

Eliciting Test-selection Strategies for a Decision-Support System in Oncology

Danielle Sent¹, Linda C. van der Gaag¹, Cilia L.M. Witteman*¹,
Berthe M.P. Aleman², Babs G. Taal²

¹Institute of Information and Computing Sciences, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
{danielle,linda,cilia}@cs.uu.nl

²The Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis,
Department of Radiation Oncology and Gastroenterology,
Plesmanlaan 121, 1066 CZ Amsterdam, The Netherlands
{baleman,bgtaal}@nki.nl

Abstract

Decision-support systems often include a strategy for selecting tests in their domain of application. Such a strategy serves to provide support for the reasoning processes in the domain. Generally a test-selection strategy is offered in which tests are selected sequentially. Upon building a system for the domain of oesophageal cancer, however, we felt that a sequential strategy would be an oversimplification of daily practice. To design a test-selection strategy for our system, we decided therefore to acquire knowledge about the actual strategy used by the experts in the domain and, more specifically, about the arguments underlying their strategy. For this purpose, we used an elicitation method that was composed of an unstructured interview to gain general insight in the test-selection strategy used, and a subsequent structured interview, simulating daily practice, in which full details were acquired. We used the method with two experts in our application domain and found that the method closely fitted in with their daily practice and resulted in a large amount of detailed knowledge.

keywords: (medical) decision-support systems, knowledge elicitation, selection of diagnostic tests

1 Introduction

Decision-support systems are being developed for a wide range of domains. To support the reasoning processes in its domain of application, such a system

*currently at Faculty of Social Sciences, Section Diagnostic Decision Making, University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands, c.witteman@ped.kun.nl

often includes a strategy for selecting tests. In the medical domain, a decision-support system may, for example, suggest a sequence of diagnostic tests to be performed in order to reduce the uncertainty about a patient's true condition. The test-selection strategy thereby provides support for the task of diagnostic reasoning. In most decision-support systems, a strategy is offered in which tests are suggested sequentially, that is, on a one-by-one basis. The system then suggests a single test to be performed and awaits the user's input; after taking the test's result into account, the system suggests a subsequent test, and so on.

With the help of two experts in gastrointestinal oncology, we have developed, over a period of more than five years, a decision-support system for the domain of oesophageal cancer [1]. Our system is based on normative principles and thus has a mathematical foundation in probability and utility theory [2, 3]. The system includes detailed knowledge about tumour growth and about the processes of invasion and metastasis. It further captures knowledge about the various different diagnostic tests that can be performed to gain insight in the often hidden condition of a patient. The system moreover contains knowledge about the beneficial effects and complications to be expected from the different treatment alternatives. Building upon this knowledge, the system provides for establishing the stage of a patient's cancer and for prognosticating the likely effects of the various therapies, based upon the patient's symptoms, signs, and test results.

Our decision-support system for oesophageal cancer at present does not support the selection of diagnostic tests. For a specific patient, the attending physician orders a number of tests, based upon his or her own judgement, and simply enters the results into the system; the system does not provide the physician with information about which tests would be relevant and should be considered next for the patient under consideration. Building upon the system's mathematical foundation, however, a sequential test-selection strategy could easily be designed. Such a strategy would select, on a one-by-one basis, the test that is the most informative, for example in terms of entropy reduction, given the already available patient specifics [4, 5]. Upon working with our system, however, we noticed that our experts do not select tests one after the other, but in packages instead. We felt that a sequential test-selection strategy would be an oversimplification of our experts' problem-solving practice and we decided to design a test-selection strategy for our system that would build upon the arguments used by the experts for deciding whether or not to order specific tests. The resulting strategy would thus more closely fit in with the strategies for test selection currently used in the domain than a standard sequential test-selection strategy.

To acquire knowledge about the actual test-selection strategy employed by our experts and about the arguments underlying their strategy more specifically, we used an elicitation method that combined several different techniques for knowledge elicitation [6]. The method consisted of two main interviews. The first of these was an unstructured interview that was aimed at providing insight in the overall strategy used by the experts. The second interview was a structured interview in which further details were acquired. In this latter inter-

view, the experts' problem-solving practice was carefully simulated by means of cards, or vignettes, describing realistic patient cases. By simulating daily routine, we aimed to exclude, as much as possible, the various different biases that could possibly originate from the elicitation method used. We note that the idea of following up an unstructured interview by a structured one has been proposed before, for example in Cognitive Task Analysis [7]. We used the elicitation method with the two experts in our domain of application. We found that the method, and the use of carefully designed patient cases more specifically, closely fitted in with the experts' daily practice. The method resulted in a large amount of detailed knowledge, not just about the actual order in which tests are selected but also about the experts' reasons for ordering certain tests and for deciding not to order other ones.

Since a test-selection strategy offered by a decision-support system should support physicians in their daily problem-solving practice, we feel that for the design of such a strategy, knowledge about the actual strategies employed in the domain of application should be elicited from experts; a standard, sequential strategy may then turn out to deviate too much from daily routines to be acceptable. Our experiences in the domain of oesophageal cancer have demonstrated that the knowledge required for the design of a tailored test-selection strategy can be feasibly acquired: with our elicitation method, we were able to elicit the arguments underlying our experts' strategy in little time.

