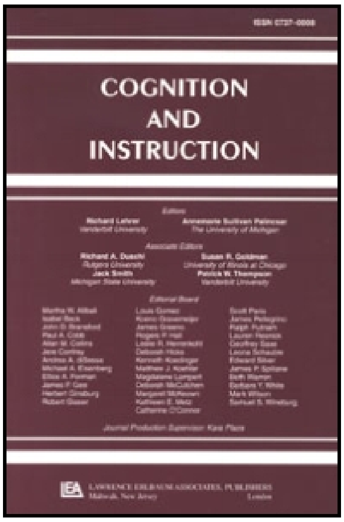

Cognition and Instruction
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775648096

The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results From an Interview Study

Eugéne J. F. M. Custers; Henny P. A. Boshuizen; Henk G. Schmidt

Online publication date: 14 December 2009

To cite this Article Custers, Eugéne J. F. M. , Boshuizen, Henny P. A. and Schmidt, Henk G.(1998) 'The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results From an Interview Study', Cognition and Instruction, 16: 4, 367 — 398

To link to this Article: DOI: 10.1207/s1532690xci1604_1
URL: http://dx.doi.org/10.1207/s1532690xci1604_1

COGNITION AND INSTRUCTION, *16*(4), 367–398

The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results From an Interview Study

Eugène J. F. M. Custers
Department of Medical Informatics
Catholic University Nijmegen
Nijmegen, The Netherlands

Henny P. A. Boshuizen
Department of Educational Research and Development
Maastricht University
Maastricht, The Netherlands

Henk G. Schmidt
Department of Psychology
Maastricht University
Maastricht, The Netherlands

In this article, we describe a study in which some current ideas about illness scripts are tested. Participants at 4 levels of medical expertise were asked to describe either a prototypical patient or the clinical picture associated with a number of different diseases. It was found that participants at intermediate levels of expertise mentioned, both absolutely and relatively, many enabling conditions (patient contextual factors such as sex, age, medical history, and occupation) when asked to describe a prototypical patient with a disease, whereas the instruction to describe the clinical picture of a disease revealed a monotonic relation with expertise level. The amount of biomedical information in the descriptions decreased with increasing expertise level

Requests for reprints should be sent to Henny P. A. Boshuizen, Department of Educational Research and Development, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: boshuizen@educ.unimaas.nl or e.custers@mie.kun.nl

for both types of instruction. In addition, a positive relation was found between number of actual patients seen with a particular disease and number of enabling conditions mentioned. These results were interpreted as supportive of the present conceptualization of the illness script theory.

The question concerning the essential differences between novices and experts in a particular domain continues to generate a large number of studies. Obviously, an important reason for this sustained interest is the fact that the expertise issue in the recent past has been approached from many domains, including cognitive psychology, artificial intelligence, education, and instructional science. In other words, expertise research has implications for both theory construction (e.g., How do expert knowledge and expert skills develop? How can expert knowledge structures best be described?) and practical educational implementation (e.g., How should educational environments be designed in order to facilitate knowledge and skill development as much as possible?). Although the origins of the study of expertise development can be found in domains in which educational and instructional benefits are not conspicuous, such as chess and problem solving (e.g., de Groot, 1946/1965; Newell, Shaw, & Simon, 1958; Newell & Simon, 1972), probably until now the bulk of the research has been performed in the domain of medicine. Nevertheless, a considerable amount of work has been done or is still in progress in other areas, for example, computer programming, industrial troubleshooting, and physics; in recent years, expertise in domains such as auditing, history, and military strategic thinking has also become the subject of study (for a review of some recent developments, see Custers, 1995).

This study, which focuses on medical expertise but has implications for many other complex domains too, has its roots in a number of findings from previous studies on medical expertise. First, it has become increasingly clear that what distinguishes medical experts from nonexperts is primarily the quality of the pertinent knowledge and not, for example, experts' presumed superior general problem-solving abilities or their better execution of prescribed diagnostic procedures (cf. Boshuizen & Schmidt, 1992; Elstein, Shulman, & Sprafka, 1978; Feltovich, Coulson, Spiro, & Dawson-Saunders, 1992; Patel & Groen, 1986; Schmidt, Norman, & Boshuizen, 1990). Second, in medical diagnostic situations, these differences in knowledge structure exert their influence from the very first moment of the consultation: Experts have what might be called a head start as far as diagnostic accuracy is concerned (Barrows, Norman, Neufeld, & Feightner, 1982; Elstein et al., 1978; Hobus, Schmidt, Boshuizen, & Patel, 1987). For example, Elstein et al. reported that, in about 90% of diagnostic consultations, a first diagnostic hypothesis is activated as soon as the patient has phrased his or her main complaint, and subsequent research has revealed a positive relation between the quality of these early hypotheses and the level of expertise of the participant (Elstein et al., 1978;

