINE)
P2 (INCLUDE P1 FAT)
P3 (OF HIGH-LEVEL FAT)
P4 (MOD INCLUDE OFTEN)
P5 (LINK-TO P2 HEART-DISEASE)

P5.S3. It is thought that high levels of dietary fat may also create a higher risk of prostate cancer.

P1 (CREATE P2 P3)
P2 (MOD CREATE POSSIBLE)
P3 (OF HIGH-LEVEL DIETARY-FAT)
P4 (OF RISK PROSTATE-CANCER)
P5 (MOD RISK HIGHER)

P6.S1. Most people in the UK would benefit from reducing their fat intake as part of a healthy diet, and in particular, reducing their saturated fat intake.

P1 (IN PERSON U.K.)
P2 (AMOUNT-OF PERSON MOST)
P3 (REDUCE $ FAT-INTAKE)
P4 (BENEFIT-FROM P1 P3)
P5 (PART-OF P3 HEALTHY-DIET)
P6 (REDUCE $ SATURATED-FAT-INTAKE)
P7 (IN-PARTICULAR P6)

P7.S1. But how can you do this?

P1 (DO PERSON THIS)
P2 (HOW P1)
P3 (BUT P2)

P8.S1. The best thing to do is avoid cooking methods that need fat such as frying and roasting.

P1 (REF BEST-THING P2)
P2 (AVOID $ COOKING-METHOD)
P3 (NEED COOKING-METHOD FAT)
P4 (AND P5 P6)
P5 (EXAMPLE-OF COOKING-METHOD FRYING)
P6 (EXAMPLE-OF COOKING-METHOD ROASTING)

P8.S2. Instead, you could try baking, steaming, poaching, or grilling foods.

P1 (INSTEAD P2)
P2 (OR (P3,P4,P5,P6))
P3 (TRY $ BAKING)
P4 (TRY $ STEAMING)
P5 (TRY $ POACHING)
P6 (TRY $ GRILLING)

P8.S3. You should also spread margarine or butter thinly on bread; and avoid adding fat to food when cooking.

P1 (AND P2 P7)
P2 (OR P2 P3)
P3 (SPREAD-ON MARGARINE BREAD)
P4 (SPREAD-ON BUTTER BREAD)
P5 (MOD SPREAD-ON THINLY)
P6 (ADD-TO FAT FOOD)
P7 (AVOID-WHEN P6 COOKING)

P9.S1. Try to eat more fish and poultry.

P1 (TRY P2)
P2 (AND P3 P4)
P3 (EAT $ FISH)
P4 (EAT $ POULTRY)
P5 (AMOUNT-OF FISH MORE)
P6 (AMOUNT-OF POULTRY MORE)

P9.S2. Avoid having red meat more than 2-3 times a week.

P1 (AVOID P2)
P2 (HAVE $ RED-MEAT)
P3 (DURATION-OF P2 MORE-THAN-2-3-TIMES/WK)

P9.S3. Cut off any fat on the meat.

P1 (CUT-OFF $ FAT)
P2 (MOD FAT ANY)
P3 (ON P1 MEAT)

P9.S4. Avoid high fat snacks such as crisps, biscuits and chocolate.

P1 (AVOID $ SNACK)
P2 (MOD SNACK HIGH-AT)
P3 (AND (P4,P5,P6))
P4 (EXAMPLE-OF SNACK CRISP)
P5 (EXAMPLE-OF SNACK BISCUIT)
P6 (EXAMPLE-OF SNACK CHOCOLATE)

P10.S1. If you maintain a normal weight for your height then you are probably eating a healthy level of dietary fat.

P1 (MAINTAIN $ NORMAL-WEIGHT)
P2 (FOR P1 HEIGHT)
P3 (EAT $ P5)
P4 (MOD EAT PROBABLY)
P5 (OF HEALTHY-LEVEL DIETARY-FAT)
P6 (IF-THEN P1 P3)

P11.S1. Antioxidants are substances that may help in the prevention of prostate cancer.

P1 (IS-A ANTIOXIDANT SUBSTANCE)
P2 (HELP ANTIOXIDANT PREVENTION)
P3 (MOD HELP POSSIBLE)
P4 (OF PREVENTION PROSTATE-CANCER)

P11.S2. Antioxidants may help to prevent the damage caused by free radicals.

P1 (CAUSE FREE-RADICAL DAMAGE)
P2 (PREVENT ANTIOXIDANT P1)
P3 (MOD PREVENT POSSIBLE)

P11.S3. Free radicals are harmful molecules that occur in the body.

P1 (IS-A FREE-RADICAL MOLECULE)
P2 (MOD MOLECULE HARMFUL)
P3 (OCCUR-IN MOLECULE BODY)

P11.S4. They can cause cell damage and can lead to diseases such as cancer and heart disease.

P1 (CAUSE FREE-RADICAL CELL-DAMAGE)
P2 (LEAD-TO FREE-RADICAL DISEASE)
P3 (AND P4 P5)
P4 (EXAMPLE-OF DISEASE CANCER)
P5 (EXAMPLE-OF DISEASE HEART-DISEASE)

P11.S5. Vitamin C is a well-known antioxidant.

P1 (IS-A VITAMIN-C ANTIOXIDANT)
P2 (MOD ANTIOXIDANT WELL-KNOWN)

P11.S6. Others include vitamin E and selenium.

P1 (REF OTHER ANTIOXIDANT)
P2 (AND P3 P4)
P3 (EXAMPLE-OF P1 VITAMIN-E)
P4 (EXAMPLE-OF P1 SELENIUM)

P11.S7. Antioxidants are found in many fruits and vegetables.

P1 (AND P2 P3)
P2 (FOUND-IN ANTIOXIDANT FRUIT)
P3 (FOUND-IN ANTIOXIDANT VEGETABLE)
P4 (AMOUNT-OF FRUIT MANY)
P5 (AMOUNT-OF VEGETABLE MANY)

P12.S1. Our knowledge on the role of these nutrients in prostate cancer is steadily growing – there has been some evidence showing a reduction in prostate cancer deaths when vitamin E was given to Finnish men who smoke, as a dietary supplement.

P1 (GROW $ KNOWLEDGE)
P2 (MOD GROW STEADILY)

P3 (ON KNOWLEDGE NUTRIENT)
P4 (OF ROLE NUTRIENT)
P5 (IN NUTRIENT PROSTATE-CANCER)
P6 (WHEN P9 P10)
P7 (SHOW EVIDENCE PROSTATE-CANCER-DEATH)
P8 (IN REDUCTION PROSTATE-CANCER-DEATH)
P9 (GIVE-TO VITAMIN-E FINNISH-MAN)
P10 (MOD FINNISH-MAN SMOKING)
P11 (AS VITAMIN-E DIETARY-SUPPLEMENT)

P13.S1. We are a long way from recommending routine Vitamin E supplementation for all men, as other studies have not shown the same benefits.

P1 (FROM LONG-WAY P2)
P2 (RECOMMEND $ VITAMIN-E-SUPPLMENTATION)
P3 (AMOUNT-OF VITAMIN-E-SUPPLMENTATION ROUTINE)
P4 (FOR P2 MAN)
P5 (AS P1 P10)
P6 (AMOUNT-OF MAN ALL)
P7 (SHOW STUDY BENEFIT)
P8 (MOD SHOW OTHER)
P9 (MOD BENEFIT SAME)
P10 (NEGATE P8)

P13.S2. There are also drawbacks.

P1 (EXIST $ DRAWBACK)
P2 (ALSO P1)

P13.S3. Some studies have shown more deaths from strokes in those groups taking vitamin supplements.

P1 (SHOW STUDY DEATH)
P2 (MOD STUDY SOME)
P3 (AMOUNT-OF DEATH MORE)
P4 (IN P6 P7)
P5 (FROM DEATH STROKE)
P6 (TAKE GROUP VITAMIN-SUPPLEMENT)

P14.S1. Selenium supplementation was also found to reduce prostate cancer in a small group of men, but again the evidence is not yet strong enough to recommend routine supplementation.

P1 (REDUCE SELENIUM-SUPPLEMENTATION PROSTATE-CANCER)
P2 (IN P1 MEN)

P3 (OF GROUP MEN)
P4 (MOD GROUP SMALL)
P5 (BUT P6)
P6 (TO P7 P9)
P7 (IS EVIDENCE STRONG)
P8 (NEGATE P7)
P9 (RECOMMEND $ ROUTINE-SUPPLEMENTATION)

P15.S1. Rich sources of vitamin E are vegetable oils and spreads (but used sparingly to help maintain a low fat intake) and nuts.

P1 (OF SOURCE VITAMIN-E)
P2 (MOD SOURCE RICH)
P3 (AND (P4,P5,P6))
P4 (EXAMPLE-OF P1 VEGETABLE-OIL)
P5 (EXAMPLE-OF P1 SPREAD)
P6 (EXAMPLE-OF P1 NUT)
P7 (USE $ SPREAD)
P8 (MOD USE SPARINGLY)
P9 (MAINTAIN P6 LOW-FAT-INTAKE)

P15.S2. Selenium can be found in fish, lentils and Brazil nuts, all of which can be incorporated into a healthy diet.