The paper is organised as follows. In Section 2, we provide some preliminaries on oesophageal cancer and its therapies. In Section 3, we give an overview of the method that we used for eliciting our experts' test-selection strategy. In Section 4, we describe the results that we obtained from the first, unstructured interview. Section 5 reports on the second, structured interview. The paper ends with our concluding observations in Section 6.

2 Preliminaries

Cancer of the oesophagus may develop as a consequence of a lesion of the oesophageal wall, for example associated with smoking habits and alcohol consumption. The primary tumour typically invades the oesophageal wall and may in time invade neighbouring organs beyond the oesophagus. When the tumour has invaded lymphatic vessels and blood vessels, it may give rise to secondary tumours, or metastases, in lymph nodes and in such organs as the liver and lungs. The latter are called haematogenous metastases, while the former are referred to as lymphatic metastases. The depth of invasion of the oesophageal tumour and the extent of its metastases are indicative of the severity of the disease, which is summarised in the cancer's stage.

In order to establish the stage of a patient's oesophageal cancer, generally a number of diagnostic tests are performed. Various different tests are available, giving insight in different aspects of the cancer. A gastroscopic examination, for example, provides information about the presentation characteristics of the primary tumour, which include its length and its location in the oesophagus.

Table 1: An overview of the diagnostic tests that give insight in the stage of an oesophageal cancer

<i>Test</i>	<i>Sensitivity</i>	<i>Specificity</i>
Biopsy	1.00	1.00
Bronchoscopy	0.92	0.96
CT: liver, loco-region, lungs, organs, truncus coeliacus	0.48 – 0.90	0.88 – 0.98
Sonography: neck	0.90	0.95
Endosonography: loco-region, mediastinum, wall, truncus coeliacus	0.51 – 0.78	0.77 – 0.86
Gastrosocopy: circumference, length, location, shape, necrosis	0.87 – 0.99	0.89 – 0.99
Laparoscopy: liver, diaphragm, truncus coeliacus	0.25 – 0.85	0.95 – 0.98
Barium swallow	0.87	0.99
X-ray	0.85	0.98
Physical examination	0.75	0.97
Interview: passage, age, weight loss	-	-

however. The main goal of a surgical procedure is to attain a better life expectancy for a patient, that is, the procedure is curative in nature. Positioning a prosthesis in the oesophagus, on the other hand, is a palliative procedure that cannot improve life expectancy: it is performed merely to relieve the patient’s swallowing problems. Radiotherapy can be administered in a curative regime, aimed at prolonging a patient’s life, as well as in a palliative regime, aimed just at improving the patient’s quality of life. The most preferred treatment in essence is to provide a curative therapy; of these curative therapies, a surgical removal of the primary tumour is preferred to a curative regime of radiotherapy. Providing a therapy, however, is often accompanied not just by beneficial effects but also by complications. These complications can be quite serious and may even prove to be fatal. The beneficial effects and complications to be expected from the different therapies for a specific patient depend on the general condition of the patient, on the characteristics of his or her primary tumour, on the depth of invasion of the tumour into the oesophageal wall and neighbouring organs, and on the extent of metastasis of the cancer. If serious complications are expected for the patient, the attending physician may decide to abstain from providing a curative therapy and to administer one of the palliative treatment alternatives.

With the help of two experts in gastrointestinal oncology, we have developed, over a period of more than five years, a decision-support system that provides for assessing the stage of a patient’s oesophageal cancer and for prognosticating the likely effects of the different treatment alternatives [1]. The kernel of our system is a probabilistic network that captures the state-of-the-art knowledge about oesophageal cancer and its treatment. The diagnostic part of the network is reproduced in Figure 1. This part captures the knowledge about the various different diagnostic tests available and is of interest to the present paper. We

would like to note that the results of a single test are often represented by a number of statistical variables in the network. For example, while a gastroscopic examination of the oesophagus is a single diagnostic test, its results are modelled by the five variables *Gastro-circumf*, *Gastro-length*, *Gastro-location*, *Gastro-shape* and *Gastro-necrosis*.

3 A Method for Eliciting Test-selection Strategies

Before describing the method that we used for eliciting test-selection strategies and before introducing the setting in which we used it, we briefly review some well-known elicitation methods.

The background

For knowledge acquisition, generally a distinction is made between methods that are aimed at eliciting object knowledge (knowing that) and methods with which to elicit problem-solving knowledge (knowing how). For finding out how our domain experts select diagnostic tests, we needed a method that focused on the latter type of knowledge. Several such methods are available, among which observation methods, methods for eliciting verbal reports, and think-aloud methods are the most well known [9, 10].

Observation amounts to recording the behaviour that is exhibited by experts while they are solving problem situations in their domain. Studying the observed behaviour results in a so-called protocol of actions. This protocol can be analysed to infer the actual problem-solving strategies used by the experts. Observation methods are most suitable for domains in which problem solving requires objects to be handled or overt actions to be performed.

The different methods in use for eliciting verbal reports are all variants of the interview. An interview may be focused but otherwise unstructured, with general questions; the topics to be addressed during the interview, but not the precise questions are prepared in advance. The interview may also be structured, containing mostly closed questions. An interview may be held orally or presented on paper, in the form of a questionnaire. Interviews are especially appropriate for domains in which it is relatively easy for experts to verbalise their knowledge, for example because their daily routines involve verbalisations of problem-solving behaviour. Unstructured, oral interviews closely resemble normal conversation. The experts are for example asked what they commonly do when they are confronted with a problem situation in their domain of expertise. This type of interview provides the interviewer with a global understanding of the structure of the knowledge domain, and of the type of strategy used for problem solving; more specifically, the interview results in an overview of the order in which the various reasoning steps generally are performed. Subsequent structured interviews are suitable to deepen the understanding, to further clarify the structure of the knowledge domain, and to zoom in on the details of the problem-solving strategies used by the experts.