Johnson, Durán, Hassebrock, Möller, & Prietula, 1981; Neufeld, Norman, Feightner, & Barrows, 1981). Third, however, it has also been found that a number of important performance measures do not consistently differentiate between experts and novices. For example, recall memory for case information and extent of pathophysiological explanations often show an "intermediate" effect: Participants at intermediate levels of expertise outperform both novices and experts (Schmidt & Boshuizen, 1993b). If displayed in graphical form, an intermediate effect is revealed by an "inverted-U" relation, indicating that participants at the intermediate levels of expertise show performance extremes on some measure (e.g., speed, accuracy, and number of statements in protocol). Although experts and novices may perform similarly on that measure, this is not a necessary condition for the effect to occur: The general effect is simply an initial increase on a variable from novice to intermediate, followed by a decrease on that variable from intermediate to expert. Supposedly, if the variable in question reflects some form of cognitive processing, the origin of the phenomenon can be found in intermediates' more extensive elaboration of their knowledge base when they are asked to perform. This hypothesis is supported by the finding that, in conditions in which case processing time is restricted, the intermediate effect gives way to a more linear relation between expertise level and performance (Schmidt & Boshuizen, 1993b). The explanation of this disappearance of the intermediate effect is, of course, that time restrictions prevent participants from elaborating on their knowledge base. In addition, Custers (1995) found essentially equal performances of preclinical students, recently graduated medical doctors (MDs), and experienced family physicians on a scrambled case reconstruction task (i.e., a task in which participants had to reconstruct a case by selecting appropriate case statements from a pool of medical information that also included "noise").

If recognition performances are similar for participants at different levels of expertise, whereas recall performances seem to peak at intermediate levels, how can it be that expert physicians show superior diagnostic performance from the very beginning of a consultation? In other words, what are the specific aspects of expert physicians' knowledge structures that enable them to outperform both novices and intermediates on diagnostic tasks but are apparently unrelated to either recall or recognition memory performance? From a behavioral point of view, the difference is mainly functional: Less experienced participants are able to recall and recognize the important features of a patient suffering from a particular disease, but they have difficulty in using this information in diagnostic settings. A possible explanation for this limited ability to exploit knowledge that is basically available for recognition and recall purposes, in diagnostic situations, is offered by the illness script theory.

According to Feltovich and Barrows (1984), illness scripts are hypothesized general knowledge structures that consist of three components: enabling condi-

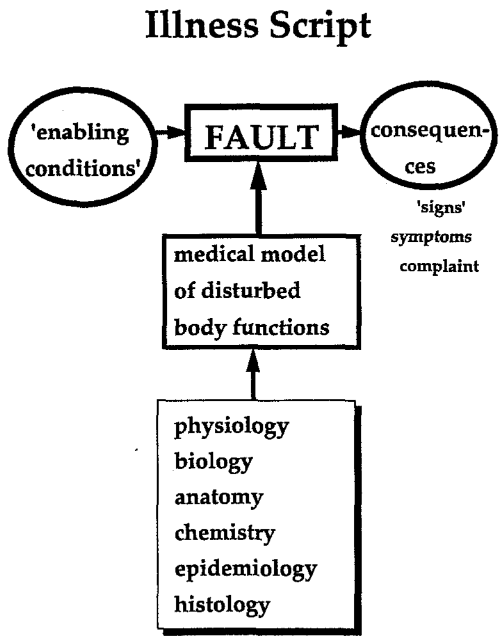

FIGURE 1 General structure of the illness script (based on Feltovich & Barrows, 1984).

tions, a fault, and consequences. *Enabling conditions* are contextual and patient background factors that influence the probability that someone gets a disease. Examples of enabling conditions are age, sex, medical history, current medication, risk behavior, hereditary factors, occupation, and living environment. These enabling conditions may contribute to the *fault,* the latter being the pathophysiological malfunctioning that constitutes the biomedical core of a disease. This fault may give rise to certain *consequences:* complaints, signs, and symptoms. Figure 1 (adopted from Feltovich & Barrows, 1984) shows a schematic example of an illness script. It is assumed that physicians possess an illness script for each disease they know, whereas the extent to which each script is "filled in" depends on the actual experience the physician has with the disease.