P1 (AND (P2,P3,P4))
P2 (FIND-IN SELENIUM FISH)
P3 (FIND-IN SELENIUM LENTIL)
P4 (FIND-IN SELENIUM BRAZIL-NUT)
P5 (INCORPORATE P1 DIET)
P6 (MOD DIET HEALTHY)
P7 (MOD INCORPORATE POSSIBLE)

P16.S1. A twelve-year study has just begun in the US to explore the links between Vitamin E and Selenium supplementation and a reduced risk of prostate cancer.

P1 (AND VITAMIN-E SELENIUM-SUPPLEMENTATION)
P2 (BETWEEN LINK P1 P3)
P3 (OF RISK PROSTATE-CANCER)
P4 (AMOUNT-OF RISK REDUCE)
P5 (STUDY EXPLORE P2)
P6 (DURATION-OF STUDY TWELVE-YEAR)
P7 (BEGIN-IN STUDY U.S.)

P16.S2. In the mean time it is probably a good idea to ensure that your diet includes antioxidants.

P1 (INCLUDE DIET ANTIOXIDANT)
P2 (REF IDEA P1)
P3 (MOD IDEA GOOD)
P4 (MOD GOOD PROBABLY)
P5 (ENSURE P1)

P17.S1. The most important source for all antioxidants is fruit and vegetables – five or more portions a day is a good target.

P1 (FOR SOURCE ANTIOXIDANT)
P2 (AMOUNT-OF ANTIOXIDANT ALL)
P3 (MOD SOURCE IMPORTANT)
P4 (DEGREE-OF IMPORTANT MOST)
P5 (REF P1 P6)
P6 (AND FRUIT VEGETABLE)
P7 (AMOUNT-OF P6 FIVE-OR-MORE-PORTIONS/DAY)
P8 (IS FIVE-OR-MORE-PORTIONS/DAY TARGET)
P9 (MOD TARGET GOOD)

P17.S2. This could be fresh, frozen or tinned fruit and vegetables.

P1 (AND P2 P3)
P2 (OR (P4,P5,P6))
P3 (OR (P7,P8,P9))
P4 (MOD FRUIT FRESH)
P5 (MOD FRUIT FROZEN)
P6 (MOD FRUIT TIN)
P7 (MOD VEGETABLE FRESH)
P8 (MOD VEGETABLE FROZEN)
P9 (MOD VEGETABLE TIN)
P10 (MOD P1 POSSIBLE)

P17.S3.1. One piece of fruit, or a slice of very large fruit (e.g., pineapple);

P1 (OR P2 P4)
P2 (OF PIECE FRUIT)
P3 (NUMBER-OF P1 ONE)
P4 (OF SLICE FRUIT)
P5 (MOD FRUIT LARGE)
P6 (MODE LARGE VERY)
P7 (NUMBER-OF SLICE ONE)
P8 (EXAMPLE-OF FRUIT PINEAPPLE)

P9 (REF P1 PORTION)

P17.S3.2. Two pieces of a small fruit,

P1 (OF PIECE FRUIT)
P2 (NUMBER-OF PIECE TWO)
P3 (MOD FRUIT SMALL)
P4 (REF P1 PORTION)

P17.S3.3. One cupful of grapes or raspberries,

P1 (OR P2 P3)
P2 (OF CUPFUL GRAPE)
P3 (OF CUPFUL RASPBERRY)
P4 (NUMBER-OF CUPFUL ONE)
P5 (REF P1 PORTION)

P17.S3.4. Two tablespoons of vegetables, raw, cooked or frozen.

P1 (OF TABLESPOON VEGETABLE)
P2 (NUMBER-OF TABLESPOON SPOON)
P3 (OR (P4,P5,P6))
P4 (MOD VEGETABLE RAW)
P5 (MOD VEGETABLE COOK)
P6 (MOD VEGETABLE FROZEN)
P7 (REF P1 PORTION)

P18.S1. Count up what you are eating now.

P1 (EAT PERSON)
P2 (TIME EAT NOW)
P3 (COUNT WHAT P1)

P18.S2. If it is less than five portions a day add in more and gradually try to build it up.

P1 (A PORTION DAY)
P2 (NUMBER-OF PORTION LESS-THAN-FIVE)
P3 (IF P2 P4)
P4 (AND P5 P6)
P5 (ADD PORTION MORE)
P6 (TRY BUILD-UP PORTION)
P7 (MOD TRY GRADUALLY)

P19.S3. Sorry, but potatoes do not count!

P1 (COUNT $ POTATO)
P2 (NEGATE P1)
P3 (BUT P2)
P4 (SORRY P3)

P19.S1. Tomatoes have had a good press recently in relation to the prevention of prostate cancer.

P1 (POSSESS TOMATO PRESS)
P2 (MOD PRESS GOOD)
P3 (TIME P1 RECENTLY)
P4 (IN-RELATION-TO P1 P5)
P5 (OF PREVENTION PROSTATE-CANCER)

P19.S2. This is because they contain not only vitamins C and E, but also another antioxidant called lycopene.

P1 (AND P2 P2)
P2 (CONTAIN TOMATO VITAMIN-C)
P3 (CONTAIN TOMATO VITAMIN-E)
P4 (MOD P1 ONLY)
P5 (NEGATE P4)
P6 (CONTAIN TOMATO ANTIOXIDANT)
P7 (REF ANTIOXIDANT LYCOPENE)
P8 (MOD ANTIOXIDANT ANOTHER)
P9 (MOD ANOTHER ALSO)
P10 (BUT P9)

P19.S3. Lycopene is the compound that gives tomatoes their red colour.

P1 (REF LYCOPENE COMPOUND)
P2 (GIVE LYCOPENE P3)
P3 (POSSESS TOMATO COLOUR)
P4 (MOD COLOUR RED)

P19.S4. There is some evidence from the USA that men who eat the most processed tomatoes, such as found on pizza toppings and pasta sauces, are less likely to develop prostate cancer.

P1 (EAT MAN PROCESS-TOMATO)
P2 (EXAMPLE-OF PROCESS-TOMATO P3)

P3 (AND P4 P5)
P4 (FIND-ON PROCESS-TOMATO PIZZA-TOPPING)
P5 (FIND-ON PROCESS-TOMATO PASTA-SAUCE)
P6 (DEVELOP $ PROSTATE-CANCER)
P7 (LESS-LIKELY P1 P6)
P8 (FROM EVIDENCE USA)
P9 (AMOUNT-OF EVIDENCE SOME)
P10 (REF EVIDENCE P7)

P19.S5. It may therefore be a good idea, whilst awaiting further evidence, to try and increase your tomato consumption to one portion a day – this may take any form, either fresh ripe tomatoes or processed tomatoes e.g. tomato juice, soup, tinned chopped or plum tomatoes.

P1 (INCREASE $ TOMATO-CONSUMPTION)
P2 (TO P1 PORTION/DAY)
P3 (NUMBER-OF PORTION/DAY ONE)
P4 (FORM-OF P5 PORTION)
P5 (OR P5 P8)
P6 (REF TOMATO-CONSUMPTION TOMATO)
P7 (MOD TOMATO RIPE)
P8 (MOD TOMATO FRESH)
P9 (REF TOMATO-CONSUMPTION PROCESS-TOMATO)
P10 (OR (P1,P12,P13,P14))
P11 (EXAMPLE-OF PROCESS-TOMATO TOMATO-JUICE)
P12 (EXAMPLE-OF PROCESS-TOMATO SOUP)
P13 (EXAMPLE-OF PROCESS-TOMATO TIN-CHOPPED)
P14 (EXAMPLE-OF PROCESS-TOMATO PLUM-TOMATO)
P15 (REF IDEA P1)
P16 (MOD IDEA GOOD)
P17 (MOD GOOD POSSIBLE)
P18 (AWAIT $ EVIDENCE)
P19 (MOD EVIDENCE FURTHER)
P20 (WHILE P1 P18)

P19.S6. Processed tomato products, such as ketchup, are particularly rich in a form of lycopene that is easier for the body to metabolise.

P1 (OF FORM LYCOPENE)
P2 (RICH PROCESS-TOMATO-PRODUCTS P1)
P3 (MOD RICH PARTICULARLY)
P4 (EXAMPLE-OF PROCESS-TOMATO-PRODUCTS KETCHUP)
P5 (EASIER LYCOPENE METABOLISE)
P6 (FOR P5 BODY)

P19.S7. Tomatoes can also contribute to your five portions of fruit or vegetables a day!

P1 (CONTRIBUTE-TO TOMATO PORTION)
P2 (OR P3 P4)
P3 (OF PORTION VEGETABLE)
P4 (OF PORTION FRUIT)
P5 (NUMBER-OF PORTION FIVE)
P6 (PER PORTION DAY)

P20.S1. Should you take supplements of lycopene?