Thinking aloud is the only method available for eliciting mental processes directly. The experts are presented with a prototypical problem situation and are asked to verbalise their reasoning processes while solving the problem without interruption. Talking out loud concurrently with solving a problem situation leaves the experts no room or time for interpretation. They therefore directly reveal the strategies they use for solving the problem. Thinking aloud appears to be quite easy for most people and has been found not to interfere with their performance [11]. Think-aloud sessions are generally tape-recorded and transcribed for further analysis.

To conclude we would like to observe that all knowledge-acquisition methods require careful preparation by the knowledge engineer. While some prior knowledge of the domain is advisable on the one hand, to understand the experts' answers and to put the right questions, too much knowledge on the other hand may cause the engineer not to listen carefully to the expert and to interpret answers in the light of his or her own views. Also, the interview questions and problem situations have to be prepared with care to guarantee that the knowledge aimed at is indeed elicited. Moreover, the setting in which the various sessions are to be held should be carefully selected. It should further be noted that, to the experts, some information or reasoning steps may be so self-evident that they are not verbalised, and consequently not acquired with any of the available methods. When analysing the acquired information, therefore, the knowledge engineer should be attentive to unjustified reasoning steps and prompt the experts for further elaboration.

The method

To acquire knowledge about the test-selection strategy used by our experts in the domain of oesophageal cancer, we employed an elicitation method that combined several of the methods reviewed above. Because the process of selecting diagnostic tests in essence is a type of problem solving in which the experts' behaviour is predominantly mental, the method of observation did not suit our purpose; with observation methods, we would not be able for example to gain insight in the experts' reasons for ordering specific tests. Also, thinking aloud with real, concurrent patients was not feasible, for evident reasons. We thus focused on interviews for eliciting verbal reports of problem-solving behaviour. We decided to conduct two consecutive interviews. The aim of the first, unstructured interview, was to obtain insight into the overall test-selection strategy employed and into the general arguments used by the experts. Since such an unstructured interview would be focused on the strategy and not on real patients, we were aware that we risked acquiring a general, text-book procedure rather than the experts' daily problem-solving routines. We decided therefore to follow up the first interview by a structured interview in which the experts were asked to think aloud while deciding, for a number of patients, which diagnostic tests to order. We felt that working with patient cases in a carefully conducted manner would closely fit in with the experts' problem-solving practice and would thus reduce possible biases from the elicitation method used. The aim of this second interview was to fill in details of the elicited test-selection strategy and, more

specifically, of the arguments underlying the strategy. We decided not to work with historical patient cases, since the experts might recall these patients and let the real final outcomes influence their test-selection behaviour. We decided to employ fictitious patient cases instead, that were designed to be as realistic as possible. Working with fictitious patient cases brought the additional advantage that it allowed us to design cases with which we were able to explore the experts' decision boundaries.

The setting

The method for eliciting test-selection strategies outlined above was used with two experts from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, who are the last two authors of the present paper. These experts had been involved in the construction of the decision-support system for oesophageal cancer from its very inception. Also present during the interviews were three researchers from the Institute of Information and Computing Sciences of Utrecht University; they are the first three authors. The first author has a background in medical computer science, providing her with some knowledge about cancer in general and about the various diagnostic tests involved. The second author has a background in mathematics. She had constructed the decision-support system, for which she had held numerous interviews. The two domain experts and this interviewer thus were very well acquainted with one another. Over the years, moreover, the second author had gained considerable knowledge about oesophageal cancer and its treatment. The third author, to conclude, has a background in cognitive science and knowledge acquisition.

The two interviews were conducted at the Netherlands Cancer Institute, the home institute of the two experts. The first author conducted the interview, asking the questions that had been prepared. We felt that the second author, because of her accumulated knowledge about oesophageal cancer and its treatment, might unknowingly and unwillingly bias the experts in their answers. She therefore did not partake in the main interview and only asked the more elaborate questions about the experts' decision boundaries that emerged during the interviews. The third author recorded the elicited knowledge and monitored the elicitation process. She typed the words from the interviews in a laptop, not just concentrating on relevant knowledge but also on remarkable meta-phrases uttered by the experts. We were aware that typing in a laptop was likely to result in a less accurate recording of the elicited verbalisations than taping with a voice recorder. Still a laptop was used instead of a voice recorder because the experts had previously indicated that they would feel hampered by the recorder. They did not seem to feel uneasy by the use of the laptop.

4 The First Interview

The first interview conducted with the two domain experts was an unstructured, oral interview. We briefly restate the main goal of the interview and the

procedure followed, before presenting the results.

The goal

The goal of the first interview was to elicit general knowledge about the selection of diagnostic tests for patients suffering from oesophageal cancer. The main issues to be addressed during the interview were:

- Are the experts guided by a standard procedure for selecting diagnostic tests, or are they mainly guided by their own experience?
- Are the various tests performed in parallel or sequentially?
- Are some tests always performed together, or one after the other? For instance, is the biopsy always combined with a gastroscopic examination of the oesophagus?
- What are the criteria that the experts use for selecting tests?
- What are the experts' criteria to stop testing?