As such, illness scripts can be considered special variants of the "classical" scripts, as described by Schank and Abelson (1977). These scripts, tailored to predominantly social situations, are defined as generalized event sequences: a description of things that may be expected to occur in situations many people are familiar with, such as going to a restaurant, flying by plane to another city, or getting up in the morning. A characteristic feature of scripts is that some aspects are fixed, such as the presence of a waiter and a bill in a restaurant situation, whereas other aspects may be variable or optional, such as the age and sex of the waiter and whether soup is served or not. In terms of Schank and Abelson, scripts consist of slots that can be filled with certain values. Another feature of scripts is that they are activated as wholes in appropriate situations and, hence, enable people to make quick predictions about what is going to happen. A basic mechanism supporting this function is the assignment of default values to empty slots. For example, unless stated otherwise, people will assume that a menu is presented after one sits down in a restaurant. The filling in of current and default values in an activated script is called *script instantiation.*

Similarly, an illness script is thought to be activated and instantiated by a physician in a diagnostic situation. For diseases, as for events, there are relatively fixed features and more variable ones. In addition, an activated illness script may enable a physician to predict what will likely happen with a particular patient in the (near) future. There are, however, some conspicuous differences between the classical scripts and illness scripts. First, unlike classical scripts, illness scripts serve no social or personal purpose and are not part of someone's plans. Obviously, in a medical diagnostic situation, both patient and physician will have plans and goals, but these are not part of the illness script structure, and they are more general (e.g., the doctor wants to diagnose accurately and treat properly, and the patient wants to be cured or at least to get rid of the complaints). In addition, illness scripts do not have slots for actors: Although both patient and physician may take actions to reveal the nature or change the course of a disease, they do not play social roles in terms of the illness script (although they do play such roles in terms of a "physician consultation script," which is a true script in the Schank & Abelson, 1977, sense; for a further elaboration of the differences between classical and illness scripts, see Custers, 1995).

Illness script theory in its present form has been developed by Schmidt et al. (1990), Schmidt, Boshuizen, and Norman (1992), Custers (1995), and Custers, Boshuizen, and Schmidt (1996) and incorporates aspects of both the Feltovich and Barrows (1984) and Schank and Abelson (1977) theories. According to this perspective, the biomedical details of the fault do not play an important role in the diagnostic process, particularly in routine diagnostic situations. The main features used to generate diagnostic hypotheses—the latter being a process that boils down to activating and instantiating illness scripts—are the enabling conditions, together with one or a few consequences, because these are the features most readily available (cf. Hobus, Boshuizen, & Schmidt, 1990; Hobus et al., 1987). However, in diagnostic situations, less experienced participants often have trouble recognizing a certain constellation of enabling conditions and consequences (i.e., they fail to activate the appropriate illness script) and have to revert to fault-based or biomedical reasoning to make sense of the features that confront them (cf. Patel, Evans, & Groen, 1989; Schmidt & Boshuizen, 1993a). More expert participants might also show biomedical reasoning, for example, on difficult diagnostic problems (cf. Patel, Groen, & Arocha, 1990) and on decontextualized cases (cf. Lesgold et al., 1988; Norman, Brooks, & Allen, 1989). According to Boshuizen and Schmidt (1992), Schmidt and Boshuizen (1993a, 1993b), and Schmidt et al.

(1992), repeated application of biomedical knowledge results in encapsulation of this knowledge: Biomedical concepts become subsumed under higher level, clinical knowledge structures, which consist of concepts in which specific enabling conditions and consequences are firmly tied together. Consequently, these concepts facilitate early recognition of diseases or, in other words, quick activation of appropriate illness scripts. Meanwhile, the lower level biomedical concepts remain available, should deeper explanation be required (a process called *unfolding*), but in relatively routine diagnostic situations, they play a limited role.