P1 (TAKE $ SUPPLEMENT)
P2 (OF SUPPLEMENT LYCOPENE)
P3 (SHOULD P1)

P20.S2. The research on lycopene is at very early stages and we may find out that there are other nutrients in tomatoes that are also beneficial.

P1 (AND P2 P5)
P2 (ON RESEARCH LYCOPENE)
P3 (AT RESEARCH EARLY-STAGE)
P4 (MOD EARLY-STAGE VERY)
P5 (FIND $ NUTRIENT)
P6 (MOD FIND POSSIBLE)
P7 (IN NUTRIENT TOMATO)
P8 (MOD NUTRIENT OTHER)
P9 (MOD NUTRIENT BENEFICIAL)

P20.S3. Therefore it is recommended that you try to increase your overall intake of fruit and vegetables including tomatoes rather than take supplements.

P1 (RATHER-THAN P2 P7)
P2 (INCREASE $ OVERALL-INTAKE)
P3 (AND P3 P4)
P4 (OF OVERALL-INTAKE FRUIT)
P5 (OF OVERALL-INTAKE VEGETABLE)
P6 (INCLUDE OVERALL-INTAKE TOMATO)
P7 (TAKE $ SUPPLEMENT)
P8 (RECOMMEND P1)

P21.S1. Soya beans have also been in the news recently due to their anti-cancer potential.

P1 (BECAUSE P2 P4)

P2 (IN SOYA-BEAN NEWS)
P3 (TIME P1 RECENTLY)
P4 (POSSESS SOYA-BEAN ANTI-CANCER-POTENTIAL)

P21.S2. Soya protein can be found in many foods including alternatives to meat such as tofu and non-dairy products such as soya yogurts and soya milk.

P1 (FIND SOYA-PROTEIN FOOD)
P2 (AMOUNT-OF FOOD MANY)
P3 (AND P4 P6)
P4 (INCLUDE FOOD MEAT-ALTERNATIVE)
P5 (EXAMPLE-OF MEAT-ALTERNATIVE TOFU)
P6 (INCLUDE FOOD NON-DAIRY-PRODUCT)
P7 (AND P7 P9)
P8 (EXAMPLE-OF NON-DAIRY-PRODUCT SOYA-YOGURT)
P9 (EXAMPLE-OF NON-DAIRY-PRODUCT SOYA-MILK)

P21.S3. The beans from which the protein is extracted contain compounds called isoflavanoids.

P1 (EXTRACT-FROM PROTEIN BEANS)
P2 (CONTAIN BEAN COMPOUND)
P3 (REF COMPOUND ISOFLAVANOID)

P21.S4. The most well known of these is genistein.

P1 (IS GENISTEIN ISOFLAVANOID)
P2 (MOD ISOFLAVANOID MOST-WELL-KNOWN)

P21.S5. This has been shown to inhibit prostate cancer cell growth in the laboratory.

P1 (INHIBIT GENISTEIN CELL-GROWTH)
P2 (MOD CELL-GROWTH PROSTATE-CANCER)
P3 (IN P1 LABORATORY)

P21.S6. Unfortunately, there is no evidence that this happens in humans, but overall inclusion of soya into the diet will not do any harm.

P1 (OCCUR-IN EVIDENCE HUMAN)
P2 (NEGATE P1)
P3 (OF INCLUSION SOYA)
P4 (MOD INCLUSION OVERALL)
P5 (IN SOYA DIET)
P6 (DO P3 HARM)
P7 (NEGATE P5)

P21.S7. It can easily be added to your diet as a glass of soya milk or soya yogurt.

P1 (ADD-TO SOYA DIET)
P2 (MOD ADD-TO EASILY)
P3 (OR P4 P5)
P4 (EXAMPLE-OF P1 GLASS-SOYA-MILK)
P5 (EXAMPLE-OF P1 SOYA-YOGURT)

P21.S8. You could eat tofu as an alternative to red meat at an evening meal.

P1 (EAT $ TOFU)
P2 (IS-A P1 ALTERNATIVE)
P3 (TO ALTERNATIVE RED-MEAT)
P4 (AT P1 EVENING-MEAL)

P21.S9. Any increase from a minimal intake will help widen your food choices.

P1 (WIDEN INCREASE FOOD-CHOICE)
P2 (MOD INCREASE ANY)
P3 (FROM INCREASE MINIMAL-INTAKE)

P22.S1. Maintain a normal weight for your height;

P1 (MAINTAIN $ WEIGHT)
P2 (MOD WEIGHT NORMAL)
P3 (FOR P1 HEIGHT)

P22.S2. Avoid fatty foods and try to decrease your fat intake;

P1 (AND P2 P3)
P2 (AVOID $ FATTY-FOOD)
P3 (DECREASE $ FAT-INTAKE)

P22.S3. Eat red meat and processed meat in moderation;

P1 (AND P2 P3)
P2 (EAT $ RED-MEAT)
P3 (EAT $ PROCESS-MEAT)
P4 (MOD P1 IN-MODERATION)

P22.S4. Include at least five portions of fruit and vegetables per day, including a regular intake of tomatoes;

P1 (INCLUDE $ PORTION)
P2 (AMOUNT-OF PORTION AT-LEAST-FIVE-PER-DAY)
P3 (AND P4 P5)
P4 (OF PORTION FRUIT)
P5 (OF PORTION VEGETABLE)
P6 (INCLUDE P7)
P7 (OF INTAKE TOMATO)
P8 (MOD INTAKE REGULAR)

P22.S5. Perhaps include soya products within your diet as an occasional food product;

P1 (INCLUDE-IN SOYA-PRODUCT DIET)
P2 (AS SOYA-PRODUCT FOOD-PRODUCT)
P3 (MOD FOOD-PRODUCT OCCASIONAL)
P4 (MOD INCLUDE PERHAPS)

P22.S6. Do not take high doses of vitamin supplements;

P1 (TAKE $ HIGH-DOSE)
P2 (OF DOSE VITAMIN-SUPPLEMENT)
P3 (MOD DOSE HIGH)
P4 (NEGATE P1)

P22.S7. Drink alcohol in moderation.

P1 (DRINK $ ALCOHOL)
P2 (IN P1 MODERATION)

P22.S8. A maximum of three units per day for men (one unit = a pub measure of spirits, a small glass of wine or half a pint of standard beer).

P1 (OF MAXIMUM UNIT)
P2 (NUMBER-OF UNIT THREE)
P3 (PER UNIT DAY)
P4 (FOR P1 MAN)
P5 (REF UNIT P6)
P6 (OR (P7,P8,P10))
P7 (OF PUB-MEASURE SPIRIT)
P8 (OF GLASS WINE)
P9 (MOD GLASS SMALL)
P10 (OF PINT STANDARD-BEER)
P11 (AMOUNT-OF PINT HALF)

Web Page 6: Prostate cancer; High Readability

P1.S1. Determining whether you have prostate cancer generally involves a series of tests and exams.

P1 (DETERMINE P2)
P2 (POSSESS PERSON PROSTATE-CANCER)
P3 (AND P4 P5)
P4 (INVOLVE $ TEST)
P5 (INVOLVE $ EXAM)
P6 (OF SERIES TEST)
P7 (OF SERIES EXAM)

P1.S2. Before starting the testing process, your physician may ask you questions about your medical history, your family history of cancer and any symptoms you may be having, particularly problems with urination.

P1 (BEFORE P2)
P2 (START $ TESTING-PROCESS)
P3 (ASK PHYSICIAN QUESTION)
P4 (MOD ASK POSSIBLE)
P5 (AND (P6,P7,P9))
P6 (ABOUT QUESTION MEDICAL-HISTORY)
P7 (ABOUT QUESTION FAMILY-HISTORY)
P8 (OF FAMILY-HISTORY CANCER)
P9 (ABOUT QUESTION SYMPTOM)
P10 (POSSESS PERSON SYMPTOM)
P11 (EXAMPLE-OF SYMPTOM P11)
P12 (WITH PROBLEM URINATION)
P13 (MOD PROBLEM PARTICULARLY)

P1.S3. Then, your doctor will most likely proceed to one or more of the tests described below.

P1 (PROCEED $ DOCTOR)
P2 (TO PROCEED TEST)
P3 (MOD PROCEED MOST-LIKELY)
P4 (AMOUNT-OF TEST ONE-OR-MORE)
P5 (DESCRIBE TEST BELOW)
P6 (THEN P1)

P2.S1. Because the prostate lies in front of the rectum, your physician can feel the prostate by inserting a gloved, lubricated finger into the rectum.

P1 (LAY $ PROSTATE)

P2 (IN LAY RECTUM)
P3 (OF FRONT RECTUM)
P4 (FEEL PHYSICIAN PROSTATE)
P5 (BY P4 P6)
P6 (INSERT PHYSICIAN FINGER)
P7 (MOD FINGER GLOVED)
P8 (MOD FINGER LUBRICATED)
P9 (INTO FINGER RECTUM)
P10 (BECAUSE P1)

P2.S2. This simple procedure is called a digital rectal examination (DRE).