The procedure

For the interview, we prepared a small number of open questions. The main question was "Can you describe the way in which you select and order diagnostic tests, starting from the very first consultation with a patient up to and including your final decision about the most suitable therapy?". Since we wanted to avoid biasing the experts, we let them talk freely and did not interrupt unless it was strictly necessary, for example when further elaboration was desired. These interruptions then only consisted of open questions such as "Why?" or "Can you describe what you are thinking right now?". The interview was conducted in the setting described in the previous section and took some 30 minutes.

The results

We found that upon first seeing a patient, the experts start with

- a physical examination of the patient;
- an interview with the patient, resulting in information about
 - the age of the patient;
 - the amount of weight loss suffered;
 - the patient's ability to swallow food.

Subsequently, independent of the results of the physical examination and interview, a number of diagnostic tests are ordered simultaneously:

- a gastroscopic examination of the oesophagus, resulting in information about
 - the shape of the primary tumour;

- the location of the tumour in the oesophagus;
- the circumference of the tumour;
- the length of the tumour;
- the presence of necrosis (substantial decay of tissue);
- a biopsy, mostly performed together with the gastroscopy, revealing
 - the histological type of the primary tumour.

In the sequel, we will refer to the physical examination, the interview, the gastroscopic examination and the biopsy together as the *starting package of tests*. The gastroscopic examination and biopsy serve to give insight in the presentation characteristics of the primary tumour. The physical examination and the interview with the patient result in an assessment of the patient's physical condition. We would like to note that, because our experts work at a highly specialised centre for cancer treatment, they generally see patients who are referred from regional hospitals where these tests have already been performed. Often, therefore, the test results are available. If the experts feel, however, that the tests have been performed too long ago they will order them to be performed anew.

After the results of the tests from the starting package have become available, the experts decide whether or not further testing is indicated. Patients with a very poor physical condition will now just receive highly palliative treatment, without having to undergo further testing. For all other patients, again a number of tests are ordered simultaneously:

- a CT-scan of the liver, lungs and thorax, resulting in information about
 - the presence of metastases in the loco-regional lymph nodes;
 - the presence of haematogenous metastases;
- an X-ray of the thorax, resulting in information about
 - the presence of haematogenous metastases in the lungs;
- a sonographic examination of the neck, providing information about
 - the presence of metastases in the lymph nodes in the neck;
- an endosonography of the local region of the primary tumour and of the mediastinum, giving insight in
 - the depth of invasion of the primary tumour into the oesophageal wall;
 - the presence of loco-regional lymphatic metastases.

These tests primarily serve to establish the extent of metastasis of the primary tumour. In the sequel we will refer to these four tests together as the *basic package of tests*. The tests from the basic package are again requested in parallel, but only after the results of the tests from the starting package have become available.

The remaining tests constitute the *extensive package of tests*:

- a bronchoscopy, resulting in information about
 - the depth of invasion of the primary tumour into the trachea and bronchi;
- a barium swallow with fluoroscopy, yielding insight in
 - the presence of a fistula (an open connection as a result of decay of tissue) between the oesophagus and the lungs;
- a laparoscopic examination of the liver, diaphragm and abdomen, resulting in information about
 - the depth of invasion of the primary tumour into the diaphragm;
 - the presence of haematogenous metastases in the liver;
 - the presence of metastases in the lymph nodes near the truncus coeliacus.

In contrast with the starting and basic packages of tests, not all tests from the extensive package are ordered just like that: one or more tests may be selected. Whether or not a specific test from the package is performed very much depends on the location of the primary tumour in the patient's oesophagus. If the tumour is located in the upper part of the oesophagus, a bronchoscopy and a barium swallow are performed to investigate whether or not the primary tumour has invaded the lungs. No laparoscopic procedures are performed, however, because the primary tumour cannot have invaded the diaphragm and, moreover, it is very unlikely that lymphatic metastases will be found in the upper abdomen. If the primary tumour is located in the lower part of the oesophagus, on the other hand, no bronchoscopy or barium swallow are performed. Laparoscopic procedures are then ordered, yet only if surgical treatment is considered.

To summarise, we found that diagnostic tests are ordered in three different packages: the starting package, the basic package, and the extensive package of tests. The tests from the starting package will always be ordered, for every patient. These tests are performed in parallel. If a patient is in a very poor physical condition, the experts will then stop testing and provide highly palliative care. For all other patients, the basic package of tests is ordered as well. In addition, one or more tests may be chosen from the extensive package, dependent on the location of the tumour and on the most preferred therapy at that particular moment in the patient's management. Figure 2 presents a