There is some evidence that the reported relative failure of inexperienced physicians to use enabling conditions in diagnostic situations (Hobus et al., 1990; Hobus et al., 1987) should be attributed specifically to a relatively poor development of the enabling conditions component of these physicians' illness scripts. As outlined previously, experts are better diagnosticians than novices because their knowledge is organized and structured in full-fledged illness scripts, with the enabling conditions (and some readily available consequences) providing the opportunity for a quick script activation and, consequently, easy access of this disease knowledge in actual diagnostic situations. It is assumed that biomedical "reasoning through" and subsequent encapsulation of knowledge are not sufficient for the development of the enabling conditions component of illness scripts but that extended experience with real patients in practical settings is also required. Therefore, for two reasons, the role of overt biomedical knowledge in clinical reasoning decreases: First, biomedical knowledge has become encapsulated, and hence, extensive biomedical reasoning to interconnect the different features that characterize a specific disease is no longer necessary. Second, with increasing practical experience, participants are able to recognize patterns consisting of mainly enabling conditions and (a few) consequences. In summary, illness script development is characterized by a shift in importance from fault-related biomedical aspects toward patient-related enabling conditions. Throughout this development, the consequences component remains important, even for very experienced participants, but receives relatively more emphasis in the earlier stages of development of diagnostic expertise.

Although one may be tempted to conclude that less experienced physicians simply lack much of the relevant knowledge concerning the enabling conditions of diseases, there is evidence that this is not the case. For example, Hobus, Boshuizen, and Schmidt (1989) asked participants to describe, for a number of diseases, prototypical patients with these diseases. Although less expert participants' descriptions, on the average, included fewer enabling conditions than expert family physicians' descriptions, the differences were relatively small: Beginning and expert physicians mentioned, on the average, 2.61 and 3.07 enabling conditions per patient, respectively. Thus, it seems more likely that the Hobus et al. (1990) and Hobus et al. (1987) findings might better be explained by nonexperts lacking the ability to use enabling conditions in a diagnostic situation rather than by complete absence of this knowledge. Consequently, it may be hypothesized that knowledge of this illness script component is not absent in intermediates[1] but that it is less well integrated into their illness script structures. In this view, intermediates' diagnostic knowledge may be described as collections of relatively scattered fragments held together by immature illness scripts. A characteristic feature of this stage appears to be that the fault and consequences components are relatively well developed, whereas the integration of the enabling conditions component into the script structure lags behind. Completing this process of integration of patient background characteristics into illness script structure may require extended experience with actual patients and, hence, may continue until well after graduation.

To test this view, a study was designed in which the procedure used by Hobus et al. (1989) was both replicated and extended. As already mentioned, in this latter study, intermediates and experienced family physicians were presented with a number of complaint–diagnosis pairs. Their task was to describe a prototypical[2] patient with that diagnosis (disease), who would present himself or herself with the particular complaint. For example, one of the descriptions was "Can you describe a prototypical patient suffering from herpes zoster, who presents himself with the complaint of a terrible pain in a specific, well-circumscribed region on the chest?" As already mentioned, the predicted effect—that experienced physicians' patient descriptions would contain more enabling conditions than those of 6th-year students—was found, although not to the point of warranting the conclusion that intermediates actually lack knowledge of enabling conditions.

It may be argued that the design of the Hobus et al. (1989) study was not sufficiently sensitive to reveal the actual developmental course of illness scripts. First, participants at only two levels of expertise took part in the study. It may be assumed that the 6th-year students, being quite advanced in both biomedical and clinical training, already hold full-fledged illness scripts for at least a number of diseases. This would explain both the relatively high number of enabling conditions mentioned by these participants and the relatively low number of biomedical statements in their descriptions, as reported by Hobus et al. Second, asking participants to describe prototypical patients for each of the diseases may have clouded the actual phenomenon of increasingly better integration of enabling conditions into illness script structures: Using this probe may have increased in less experienced participants the awareness of patient background factors, which in turn may have resulted in a spuriously high number of enabling conditions in their descrip-

[1]In the Hobus et al. (1989) study, the least experienced participants were beginning physicians who had recently graduated from medical school. Because these participants obviously cannot be considered novices, we will refer to them as intermediates.

[2]In the Hobus et al. (1989) experiment, as in the one to be presented in this article, participants were asked to describe a prototypical patient. Although what was intended was a "typical" patient, the expression "prototypical" was used for two reasons: (a) because a strong association with the prototypicality theory of categorization was assumed to hold, and (b) the word *typical* has a somewhat different flavor in Dutch compared to English. In Dutch, a "typical patient" might be interpreted as a "peculiar case."

tions. If the key difference between experts' and nonexperts' knowledge is that the experts have integrated enabling conditions into their illness scripts, whereas nonexperts have not, then it may be expected that experienced physicians will always activate enabling conditions, regardless of the way their knowledge is probed or accessed, and intermediates will only mention enabling conditions if explicitly encouraged to, for example, by being asked to describe a prototypical patient.