P1 (REF PROCEDURE DIGITAL-RECTAL-EXAMINATION)
P2 (LABEL DIGITAL-RECTAL-EXAMINATION DRE)
P3 (MOD PROCEDURE SIMPLE)

P2.S3. It allows your physician to determine whether the prostate is enlarged or has lumps or other types of abnormal texture.

P1 (ALLOW DRE PHYSICIAN)
P2 (DETERMINE PHYSICIAN P3)
P3 (OR (P4,P5,P6))
P4 (IS PROSTATE ENLARGED)
P5 (POSSESS PROSTATE LUMP)
P6 (POSSESS PROSTATE TEXTURE)
P7 (MOD TEXTURE ABNORMAL)
P8 (OF TYPE TEXTURE)
P9 (MOD TYPE OTHER)

P3.S1. Used in addition to the DRE, a PSA test increases the likelihood of prostate cancer detection.

P1 (USE $ DRE)
P2 (IN-ADDITION-TO P3 P1)
P3 (INCREASE PSA DETECTION)
P4 (MOD DETECTION PROSTATE-CANCER)
P5 (OF LIKELIHOOD PROSTATE-CANCER)

P3.S2. PSA is the abbreviation for prostate-specific antigen, a substance produced by the prostate cells.

P1 (IS PSA ABBREVIATION)
P2 (FOR ABBREVIATION PROSTATE-SPECIFIC-ANTIGEN)
P3 (IS PROSTATE-SPECIFIC-ANTIGEN SUBSTANCE)
P4 (PRODUCE PROSTATE-CELL SUBSTANCE)

P3.S3. A PSA test measures the level of PSA in the bloodstream and is reported as nanograms per milliliter, or ng/mL.

P1 (MEASURE PSA-TEST PSA)
P2 (OF LEVEL PSA)
P3 (IN PSA BLOODSTREAM)
P4 (OR P5 P6)
P5 (REPORT P2 NANOGRAM-PER-MILLILITER)
P6 (REPORT P2 NG/ML)

P3.S4. Very little PSA escapes from a healthy prostate into the bloodstream, but certain prostate conditions can cause larger amounts of PSA to leak into the blood.

P1 (ESCAPE $ PSA)
P2 (MOD PSA LITTLE)
P3 (MOD LITTLE VERY)
P4 (FROM P1 PROSTATE)
P5 (MOD PROSTATE HEALTHY)
P6 (INTO P1 BLOODSTREAM)
P7 (BUT P1 P8)
P8 (CAUSE PROSTATE-CONDITION P10)
P9 (MOD PROSTATE-CONDITION CERTAIN)
P10 (LEAK PSA BLOOD)
P11 (OF PSA AMOUNT)
P12 (MOD AMOUNT LARGE)

P4.S1. 1) a benign noncancerous enlargement of the prostate called benign prostatic hyperplasia (BPH)

P1 (OF ENLARGEMENT PROSTATE)
P2 (MOD PROSTATE NONCANCEROUS)
P3 (MOD PROSTATE BENIGN)
P4 (REF P1 BENIGN-PROSTATIC-HYPERPLASIA)
P5 (LABEL BENIGN-PROSTATIC-HYPERPLASIA BPH)

P5.S1. A high level of PSA in the bloodstream is a warning sign that prostate cancer may be present.

P1 (OF PSA LEVEL)
P2 (MOD LEVEL HIGH)
P3 (IN PSA BLOODSTREAM)
P4 (IS P1 WARNING-SIGN)
P5 (IS WARNING-SIGN P6)
P6 (PRESENT $ PROSTATE-CANCER)

P7 (MOD PRESENT POSSIBLE)

P5.S2. But since other kinds of prostate disease can also cause high PSA levels, PSA testing by itself cannot confirm the presence of prostate cancer.

P1 (OF KIND PROSTATE-DISEASE)
P2 (MOD KIND OTHER)
P3 (CAUSE P1 PSA-LEVEL)
P4 (MOD PSA-LEVEL HIGH)
P5 (SINCE P6 P1)
P6 (CONFIRM PSA-TEST PROSTATE-DISEASE)
P7 (OF PRESENCE PROSTATE-DISEASE)
P8 (MOD PSA-TEST ALONE)

P5.S3. A high PSA level only indicates the possibility of prostate cancer and the need for additional evaluation by your physician.

P1 (INDICATE PSA-LEVEL PROSTATE-CANCER)
P2 (MOD INDICATE ONLY)
P3 (OF POSSIBILITY PROSTATE-CANCER)
P4 (INDICATE PSA-LEVEL P5)
P5 (FOR NEED EVALUATION)
P6 (BY EVALUATION PHYSICIAN)
P7 (MOD EVALUATION ADDITIONAL)

P5.S4. Conversely, a low PSA level does not always mean that prostate cancer is not present.

P1 (MEAN PSA-LEVEL P4)
P2 (MOD PSA-LEVEL LOW)
P3 (NEGATE P1)
P4 (IS PROSTATE-CANCER PRESENT)
P5 (NEGATE P4)
P6 (CONVERSELY P1)

P6.S1. According to the American Cancer Society, men aged 50 and older, and those over the age of 45 who are in high-risk groups, blood test and digital rectal exam (DRE) once every year.
prostate cancer, should have a prostate-specific antigen (PSA) such as African-American men and men with a family history of

P1 (MOD AMERICAN-CANCER-SOCIETY ACCORDING-TO)
P2 (AGE MEN 50-AND-OLDER)
P3 (AGE MEN AGE-45)
P4 (IN AGE-45 HIGH-RISK-GROUP)

P5 (EXAMPLE-OF HIGH-RISK-GROUP AFRICAN-AMERICAN)
P6 (EXAMPLE-OF HIGH-RISK-GROUP P7)
P7 (WITH MEN FAMILY-HISTORY)
P8 (OF FAMILY-HISTORY PROSTATE-CANCER)
P9 (AND P10 P11)
P10 (POSSESS MEN PSA-BLOOD-TEST)
P11 (POSSESS MEN DIGITAL-RECTAL-EXAM)
P12 (DURATION-OF PSA-BLOOD-TEST ONCE-EVERY-YEAR)
P13 (DURATION-OF DIGITAL-RECTAL-EXAM ONCE-EVERY-YEAR)

P6.S2. Any man who develops persistent urinary symptoms should contact his physician.

P1 (DEVELOP MAN URINARY-SYMPTOM)
P2 (MOD MAN ANY)
P3 (MOD URINARY-SYMPTOM PERSISTENT)
P4 (SHOULD P1 P5)
P5 (CONTACT MAN PHYSICIAN)

P7.S1. Transrectal Ultrasound (TRUS) is the use of soundwaves to create an image of the prostate.

P1 (REF TRANSRECTAL-ULTRASOUND P3)
P2 (LABEL TRANSRECTAL-ULTRASOUND TRUS)
P3 (OF USE SOUNDWAVE)
P4 (CREATE SOUNDWAVE IMAGE)
P5 (OF IMAGE PROSTATE)

P7.S2. As the waves bounce off the prostate, they create a pattern that is converted into a picture by a computer.

P1 (BOUNCE-OFF WAVE PROSTATE)
P2 (CREATE WAVE PATTERN)
P3 (CONVERT $ PATTERN)
P4 (INTO PATTERN PICTURE)
P5 (CONVERT COMPUTER PATTERN)

P7.S3. TRUS is used to detect abnormal prostate growth and to guide a biopsy of the abnormal prostate area.

P1 (USE $ TRUS)
P2 (AND P3 P5)
P3 (DETECT TRUS PROSTATE-GROWTH)
P4 (MOD PROSTATE-GROWTH ABNORMAL)
P5 (GUIDE TRUS BIOPSY)
P6 (OF BIOPSY PROSTATE-AREA)

P7 (MOD PROSTATE-AREA ABNORMAL)

P8.S1. A biopsy is the removal of a sample of tissue, which is then examined under a microscope to check for cancerous changes.

P1 (IS BIOPSY P2)
P2 (OF REMOVAL TISSUE)
P3 (OF SAMPLE TISSUE)
P4 (EXAMINE $ SAMPLE)
P5 (UNDER P4 MICROSCOPE)
P6 (CHECK SAMPLE CHANGE)
P7 (MOD CHANGE CANCEROUS)
P8 (AMOUNT-OF TISSUE-SAMPLE MUTLIPLE)

P8.S2. Only a biopsy can definitely confirm prostate cancer.

P1 (CONFIRM BIOPSY PROSTATE-CANCER)
P2 (MOD BIOPSY ONLY)
P3 (MOD CONFIRM DEFINITELY)

P9.S1. Typically, the physician takes multiple tissue samples for biopsy.

P1 (TAKE PHYSICIAN TISSUE-SAMPLE)
P2 (MOD TAKE TYPICALLY)
P3 (AMOUNT-OF TISSUE-SAMPLE MUTLIPLE)
P4 (FOR TISSUE-SAMPLE BIOPSY)

P9.S2. Keep in mind that it is still possible to have cancer, even if the biopsy is negative.