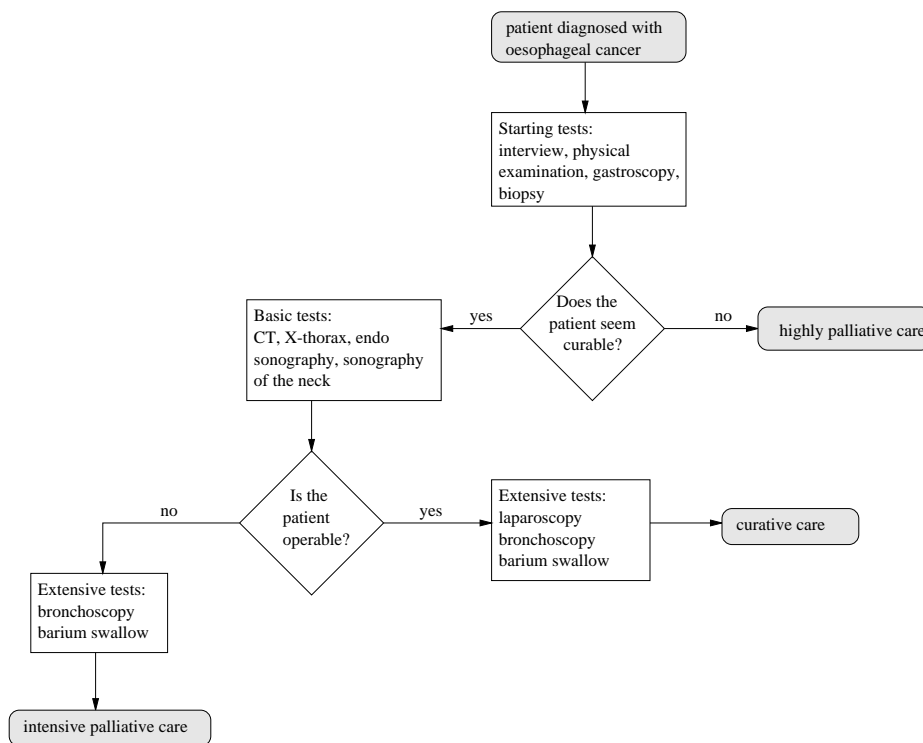


Figure 2: A flowchart summarising the knowledge from the first interview

flowchart summarising the knowledge acquired from the first interview. In the flowchart, the grey boxes with rounded corners describe the beginning and end points of the experts' strategy. The white rectangular boxes capture requests to perform specific tests; the diamonds represent alternative choices and capture the appropriate questions to decide upon which tests to order. The arrows in the chart indicate the sequential order in which the various decisions have to be taken and choices have to be made.

A discussion

From the first interview, various distinguishing features of the test-selection strategy employed by our experts emerged. We found that diagnostic tests are not ordered sequentially, where the decision whether or not to order a specific test depends on the result of a previous test. Instead, the tests are ordered simultaneously, in packages. The most important argument underlying the experts' strategy of parallel testing is the loss of time that would be incurred by sequential testing. It may take several weeks before the results of a test become available. The tumour may have progressed within that time and may thereby render the results from earlier tests obsolete. Moreover, patients often are in such a poor physical condition that it is preferable not to have them return to the hospital too often for yet another test. And, even more importantly, the loss of time may make the difference between a curable cancer and an incurable one.

As a consequence of ordering diagnostic tests simultaneously, however, more tests are likely to be performed than are strictly necessary. When questioned about such unnecessary testing, our experts indicated they did not see it as a problem, as the tests are not inconvenient for patients and gaining time is of primary importance:

"Ordering tests in packages is time saving. You might perform too many tests, but time is so important that you order all the tests anyway."

Our experts' strategy of ordering diagnostic tests in packages thus is supported by a strong argument. This argument in fact indicates that a sequential test-selection strategy for our decision-support system would be an unacceptable oversimplification of problem-solving practice. Our system should offer a strategy that is able to select tests in packages, based upon the argument reviewed above.

A second feature that we noticed of our experts' test-selection strategy pertains to the role of the stage of a patient's cancer. The experts had indicated before that they first establish the most likely stage for a cancer before deciding upon an appropriate therapy. We found, however, that a cancer's stage only very indirectly plays a role in the selection of diagnostic tests. The decision which tests to order appears in fact to be based upon the experts' current idea about the most suitable therapy for a patient rather than on the uncertainty about the stage of his or her cancer. For example, if the tests from the basic package reveal lymphatic metastases in distant lymph nodes, then surgical removal of the primary tumour is no longer a feasible treatment option for the patient under consideration. Invasive laparoscopic procedures for establishing the exact extent of the cancer's metastasis then are not performed, even though the stage of the patient's cancer is still uncertain. To establish the most appropriate treatment alternative for a patient, the experts appear to gather information that helps them weigh the beneficial effects and complications to be expected for each alternative. Starting with the best possible alternative, that is to perform a surgical procedure, they order tests to see whether or not such a procedure is contra-indicated. Only if the test results indicate that a treatment alternative is not feasible, do they focus their attention on the next-best alternative. We note that overtreatment of a patient may easily prove to be fatal; Under-treatment on the other hand may cause a patient to die prematurely from a cancer that might have been curable.

"You definitely do not want to miss anything. Missing and over-diagnosing are both bad. For physicians, however, missing feels worse than over-diagnosing as you want to do no harm. [...] When your feeling tells you something, but you cannot find it, you continue to look for it."

The test-selection strategy to be offered by our decision-support system should therefore take into account the impact of the possible results of a diagnostic test on the appropriateness of the available therapies rather than just its influence

on the uncertainty of the most likely stage of a patient’s cancer.