Therefore, in this study, the Hobus et al. (1989) design was extended with regard to two important aspects. First, two additional expertise levels were included: 4th-year students, who had extensive knowledge of biomedical sciences but virtually no practical clinical experience, and interns, that is, postgraduate medical students in training as family physicians. We hypothesized that the 4th-year students, lacking practical experience with patients, would tend to describe diseases mainly in terms of consequences and underlying biomedical malfunctions. The interns, on the other hand, should be located, as far as their illness script development is concerned, somewhere between the 6th-year students and the experienced physicians. Second, in this study, two probes were employed: One half of the participants, like those in the Hobus et al. study, was asked to describe a prototypical patient with a specific disease; the other half was asked to describe the clinical picture of the disease. It was assumed that this latter probe would activate the same signs and symptoms for a particular disease as the instruction to describe a prototypical patient but would certainly not focus participants' attention specifically on the contextual factors. For example, factors such as age, sex, and medical history will probably be part of the description of a prototypical patient, but it may be less obvious to include these factors in a description of the clinical picture of a disease. Nevertheless, it may be expected that the presence of full-fledged illness scripts will be characterized by the inclusion of some information about enabling conditions regardless of the specific kind of probe; hence, experts will always provide at least some information about enabling conditions. Because intermediates' illness scripts still lack the necessary coherence, they will activate their knowledge of enabling conditions only under certain circumstances, for example, when being asked to describe a prototypical patient, but will be less inclined to do so when being probed in a different way. Participants at the lowest levels of expertise, that is, preclinical students, will not be strongly inclined to mention enabling conditions regardless of type of probe; they will describe prototypical patients and clinical pictures mainly in terms of consequences and fault-related (i.e., biomedical) aspects.

In addition to these extensions, some aspects of the procedure in the Hobus et al. (1989) study were modified in this experiment. For example, we provided participants with only the diagnoses (names of diseases); we gave no information about the presenting complaint in either of the conditions. Although this modification may decrease the comparability between the two studies, it should be emphasized that, according to illness script theory, the presenting complaint is an important consequence of the disease, which we did not want to "give away" immediately. Moreover, providing participants with the presenting complaint in addition to the probe would reduce the difference between the two experimental conditions, and consequently, the expected differential effect of the probes might also decrease. Thus, compared to the Hobus et al. study, it may be expected that participants in this experiment will include more consequences in their descriptions, particularly because it is likely that the usual presenting complaints of a disease are relatively salient consequences.

As stated previously, in our view, experience with actual patients plays an important role in illness script development. Hence, it may be expected that participants' descriptions of clinical pictures or prototypical patients are influenced by their actual experience with the diseases in question. To investigate this issue, a simple inventory was devised. For each disease included in the study, participants had to estimate the number of actual patients they had seen. It was expected that this inventory, as opposed to the rather crude measure of general level of expertise, would enable a more fine-grained analysis of the influence of actual experience on the information provided by participants in an interview study such as this one.

In summary, if enabling conditions become increasingly integrated in illness script structures, then the number or proportion of enabling conditions produced in a free production task will increase monotonically with expertise level. Conversely, the contribution of biomedical, or fault-related, information in the free production protocols will decline with increasing expertise level because physicians use this knowledge in encapsulated form, at least in routine diagnostic situations. Furthermore, because enabling conditions are fully integrated into experts' illness scripts but not into those of intermediates, a statistical interaction between expertise level and probe on especially the enabling conditions information produced will be expected: Experienced physicians will provide some patient contextual and background information regardless of the way they are probed, whereas participants at intermediate levels will be more inclined to access their knowledge in accordance with the content of the probe. In other words, intermediates will include enabling conditions in their descriptions if they are accessed directly by the probe but not if the probe is more tangential to this illness script component. Compared to intermediates, participants at the lower extreme end of the expertise scale will not, or to a lesser extent, be influenced by type of probe: They will just tell what they know, primarily knowledge of biomedical and consequences aspects of diseases.

In addition, it is predicted that a positive relation will be found between number of patients seen with a specific disease and the number of enabling conditions mentioned. Similarly, a negative relation will exist between number of patients seen and amount of biomedical, fault-related information in participants' descriptions. Thus, apart from expert physicians having, in general, more full-fledged illness scripts than less experienced participants do, intermediate-level participants may have full-fledged illness scripts for some diseases they are already familiar with but not for other, less common ailments.