P1 (POSSESS $ CANCER)
P2 (MOD POSSESS POSSIBLE)
P3 (MOD POSSIBLE STILL)
P4 (IS BIOPSY NEGATIVE)
P5 (IF P4)
P6 (MOD IF EVEN)
P7 (KEEP-IN-MIND P1 MIND)

P9.S3. This is because, even though multiple samples are taken during a biopsy, it can still miss some cancers.

P1 (TAKE $ SAMPLE)
P2 (AMOUNT-OF SAMPLE MULTIPLE)
P3 (DURING P1 BIOPSY)
P4 (MISS BIOPSY CANCER)
P5 (AMOUNT-OF CANCER SOME)

P6 (THOUGH P1 P4)
P7 (MOD THOUGH EVEN)
P8 (BECAUSE P6)

P10.S1. If the biopsy is taken and prostate cancer is found, the tumor is graded in the medical lab.

P1 (IF P5 P2)
P2 (AND P3 P4)
P3 (TAKE $ BIOPSY)
P4 (FIND $ PROSTATE-CANCER)
P5 (GRADE $ TUMOUR)
P6 (IN GRADE MEDICAL-LAB)

P10.S2. The grade estimates how aggressive a prostate cancer is; that is, how fast it is growing and the likelihood of its spreading.

P1 (ESTIMATE GRADE PROSTATE-CANCER)
P2 (IS PROSTATE-CANCER AGGRESSIVE)
P3 (HOW P1 P2)
P4 (GROW $ PROSTATE-CANCER)
P5 (MOD GROW FAST)
P6 (SPREAD $ PROSTATE-CANCER)
P7 (AND P4 P6)
P8 (HOW P7)

P10.S3. Sometimes you will hear the grade referred to as the Gleason grade.

P1 (REFER $ GRADE)
P2 (AS GRADE GLEASON-GRADE)
P3 (HEAR $ PERSON)
P4 (MOD HEAR SOMETIMES)

P11.S1. Once diagnosis is made, prostate cancer is categorized into stages based on the size and spread of the disease.

P1 (MAKE $ DIAGNOSIS)
P2 (CATEGORIZE $ PROSTATE-CANCER)
P3 (INTO CATEGORIZE STAGE)
P4 (AND P5 P7)
P5 (BASE-ON STAGE SIZE)
P6 (OF SIZE DISEASE)
P7 (BASE-ON STAGE SPREAD)
P8 (OF SPREAD DISEASE)
P9 (ONCE P2 P1)

P11.S2. Learn more about grading and staging of prostate cancer.

P1 (ABOUT LEARN GRADING)
P2 (ABOUT LEARN STAGING)
P3 (AMOUNT-OF LEARN MORE)
P4 (AND P5 P6)
P5 (OF GRADING PROSTATE-CANCER)
P6 (OF STAGING PROSTATE-CANCER)

Appendix E: Samples of Protocol Propositional Analysis

Female; Colorectal cancer, Low readability (Participant 3)

Q. *Okay so let's go to the general article about colon cancer and similar question, if you were going to tell someone about the information in this article, what is this article about?*

If I was going to talk about colon cancer with someone I would again hand it to them and say here read this, it might be helpful to you.

Q. *So what type of information is in here?*

The risk factors and how it can affect your life, the biopsy, it just explained the certain factors, the prognosis, chances of recovery are in here. The risk factors and it explains too about the parts of your bowel and how they all work and then it talks about the big intestine, and the small intestine and the rectum and all of that.

P1	AFFECT	RISK-FACTOR	LIFE
P2	HOW	P1	
P3	EXPLAIN	$	FACTOR
P4	MOD	FACTOR	CERTAIN
P5	EXPLAIN	$	PROGNOSIS
P6	EXPLAIN	$	RECOVERY
P7	OF	CHANCE	RECOVERY
P8	EXPLAIN	$	BOWEL
P9	OF	PART	BOWEL
P10	WORK	P9	
P11	HOW	P10	
P12	EXPLAIN	$	BIG-INTESTINE
P13	EXPLAIN	$	SMALL-INTESTINE
P14	EXPLAIN	$	RECTUM

Q. *What are two or three risk factors for developing colorectal cancer that are listed in the article?*

Family history, age and I believe diet; I'm not sure about the diet, history of polyps which I really didn't understand. I don't know what a polyp is other than it's a non malignant growth but how would you find it if you weren't looking for cancer, like (probably similar signs, or symptoms and then they do a test, that's why screening is so important I guess).

P1	IS	FAMILY-HISTORY	RISK-FACTOR
P2	IS	AGE	RISK-FACTOR
P3	IS	DIET	RISK-FACTOR
P4	MOD	RISK-FACTOR	POSSIBLE
P5	OF	HISTORY	POLYPS
P6	IS	POLYPS	GROWTH
P7	MOD	GROWTH	NON-MALIGNANT

Q. *So what are some possible symptoms of colorectal cancer, you mentioned a few?*

Blood in your stool, can I look (if you can remember off the top of your head) I should have read it but I should have gone over it again. I think you're tired and I think you have a tendency to throw up, I'm not sure (great).

P1	IS	SYMPTOM	BLOOD
P2	IN	BLOOD	STOOL
P3	IS	SYMPTOM	TIREDNESS
P4	IS	SYMPTOM	VOMIT

Female; Colorectal cancer, High readability (Participant 1)

Q. *Okay so we're going to move onto the next article here, "Can Colorectal Cancer be Prevented" and again I know we already touched on this article but pretend that you're telling a friend or family member about the information, what would you describe to them?*

Well it can be prevented if you see to your general health and wellbeing, that's a protection, it can't protect you if it's a genetic problem, that may happen anyway even if you do all the things, but if you keep track of, if you get check-ups, if you know you have a genetic predisposition or a bowel predisposition you should get tested, and I guess get advice from a doctor of things to do, many of the things that ___ and it didn't stress this in the article but fresh, clean water is very important _____ so I would stress water as well as vegetables and, and is this the article where, I don't think I read very much of that one so it must be this one where they said vitamin D (yeah) so you should have vitamins especially folic acid and you should have vitamin D and not too much sun because there's problems with skin cancer if you get too much sun so it's better to get it from a pill, fresh fruits and vegetables are good, whole grain cereals which I have to be careful with because I'm a little bit sensitive to them, bran and wheats I've got to stay away from but and there maybe other people that are sensitive like that too, got to be careful and I think for inflammation and stuff like that, I think maybe that what you're allergic to at some point you pay attention to that and that's not a general ___but I think it said that specifically in here.

P1	IS	GENERAL-HEALTH	GOOD
P2	IS	WELLBEING	GOOD
P3	AND	P1	P2
P4	PREVENT	$	COLORECTAL-CANCER
P5	IF	P4	P3
P6	IS	P3	PROTECTION
P7	IS	COLORECTAL-CANCER	GENETIC-PROBLEM
P8	NEGATE	P4	
P9	IF	P8	P7
P10	GET	$	TEST
P11	POSSESS	$	GENETIC-PREDISPOSITION
P12	POSSESS	$	BOWEL-PREDISPOSITION
P13	OR	P11	P12
P14	IF	P13	P10
P15	POSSESS	$	VITAMIN
P16	EXAMPLE-OF	VITAMIN	FOLIC-ACID
P17	MOD	FOLIC-ACID	ESCECIALLY
P18	EXAMPLE-OF	VITAMIN	VITAMIN-D
P19	IS	FRUIT	GOOD
P20	MOD	FRUIT	FRESH
P21	IS	VEGETABLE	GOOD
P22	IS	CEREAL	GOOD
P23	MOD	CEREAL	WHOLE-GRAIN

I'm reading here calcium I forgot to mention that one, yeah, I met a wonderful lady. We went to a speech that she gave. She was 83 and she had been diagnosed with osteoporosis. She said she drank four glasses of milk every day and did a lot of exercise. I mean hours not just ½ hour and she was in much better shape today than when they diagnosed her and she actually reversed the osteoporosis so that has nothing much to do with colorectal cancer but I think it goes into that point about your general health and wellbeing and activity. If you're feeling well in a healthy way, if your general health is good, you will be more active and that's a good thing to avoid colorectal cancer, if you are active. I don't think you have to do gym stuff necessarily but walking, and dancing. I have some older friends who do dancing and they find it good; I've never done it, so the other two things that I didn't mention were these non steroidal drugs, and the female hormones and the female hormones are up for grabs presently and the non steroidal drugs. I mean it sounds good from the article here, it's just my own particular and I'm thinking that I mean and that sounds ridiculous to everybody I've said it to but that was my experience and so I would stay away from those ___, but I'm willing to get more information as things come forward and they might have something better so I would look into it anyway and see what's developing. But presently I'm not in favour of ___.