Closely related to the role of the stage of a patient’s cancer is the role of the reliability characteristics of the various diagnostic tests in the test-selection strategy employed by our experts. As we have briefly mentioned in Section 2, the literature on medical decision making generally stresses the importance of taking the sensitivity and specificity characteristics of diagnostic tests into consideration upon test selection [8]. We found, however, that these characteristics play no role with our experts, since they do not gather information to reduce their uncertainty about the most likely stage of a patient’s cancer. They also appear not to take the informativeness of the remaining available tests into account when they decide whether or not to stop testing. Their stopping criterion for ordering tests instead seems to be their certainty about the most appropriate treatment alternative for the patient at hand.

5 The Second Interview

From the first interview we had gained general insight in the test-selection strategy used by our two domain experts. Building upon the acquired knowledge, we followed up on the first interview with a second, more focused interview. Once again we briefly restate the goal of the interview and the procedure followed, before presenting the main results.

The goal

The goal of the second interview was to fill in the details of the general test-selection strategy that had been acquired from the two experts during the first interview. More specifically, we aimed at eliciting the exact arguments used by the experts in their decisions to order a new package of tests or to refrain from further testing.

The procedure

For this second interview, we wanted to walk through the entire process of test selection for a specific patient, from the very first moment the experts see the patient up to and including the selection of the most suitable therapy. To this end, we designed a structured interview in which we carefully simulated the experts’ daily problem-solving practice, by means of realistic patient cases. The experts were asked to think aloud while deciding for these patients, which tests to order.

For the interview, we created eight fictitious patient cases; The specifics of these are summarised in Figure 3. Because we wanted to obtain as many details as possible about the test-selection strategy used by our experts, we created both patients for whom the most appropriate therapy seemed obvious and patients for whom the best therapy was not so evident. The first patient case mentioned in Figure 3 is an example of a patient for whom the most suitable therapy is quite evident. We expected that the experts would decide to administer highly palliative care for this patient. We further expected that the experts would

not order the basic package of tests, nor any tests from the extensive package. For patient 6, however, it is not so clear what the best treatment alternative would be. The patient has a moderately-sized primary tumour and is in a very good physical condition. In fact, from the results of the tests from the starting package, the patient appears to be curable. We therefore expected that the experts would consider surgery and would order the basic package of tests. The results of the tests from the basic package reveal distant metastases. With this evidence and the primary tumour being distal, we expected that the experts would refrain from further testing and would decide to administer intensive palliative care.

The eight patient cases were carefully designed as illustrated above, by means of the flowchart that had resulted from the first interview. For each patient case, moreover, we prepared a small number of questions that allowed us to more closely investigate the experts' exact decision boundaries. For patient 7, for example, we expected that the experts would pronounce him to be incurable, mainly as a consequence of his poor physical condition. The patient's tumour, however, is very small and seems to be resectable. To investigate under which conditions the patient would no longer be deemed incurable, we prepared various what-if questions, such as "What would you do if the patient were just 58 years of age?" and "What would you do if the patient were in moderate health?".

For each patient case we prepared three cards, or vignettes with the results from the three different packages of tests. We created the three cards for every patient, also if it were very unlikely that even the basic package would be selected. The experts were informed that there were three cards per patient irrespective of whether or not we expected them to request the second package of tests or order tests from the extensive package. On each card, a result was indicated for each test from the package under consideration. Since the experts might feel that the first test results on a card are the most important, we decided to simply list the results in alphabetical order. We informed the experts about this alphabetical order to avoid biasing their problem-solving behaviour. Since the tests from a single package would be ordered in parallel, we decided to present the results of these tests simultaneously, and not one by one. As an example, Figure 4 shows the three cards that we prepared for patient 6.

We asked the experts to discuss each patient case aloud. We asked them more specifically to verbalise their subsequent reasoning steps in ordering tests and to conclude the discussion of a patient case with an indication of the most appropriate therapy. We asked them to pretend that they were ordering real tests for real patients. For each patient, the card with the results of the tests from the starting package was presented first. Only when the experts indicated that they would order additional tests, would we show the second card with the test results from the basic package. Upon studying the second card, the experts also had access to the first card; they could thus survey the accumulated patient data. When still further testing was desired, the last card was presented. If the experts did not order any test from the extensive package or even from the basic package, the associated cards were not shown. The interview was conducted in the setting described in Section 3 and took approximately two hours for the

Patient 1: An 87-year old male with a very poor physical condition and a large primary tumour.

Patient 2: A male of 76 years old who has a large primary tumour, yet is in good physical condition

Patient 3: A 57-year old male having a very small primary tumour who is in an excellent physical condition. The tumour is located in the upper part of the oesophagus (a proximal tumour).

Patient 4: This 60-year old male is in excellent condition and has a very small primary tumour located in the lower part of the oesophagus (a distal tumour).

Patient 5: A 64-year old male in good physical condition with a moderately-sized primary tumour. The patient has lymph node metastases in his neck.

Patient 6: This 67-year old male has a moderately-sized distal primary tumour and is in a good physical condition. He has both proximal and distal lymphatic metastases.

Patient 7: An 80-year old male in a very poor physical condition, having a very small primary tumour.

Patient 8: A 59-year old male with a large primary tumour, in a very good physical condition.

Figure 3: The eight patient cases designed for the second interview

- **Age:** 67 years
- **Biopsy:** adenocarcinoma
- **Gastroscopy:**
 - **Circumference:** circular
 - **Length:** between 5 and 10 centimeter
 - **Location:** distal
 - **Necrosis:** yes
 - **Shape:** scirrheus
- **Passage:** puree
- **Physical condition:** no COPD, normal heart condition
- **Weightloss:** less than 10%

(a)

- **CT:**
 - **Liver:** no
 - **Loco regio:** yes
 - **Lungs:** no
 - **Organs:** diaphragm
 - **Truncus coeliacus:** yes
- **Endosonography:**
 - **Loco regio:** yes
 - **Mediastinum:** no
 - **Wall:** T3
- **Sonography neck:** yes
- **X-lungs:** yes

(b)

- **Barium swallow:** no
- **Bronchoscopy:** no
- **Laparoscopy:**
 - **Diaphragm:** yes
 - **Liver:** no
 - **Truncus coeliacus:** no

(c)

Figure 4: The three cards representing the results of the tests from the starting package (a), the basic package of tests (b), and the extensive package (c)

eight patient cases.

The results

We present some fragments of the dialogue between the two experts while they were discussing patient 6, for whom the three cards shown in Figure 4 were created. Upon being presented with the first card, with the results of the tests from the starting package, the experts reasoned as follows:

"Oh, an average patient! We regularly see this type of patient. As his physical condition is quite good and the tumour is of moderate size, we might wish to consider surgery, so let's do the basic package of tests."

We presented the second card, with the results of the tests from the basic package:

"Mmm, there is some discrepancy here. Ah, well, we see them like this. Both proximal and distal metastases with a distal tumour. However, his condition is still quite good, so we would prefer to do something. We could consider palliative radiotherapy, also because he is not so old. [...] If the result from the sonography of the neck had been negative, there would only be metastases near the truncus coeliacus, which would make surgery a feasible option. Then, laparoscopy might be interesting. Now that the result of the sonography is positive, laparoscopy is no longer necessary."

The experts indicated that no further tests would be ordered.

The two experts discussed the eight patient cases at length. From the knowledge thus acquired, we constructed a new flowchart to capture the experts' test-selection strategy; the resulting flowchart is shown in Figure 5. To summarise, our observation from the first interview that diagnostic tests are ordered simultaneously in three different packages, was not contradicted by the second interview. Tests from the starting package are always performed. Only if a patient's physical condition is quite poor will the experts refrain from further testing and provide highly palliative care by positioning a prosthesis in the patient's oesophagus. For all other patients, the basic package of tests is ordered as well. If the results of the tests from this package reveal distant metastases, the experts will not order any new tests to be performed and start with local palliative care, that is, either a prosthesis is positioned in the patient's oesophagus or a palliative regime of radiotherapy is administered. If the metastases of the primary tumour are loco-regional, on the other hand, further tests will be selected from the extensive package to investigate whether or not surgical removal of the oesophagus is a feasible treatment alternative. If the tumour appears to be resectable, the appropriate tests from the extensive package will be performed. If it is evident that the tumour is not resectable, or if contra-indications for surgery have been found, such as a relatively poor heart condition, the laparoscopic procedures

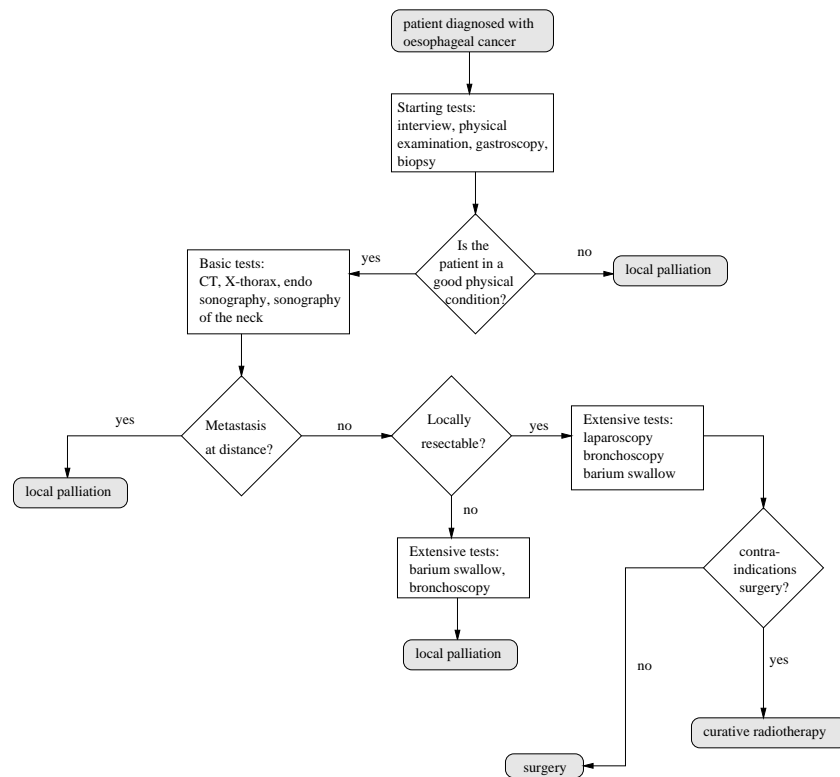


Figure 5: A flowchart summarising the knowledge from the second interview

will not be ordered.