P1	IS	GENERAL-HEALTH	GOOD
P2	IS	PERSON	ACTIVE
P3	IF	P1	P2
P4	AVOID	$	COLORECTAL-CANCER
P5	HELP	P2	P4

Q. *So some specific questions, which we've already touched on in your explanation, does the article discuss prevention or treatment?*

Well it says, I guess you could call it that, treatment, once you've, it doesn't say anything about cutting out. It just said that you have to have it cut out if you're 20 and you have that particular genetic predisposition so I guess you could call that a treatment and if they last longer. It's pretty extreme but it might be better than what they would have otherwise so yes there's some treatments there and what was the other question?

P1	CUT-OUT	$	COLON
P2	AT	PERSON	AGE-20
P3	POSSESS	$	GENETIC-PREDISPOSITION
P4	AND	P2	P3
P5	IF	P4	P1
P6	IS	P1	TREATMENT

Q.*Prevention?*

Well prevention yeah, good water, good food and it doesn't say anything about it here but you know just having this, you know just doing things that you enjoy because that gets you focused on making

motions of various kinds which are good for you, (laugh) you know like my friends who are doing dancing, well, they don't do it because it's for their health, they do it because they enjoy doing it so, the kind of activity that you enjoy doing you shouldn't avoid is what I'm saying because anything that makes you active and if you're enjoying it too, it can't but be good for your health.

Q. *What types of food are recommended to reduce people's risk of colorectal cancer according to the article?*

The whole grains and fresh fruits and vegetables and vitamin D and the calcium, I can't think of anything else offhand but, oh wait a minute, there was folic acid.

P1	REDUCE	WHOLE-GRAIN	RISK
P2	REDUCE	FRUIT	RISK
P3	MOD	FRUIT	FRESH
P4	REDUCE	VEGETABLE	RISK
P5	REDUCE	VITAMIN-D	RISK
P6	REDUCE	CALCIUM	RISK
P7	REDUCE	FOLIC-ACID	RISK

Q. *What is the recommended amount of weekly physical activity?*

I think they say 30 minutes a day for four days, something like that. I haven't even done that yet but I'm going to. I'm already improving my water drinking so activity is the next thing on my list.

P1	AMOUNT-OF	EXERCISE	30MIN-PER-DAY
P2	FOR	P1	FOUR-DAY

Male; Prostate cancer, Low Readability (Participant 15)

Q. *So now to the article on diet (okay, thought I might need it, I got it) this has some different information from the other two (right) in general if you were going to tell someone about it what would you say that this article is about?*

The role of diet in assessing you know whether, it talks about diet, but it also talks about, it focuses on you know the headings are geography and diets, and you know fat in your diet, and then it also talks about antioxidants and licopene and Soya, in other words what foods and what type of diet that we have in North America and in other parts of the world that seem to be somewhat related.

P1	AND	GEOGRAPHY	DIET
P2	IN	FAT	DIET
P3	EXAMPLE-0F	DIET	ANTIOXIDANT
P4	EXAMPLE-0F	DIET	LICOPENE
P5	EXAMPLE-0F	DIET	SOYA
P6	IN	FOOD	NORTH-AMERICA
P7	IN	DIET	NORTH-AMERICA
P8	OF	TYPE	DIET
P9	OF	PART	WORLD
P10	MOD	PART	OTHER
P11	IN	FOOD	P9
P12	IN	DIET	P9

There's not a one to one correspondence but there seems to be a relationship between what we eat and the incidences of the cancer in North America and the world.

P1	EAT	$	PEOPLE
P2	OF	INCIDENCE	CANCER
P3	IN	CANCER	NORTH-AMERICA
P4	IN	CANCER	WORLD
P5	AND	P1	P2
P6	BETWEEN	RELATIONSHIP	P5

Q. *Okay so some specific questions now, what is the benefit of a balanced diet?*

I assume that you're not setting yourself up for things like, if you have a preponderance of fat in your diet then you're putting yourself at greater risk, you know, I guess you're in-taking these antioxidants, and you know just basically I would think you would be healthier, you know you're reducing your risk, you're probably going to have more energy, feel better, you know blah, blah, blah (okay).

P1	OF	PREPONDERANCE	FAT
P2	IN	FAT	DIET
P3	AT	PERSON	RISK
P4	AMOUNT-OF	RISK	GREATER
P5	IF	P3	P1

P6	POSSESS	$	ANTIOXIDANT
P7	IS	PERSON	HEALTHIER
P8	CAUSE	P6	(P7,P9,P10,12)
P9	REDUCE	$	RISK
P10	POSSESS	$	ENERGY
P11	MOD	ENERGY	MORE
P12	FEEL	PERSON	BETTER

Q. *Why do people in eastern countries less likely than westerners to develop cancer according to the article?*

Well it seems to imply that there is some sort of genetic issue and then it goes on to say that when they move these people from the area in which they lived in and they moved them to North America and subjected them to or they subjected themselves to the North American diet then the diet seemed to have a negative influence on the incidents of cancer, so it touches on the issues of genetics and diet, like it shows that they thought that the genetics was an issue and it still may be but there is also some pretty strong evidence to show that diet is really more of a factor (great).

P1	EXIST	$	GENETIC-ISSUE
P2	MOVE-TO	EASTERNER	NORTH-AMERICA
P3	EAT	EASTERNER	DIET
P4	MOD	DIET	NORTH-AMERICAN
P5	INFLUENCE	DIET	CANCER
P6	MOD	INFLUENCE	NEGATIVE
P7	OF	INCIDENT	CANCER
P8	IS	GENETIC	ISSUE
P9	IS	DIET	FACTOR
P10	MOD	FACTOR	MORE
P11	EXIST	$	EVIDENCE
P12	MOD	EVIDENCE	STRONG
P13	SHOW	EVIDENCE	P9
P14	BUT	P8	P13

Q. *And what diseases are linked with dietary fat intake?*

Prostate cancer obviously, I don't if it says about it a great deal, I don't see much about heart disease but I think there's probably something in there (yeah) about it. I think, is heart disease #1 (yeah) and prostate is #2, so those two in particular I guess, heart and prostate.

P1	IS	HEART-DISEASE	#1
P2	IS	PROSTATE-CANCER	#2
P3	AND	P1	P2
P4	LINK-WITH	P3	DIETARY-FAT-INTAKE

Q. *Okay, what are antioxidants and can you list a couple of foods that contain them?*

Tomatoes and fruit, and vegetables.

P1	CONTAIN	TOMATO	ANTIOXIDANT
P2	CONTAIN	FRUIT	ANTIOXIDANT
P3	CONTAIN	VEGETABLE	ANTIOXIDANT

Antioxidant is something that in the body that fights free radicals.

P1	IN	ANTIOXIDANT	BODY
P2	FIGHT	ANTIOXIDANT	FREE-RADICAL

Free radicals is a, I don't know the chemistry of it, but I know it's bad for you. If you have a lot of free radicals then the risk is enhanced for cancer.

P1	IS	FREE-RADICAL	BAD
P2	FOR	BAD	PERSON
P3	POSSESS	$	FREE-RADICAL
P4	AMOUNT-OF	FREE-RADICAL	A-LOT
P5	ENHANCE	P4	RISK
P6	FOR	RISK	CANCER
P7	IF	P5	P3

The antioxidants neutralize these free radicals which in turn lessen your risk of getting cancer (okay).

P1	NEUTRALIZE	ANTIOXIDANT	FREE-RADICAL
P2	REDUCE	P1	RISK
P3	OF	RISK	CANCER
P4	IMPLY	P1	P2

Male; Prostate cancer, High Readability (Participant 13)

Q. *So let's move onto this article here, this is one page, written in bigger writing. Okay so if you were going to tell a friend or family member about the information that you read what would you tell them that this article is about?*

If you have any symptoms you have to go to a doctor to get the first exam, the digital rectal exam and if the doctor recommends the next step, you have to go to the next step, to have a specific antigen test and from that point there the exams they show up where you are, you have to go further, if it shows everything clear you don't have to bother, that's it, but you have to do both of those, the first tests, the digital and the antigen tests, those two and if you go the second one there, and they find something at the laboratory, finds something that's there, which stage, the first stage, how far progressive every year, you know you have to go once a year and make a check and so on to see you progress.

P1	POSSESS	$	SYMPTOM
P2	VISIT	$	DOCTOR
P3	IF	P2	P1
P4	GET	$	DIGITAL-RECTAL-EXAM
P5	FROM	P4	DOCTOR
P6	REF	DIGITAL-RECTAL-EXAM	FIRST-EXAM
P7	RECOMMEND	DOCTOR	STEP
P8	MOD	STEP	NEXT
P9	GET	$	SPECIFIC-ANTIGEN-TEST
P10	IF	P9	P7
P11	REF	STEP	SPECIFIC-ANTIGEN-TEST
P12	IS	SPECIFIC-ANTIGEN-TEST	CLEAR
P13	GET	$	TEST
P14	MOD	TEST	FURTHER
P15	NEGATE	P13	
P16	IF	P15	P12
P17	DETECT	SPECIFIC-ANTIGEN-TEST	SOMETHING
P18	OF	STAGE	SOMETHING
P19	OF	PROGRESS	SOMETHING
P20	VISIT	$	DOCTOR
P21	AMOUNT-OF	VISIT	ONCE-YEAR
P22	IF	P20	P17

Female; Breast Cancer, Low Readability (Participant 6)

Q. *Let's go to the purple, this one here on Detection, and again if you were going to tell someone what this article was about what would you tell them?*

Well basically it just outlines approximately ten different problems or symptoms that you might observe that would warrant going to your doctor for a further check and you want me to list them? (You can go ahead). I'll try and do what I can (sure).

P1	OBSERVE	$	PROBLEM
P2	OBSERVE	$	SYMPTOM
P3	OR	P1	P2
P4	MOD	PROBLEM	DIFFERENT
P5	NUMBER-OF	PROBLEM	TEN
P6	MOD	OBSERVE	POSSIBLE
P7	GO-TO	$	DOCTOR
P8	FOR	DOCTOR	CHECK
P9	MOD	CHECK	FURTHER

You know it's (whatever you can) text anxiety, well things like lumps or swelling or discharge from the nipple area, just general fatigue, you go and get it checked out for sure (okay, great).

P1	IN	LUMP	NIPPLE-AREA
P2	IN	SWELLING	NIPPLE-AREA
P3	FROM	DISCHARGE	NIPPLE-AREA
P4	OR	(P10,P11,P12)	
P5	EXAMPLE-OF	GENERAL-FATIGUE	SYMPTOM
P6	EXAMPLE-OF	P11	SYMPTOM
P7	EXAMPLE-OF	P12	SYMPTOM
P8	EXAMPLE-OF	P13	SYMPTOM

Q. *The article talks about some screening methods, can you name a couple of them?*

Yeah well there's the monthly self breast examination, there's regular clinical examinations by your physician or a practicing nurse or whatever, and your mammograms and possibly ultrasound if it's warranted.

P1	EXAMPLE-OF	BREAST-SELF-EXAMINATION	SCREENING-METHOD
P2	MOD	BREAST-SELF-EXAMINATION	MONTHLY
P3	EXAMPLE-OF	CLINICAL-EXAMINATION	SCREENING-METHOD
P4	MOD	CLINICAL-EXAMINATION	REGULAR
P5	BY	CLINICAL-EXAMINATION	PHYSICIAN
P6	BY	CLINICAL-EXAMINATION	NURSE
P7	OR	P5	P6

P8	EXAMPLE-OF	MAMMOGRAM	SCREENING-METHOD
P9	EXAMPLE-OF	ULTRASOUND	SCREENING-METHOD
P10	MOD	ULTRASOUND	POSSIBLE
P11	WARRANT	$	ULTRASOUND
P12	IF	P9	P11

Q. *And at what age does it suggest that women should begin to have mammograms?*

Probably around, at 40 I believe (okay), every two years and possibly annually or semi-annually if there's something that they're watching.

P1	BEGIN	$	MAMMOGRAM
P2	AT	MAMMOGRAM	40
P3	AMOUNT-OF	MAMMOGRAM	EVERY-TWO-YEAR
P4	MOD	BEGIN	POSSIBLE
P5	WATCH	$	SOMETHING
P6	AMOUNT-OF	MAMMOGRAM	ANNUAL
P7	AMOUNT-OF	MAMMOGRAM	SEMI-ANNUAL
P8	OR	P6	P7
P9	IF	P8	P5

Female; Breast Cancer, High Readability (Participant 8)

Q. *Okay so let's go to the second breast cancer article and again if you were going to tell a friend or a family member about the information what would you tell them about this article, just in general?*

Well it talked about women being exposed to two different types of a gene and it went into detail about what each gene was and if you are predisposed to that type of gene, the type of cancer that could possibly follow, and then they followed it up with the MRI treatment to pick it up early so that you could have it treated, lots of heavy reading basically.

P1	EXPOSE	$	WOMAN
P2	TO	EXPOSE	GENE
P3	MOD	GENE	DIFFERENT
P4	NUMBER-OF	GENE	TWO
P5	PREDISPOSE	$	WOMAN
P6	TO	PREDISPOSE	GENE
P7	FOLLOW	$	CANCER
P8	OF	TYPE	CANCER
P9	IF	P7	P6
P10	DETECT	$	MRI
P11	MOD	DETECT	EARLY
P12	TREAT	$	CANCER
P13	THEREFORE	P10	P12

Q. *Okay so does the article discuss more prevention or treatment for breast cancer?*

I would say it talks more about treatment from the MRI standpoint to zero right in and get rid of it before it spreads.

P1	OF	TREATMENT	MRI
P2	RID	$	CANCER
P3	SPREAD	$	CANCER
P4	BEFORE	P2	P3

References

Armbruster, B. B. (1986). Schema theory and the design of content-area textbooks. *Educational Psychologist, 21*(4), 253-267.

Arocha, J. F. (2005). *Paradigms for research*. Unpublished manuscript.

Arocha, J. F., & Patel, V. L. (1995). Novice diagnostic reasoning in medicine: Accounting for clinical evidence. *Journal of the Learning Sciences, 4*, 355-384.

Arocha, J. F., Wang, D., & Patel, V. L. (2005). Identifying reasoning strategies in medical decision making: A methodological review. *Journal of Biomedical Informatics, 38*, 154-171.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*, 471-481.

Bauman, A. (1997). The comprehensibility of asthma education materials. *Patient Education and Counseling, 32*, S51-S59.

Bernhardt, J. M., & Cameron, K. A. (2003). Accessing, understanding, and applying health communication messages: The challenge of health literacy. In T. L. Thomson (Ed.), *Handbook of Health Communication* (pp. 583-605). Mahwah, NJ: Lawrence Erlbaum.

Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), *Understanding Expository Text: A Theoretical and Practical Handbook for Analyzing Explanatory Text*. Hillsdale, NJ: Lawrence Erlbaum.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Education Psychology, 83*, 329-345.

Bull, F. C., Kreuter, M. W., & Shariff, D. P. (1999). Effects of tailored, personalized, and general materials on physical activity. *Patient Education and Counseling, 36*, 181-192.

Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior, 19,* 135-145.

Canadian Cancer Society. (2005). *Letrozole after tamoxifen reduces recurrence of breast cancer*. Retrieved December 3, 2005 from http://www.cancer.ca/ccs/internet/mediaresleaselist/0,3208,3172_343093094_507768238_langId-en,00.html.

Cicourel, A. V., & Katz, A. N. (1996). Ecological validity in pragmatic research: Pragmatics & Cognition. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Cooley, M. E., Moriarty, H., Berger, M. S., Selm-Orr, D., Coyle, B., & Short, T. (1995). Patient literacy and the readability of written cancer education materials. *Oncology Nursing Forum, 22*, 1345-1351.

Davis, T. C., Crouch, M. A., Wills, G., & Miller, S. (1990). The gap between patient reading comprehension and the readability of patient education materials. *Journal of Family Practice, 31*, 533-538.

Doak, C. C., Doak, L. G., Friedell, G. H., & Meade, C. D. (1998). Improving comprehension for cancer patients with low literacy skills: strategies for clinicians. *Cancer Journal for Clinicians, 48*, 151-162.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension. *Applied Psychological Measurement, 11*(2), 175-193.

Farand, L., Lafrance, J.-P., & Arocha, J. F. (1988). Collaborative problem-solving in telemedicine and evidence interpretation in a complex clinical case. *International Journal of Medical Informatics, 51*, 153-167.

Fielding, L., & Pearson, P. D. (1994). Reading comprehension: what works. *Educational Leadership, 51*, 62-68.

Fincher-Kiefer, R., Post, T. A., Greene, T. R., & Voss, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory and Language, 27*(416-428).

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes, 25*, 285-307.

Frederiksen, C. H. (1975). Logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458.

French, K. S., & Larrabee, J. H. (1999). Relationships among educational material readability, client literacy, perceived beneficence, and perceived quality. *Journal of Nursing Care Quality, 13*, 68-82.

Friedman, D. B., Hoffman-Goetz, L., & Arocha, J. F. (2004). Readability of cancer information on the Internet. *Journal of Cancer Education, 19*, 117-122.

Friedman, D. B., Hoffman-Goetz, L., & Arocha, J. F. (in press). Health literacy and the World Wide Web: Comparing the readability of leading incident cancers on the Internet. *Medical Informatics and the Internet in Medicine.*

Friedman, DB., Hoffman-Goetz, L., & Arocha, J.F. (In press). Health literacy and the World Wide Web: Comparing the readability of leading incident cancers on the Internet. *Medical Informatics and the Internet in Medicine.*

Ghiglione, R., Landré, A., & Bromberg, M. D., A. (1990). "Hear say" or mediated political discourse/"Écoute voir" ou le discours politique médiatisé. *Psychologie Française, 35*, 123-142.