A discussion

From the first interview we had learned various distinguishing features of the test-selection strategy employed by our experts. The second interview served to corroborate and further detail our previous observations. It provided additional insight especially in the way in which the experts used the results from the different tests as arguments for their subsequent decisions. The second interview further most prominently demonstrated that, from the very first moment of seeing a patient, the experts think in terms of appropriate treatment alternatives. In fact, our observations strongly suggest that it is the task of selecting an optimal therapy that drives the ordering of diagnostic tests, rather than the task of establishing the stage of a patient's cancer. For example, when discussing one of the patient cases, the experts mentioned:

"We are thinking of surgery as the best therapy right now. Let's see how deep the tumour has invaded into neighbouring organs to see if surgery really is an option."

6 Conclusions

Upon working with our decision-support system for oesophageal cancer, we felt that using the sequential test-selection strategies commonly proposed in the decision-making literature would be an oversimplification of our experts' daily problem-solving practice. We decided to acquire knowledge about the actual strategy used by the experts to provide for the design of a tailored test-selection strategy. For this purpose we used an elicitation method that was composed of two focused interviews: an unstructured interview, followed up by a structured one. With the first, unstructured interview, we found that tests were ordered not sequentially but in three different packages, with the tests from a single package ordered simultaneously. With the second, highly structured interview, we were able to fill in the details of the general strategy that we had elicited. More specifically, we were able to establish the different arguments underlying the experts' test-selection decisions.

With the structured interview, we carefully simulated the experts' problem-solving practice through the use of fictitious patient cases. Each patient case was captured by three different cards, or vignettes, with the results of the three packages of tests. These cards were presented sequentially to the experts. We found that our approach indeed closely fitted in with our experts' daily practice. In fact, they explicitly mentioned that using the cards was very intuitive:

"These cards and the way we discuss them are very similar to how patients are presented during the sessions we have with colleagues when we discuss patients."

The use of the cards thus worked quite well. The only difference with the experts' daily problem-solving practice may have been that in the current interview setting, not facing a real patient and with less time pressure, the experts were more consistent and more thorough in their decisions than they usually are. One of the experts mentioned:

"Perhaps we are now more consistent than we normally are in practice."

We felt that linking up with practice was highly advantageous for the purpose of acquiring knowledge of the test-selection strategy used by our experts. We would like to note, however, that especially for the set-up of the second, structured interview, prior knowledge appeared to be imperative. Without prior knowledge, we would not have been able to design the fictitious patient cases in a way that allowed us to explore the experts' decision boundaries.

To conclude, we feel that a test-selection strategy offered by a decision-support system should support physicians in their daily problem-solving practice and should therefore be based upon the argument experts use in their decisions to order specific tests. We feel that to design such a strategy, knowledge about the actual test-selection strategy used should be elicited from experts in the domain of application. A standard sequential strategy may then turn out to be unacceptable to the physicians who are the projected users of the system. In this paper, we have demonstrated that eliciting test-selection knowledge from experts can indeed be feasible and is likely to result in a wealth of detailed information that can provide for the carefully tailored test-selection strategy.

Acknowledgements. The research of the first two authors was (partly) supported by the Netherlands Organisation for Scientific Research (NWO). We are most grateful to Eveline Helsper and Silja Renooij for their useful comments on an earlier draft of this paper.

References

- [1] L.C. van der Gaag, S. Renooij, B.M.P. Aleman, B.G. Taal, 'Evaluation of a probabilistic model for staging of oesophageal carcinoma', in: A. Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell, H.-U. Prokosch (Editors) *Medical Infobahn for Europe*, Proceedings of MIE2000 and GMDS2000, IOS Press, Amsterdam, pp. 772 – 776, 2000.
- [2] P.C. Fishburn, *Utility Theory for Decision Making*, John-Wiley & Sons, New York, 1970.
- [3] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Statistics for Engineering and Information Science, Springer, New York, 2001.
- [4] S. Andreassen, 'Planning of therapy and tests in causal probabilistic networks', *Artificial Intelligence in Medicine*, vol. 4, pp. 227 – 241, 1992.

- [5] P. Doubilet, 'A mathematical approach to interpretation and selection of diagnostic tests', *Medical Decision Making*, vol. 3, pp. 177 – 195, 1983.
- [6] J.St.B.T. Evans, 'The knowledge elicitation problem: a psychological perspective', *Behaviour and Information Technology*, vol. 7, 2, pp 111 – 130, 1988.
- [7] J.M. Schraages, S.F. Chipman, V.L. Shelin (Editors), *Cognitive Task Analysis*, Lawrence Erlbaum Associates, Mahwah, NJ, 2000.
- [8] H.C. Sox, M. Blatt, M.C. Higgins, K.I. Marton, *Medical Decision Making*, Butterworth, Kent, 1987.
- [9] A.Th. Schreiber, J.M. Akkermans, A.A. Anjewierden, R. de Hoog, N.R. Shadbolt, W. van de Velde, B.J. Wielinga, *Knowledge Engineering and Management: The CommonKADS Methodology*, MIT Press, Massachusetts, 2000.
- [10] M.W. van Someren, Y. Barnard, J. Sandberg, *The Think Aloud Method – a Practical Approach to Modelling Cognitive Processes*, Academic Press, London, 1994.
- [11] K.A. Ericsson, H.A. Simon, *Protocol Analysis: Verbal Reports as Data*, revised edition, MIT Press, Cambridge, 1993.