Glazer, H. R., Kirk, L. M., & Bosler, F. E. (1996). Patient education pamphlets about prevention, detection, and treatment of breast cancer for low literacy women. *Patient Education and Counseling, 27*, 185-189.

Glazer-Waldman, H., Hall, K., & Weiner, M. F. (1985). Patient education in a public hospital. *Nursing Research, 34*, 184-185.

Glenberg, A. M., & Mathew, S. (1992). When minimalism is not enough: Mental models in reading comprehension. *Psycoloquy, 3*(64), Reading Inference (1).

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69-83.

GraphViz [Open source computer software]. (2005). http://www.research.att.com/sw/tools/graphviz/

Griffiths, M., & Leek, C. (1995). Patient education needs: opinions of oncology nurses and their patients. *Oncology Nursing Forum, 22,* 139-144.

Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Hoskyn, M., & Swanson, H. L. (2003). The relationship between working memory and writing in younger and older adults. *Reading & Writing, 16*(8), 759-784.

Jimison, H. B., & Sher, P. P. (2000). Advances in presenting health information to patients. In G. B. Chapman & F. A. Sonnenberg (Eds.), *Decision Making in Health Care: Theory, Psychology, and Applications*. Cambridge: U. K.: Cambridge University Press.

Kardash, C. M., & Noel, L. K. (2000). How organizational signals, need for cognition, and verbal ability affect text recall and recognition. *Contemporary Educational Psychology, 25,* 317-331.

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in younger and older adults. *Psychology and Aging, 16,* 312-322.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294-303.

Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., & Keenan, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5*, 257-274.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14*, 196-214.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kintsch, W., & Vipond, D. (1977). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-366). Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives on Memory Research* (pp. 329-366). Hillsdale, NJ: Erlbaum.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*(2), 133-159.

Klare, G. R. (1974-5). Assessing readability. *Reading Research Quarterly, 10*, 62-102.

Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Reading Behavior, 8*, 129-152.

Kools, M., Ruiter, R. A. C., van de Wiel, M. W. J., & Kok, G. (2004). A qualitative study into how professional writers make texts coherent. *Health Education & Behavior, 31*, 720-740.

Kozminsky, E. (1977). Altering comprehension: The effect of biasing titles on text comprehension. *Memory & Cognition, 5*, 482-490.

Landauer, T. K., Foltz, P., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Le Bas, J. (1989). Comprehensibility of patient education literature. *Australian & New Zealand Journal of Psychiatry, 23*, 542-546.

Mannes, S., & St George, M. (1996). Effects of prior knowledge on text comprehension: A simple modeling approach. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 115-139). Hillsdale, NJ: Lawrence Erlbaum.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction, 14*, 1-43.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

McNamara, D. S., & O'Reilly, T. (2002). Learning: Knowledge acquisition, representation, and organization. In J. W. Guthrie (Ed.), *The Encyclopedia of Education*. New York, NY: Macmillan Reference.

McNeal, B., Salisbury, Z., Baumgardner, P., & Wheeler, F. (1984). Comprehension assessment of diabetes education program participants. *Diabetes Care, 7*, 232-235.

Meade, C., & Byrd, J. C. (1989). Patient literacy and the readability of smoking education literature. *American Journal of Public Health, 79*, 204-206.

Meade, C., Diekman, J., & Thornhill, D. (1992). Readability of American Cancer Society patient education literature. *Oncology Nursing Forum, 19*, 51-55.

Meade, C., & Smith, C. F. (1991). Readability formulas: cautions and criteria. Patient Education & Counseling, 17, 153-158.

Means, M., & Voss, J. (1985). Star Wars: A developmental study of expert and novice knowledge structures. *Memory and Language, 24*, 746-757.

Michielutte, R., Bahnson, J., Dignan, M. B., & Schroeder, E. M. (1992). The use of illustrations and narrative text style to improve readability of a health education brochure. *Journal of Cancer Education, 7*, 251-260.

Miller, J. R., & Kintsch, E. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(4), 335-354.

Mitchell, J. V. (Ed.). (1985). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

National Cancer Institute of Canada. (2003). *Canadian Cancer Statistics 2003*. Retrieved December 2, 2005 from http://www40.statcan.ca/l01/cst01/health25a.htm.

Park, H.-S. (2000). The effect of item text characteristics on children's growth in reading. *Dissertation Abstracts International, 61*(2-A), 548.

Pask, G. (1988). Learning strategies, teaching strategies, and conceptual or learning style. In R.R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 83–99). New York: Plenum Press.

Patel, V. L., & Arocha, J. F. (1995). Cognitive models of clinical reasoning and conceptual representation. *Methods of Information in Medicine, 34*, 47-56.

Patel, V. L., Arocha, J. F., Diermeier, M., Greenes, R. A., & Shortliffe, E. H. (2001). Methods of cognitive analysis to support the design and evaluation of biomedical systems: the case of clinical practice guidelines. *Journal of Biomedical Informatics, 34*, 52-66.

Patel, V. L., Groen, G. J., & Frederiksen, C. H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education, 20*, 3-9.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2000). Conceptual change in the biomedical and health sciences domain. In R. Glaser (Ed.), *Advances in Instructional Psychology: Educational design and cognitive science* (pp. 329-392). Mahwah, NJ: Lawrence Erlbaum.

Payne, S., Large, S., Jarrett, N., & Turner, P. (2000). Written information given to patients and families by palliative care units: a national survey. *Lancet, 355*, 1792.

Potelle, H., & Rouet, J.-F. (2003). Effects of content representation and readers' prior knowledge on the comprehension of hypertext. *International Journal of Human-Computer Studies, 58*, 327-345.

Protégé [Computer software]. (2000). Stanford, CA: Stanford Medical Informatics.

Prostate Cancer Research Foundation of Canada. (2006). Retrieved December 3, 2005 from http://www.prostatecancer.ca/english/home/.

Ratzan, S. C. (2001). Health literacy: communication for the public good. *Health Promotion International, 16*, 207-214.

Rice, G. E., & Meyer, B. J. (1986). Prose recall: effects of aging, verbal ability, and reading behavior. *Journal of Gerontology, 41*, 469-480.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1999). Self-reference and the encoding of personal information. In R. F. Baumeister (Ed.), *The self in social psychology. Key readings in social psychology* (pp. 139-149). New York, NY: Psychology Press.

Romano R. (1979). *Readability in Cancer Communications: Methods, Examples and Resources for Improving the Readability of Cancer Messages and Materials*. Bethesda, MD: U.S. Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, National Cancer Institute.

Rudd, R. E., Moeykens, B. A., & Colton, T. C. (1999). Health and literacy: a review of medical and public health literature. In J. Comings, B. Garners & C. Smith (Eds.), *Annual Review of Adult Learning and Literacy*. New York, NY: Jossey-Bass.

Runkel, P. J. (1990). *Casting nets and testing specimens : two grand methods of psychology*. New York: NY: Praeger.

School Renaissance Institute. (2000). *The ATOS readability formula for books and how it compares to other formulas*. Retrieved Jan 17, 2005 from http://research.renlearn.com/research/62.asp.

Skinner, C. S., Strecher, V. J., & Hospers, H. (1994). Physician recommendations for mammography: Do tailored messages make a difference? *American Journal of Public Health, 84*, 43-49.

Smith, E. E., & Swinney, D. A. (1992). The Role of Schemas in Reading Text: A Real-Time Examination. *Discourse Processes, 15*, 303-316.

Smith, H., Gooding, S., Brown, R., & Frew, A. (1998). Evaluation of readability and accuracy of information leaflets in general practice for patients with asthma. *British Medical Journal, 317*, 263-264.

Statistics Canada. (2004, July 8). *Household Internet Survey*. Retrieved November 3, 2004 from http://www.statcan.ca/Daily/English/040708/d040708a.htm

Stine, E. A. L., & Wingfield, A. (1990). How much do working memory deficits contribute to age differences in discourse memory? *European Journal of Cognitive Psychology, 2*, 289-304.

Strecher, V. J., Kreuter, M. W., Den Boer, D. J., Kobrin, S., Hospers, H. J., & Skinner, C. S. (1994). The effects of computer-tailored smoking cessation messages in family practice settings. *Journal of Family Practice, 39,* 262-270.

Thagard, P., & Verbeurgt, K. (1997). Coherence as constraint satisfaction. *Cognitive Science, 22,* 1-24.

United States Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2000). *Healthy People 2010*. Retrieved Nov 8, 2004 from the World Wide Web at http://www.health.gov/healthypeople.

van den Broek, P., Virtue, S., Graddy, M., Tzeng, Y., & Sung, Y.-C. (2002). Comprehension of memory of science texts: inferential processes and the construction of a mental representation. In J. Otero, J. Leon & A. C. Graesser (Eds.), *The Psychology of Science Text Comprehension* (pp. 131-154). Mahwah, NJ: Lawrence Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press Inc.

Zwaan, R. A., & Graesser, A. C. (1993). Reading goals and situation models: A commentary on Glenberg & Mathew. *Psycoloquy, 4*(5), Reading Inference (4).