METHOD

Participants

Participants were 23 fourth-year medical students, 22 sixth-year medical students, 23 second-year interns, and 22 family physicians. The 4th-year students had no or negligible experience in practical clinical settings, whereas the 6th-year students, interns, and physicians had an average of 2 years, 5.5 years, and 13.9 years of clinical experience, respectively. All participants were either studying at the University of Limburg at Maastricht, The Netherlands, or practicing in the Maastricht area, including a nearby practice in Belgium.

Materials

The stimulus material consisted of 20 names of diseases used in the Hobus et al. (1989) study. Earlier experiments showed that contextual information could play a facilitative role in the activation of accurate diagnostic hypotheses for these diseases (Hobus et al., 1990; Hobus et al., 1987). Furthermore, the diseases were selected to display substantial variance in seriousness of the illness, afflicted organ system, and frequency of occurrence in real-life situations. Table 1 lists the names

TABLE 1
Diseases Used in the Study

A.	Metastatic sigmoid cancer (metastases in the lungs)
B.	Dyspepsia on a nervous basis (also called nervous gastritis)
01.	Aneurysm of the aortic artery (threatening rupture)
02.	Urosepsis
03.	Dermatitis peri-oralis
04.	Vaginal candidiosis
05.	Perforated otitis media
06.	Kidney stones colic
07.	Carcinoma of the head of the pancreas
08.	Stomatitis aftosa (multiple small ulcera in the mouth)
09.	Secondary enuresis nocturna
10.	Digitalis intoxication
11.	Epidural hematoma
12.	Nervous abdominal pain
13.	Pediculosis pubis
14.	Herpes zoster
15.	Meningitis or encephalitis as a complication of mumps
16.	Hepatitis A
17.	Monilia of the mouth
18.	Pre-infarct syndrome

Herpes zoster	0	1	2	3	4	5	6–10	11–15	16–20	20+

FIGURE 2 Rating scale item of the patient frequency inventory.

of the diseases. The first two diseases (A and B in Table 1) were used for practicing purposes only.

The materials also included a patient frequency inventory, a sheet of paper on which a rating scale for each of the 20 selected diseases was printed. A 10-point rating scale was used, with values ranging from 0 to 20 or more patients seen. Participants were to indicate on each scale how many patients with that particular disease they had seen; if they were unable to provide an exact number, they were asked to estimate. In case of doubt, the participant could mark two adjacent points on a scale. Figure 2 shows a rating scale used in the inventory. Participants were asked to consider only those cases seen in an actual clinical situation, disregarding all other instances (e.g., relatives or friends who might have had the disease). In addition, a patient seen two or more times should be counted as one instance of the disease.

Procedure

Participants were, in order of participation, assigned to one of the two conditions. Participants in the *prototype* condition were asked to describe prototypical patients; those in the *clinical picture* condition were required to describe the clinical picture of each of the 20 diseases. A small pilot study had shown that these instructions were generally well comprehended; if participants asked informative questions (e.g., "Should I tell something about the treatment?"), a relatively frequent occurrence, the experimenter's standard reply was "Yes, if you think that is part of the clinical picture [prototypical patient]" (depending on the experimental condition). In line with Clancey's (1984) report that his expert physician found it easy to describe a typical case for the main diagnoses in NEOMYCIN (Clancey's intelligent tutoring system), we also found that participants, even at the lower end of the expertise scale, had no trouble in describing typical cases. Appendix A shows some examples of participants' descriptions. If the beginning of a description suggested that the participant had a wrong disease in mind, he or she was interrupted and corrected by the experimenter; however, this was rarely necessary. To avoid misunderstandings, some of the diseases were announced by both their medical name and a more colloquial expression. Any information given by participants about diseases not included in this study was discarded from the analysis.

Participants were tested individually. All narratives were audiotape recorded; there were no time constraints on the duration of either the entire session or the individual disease descriptions. The modal duration of a session was 20 to 30 min. At the end of the session, the patient frequency inventory was presented to the

6th-year students, the interns, and the family physicians. The 4th-year students had virtually no practical experience with any of the diseases.

Analysis

Because there was a major problem with respect to the nature of secondary enuresis nocturna (Disease 9 in Table 1), all data on this disease were excluded from the analysis.[3] For the remaining 17 experimental diseases, verbatim protocols were derived from the audiotape recordings. The protocols were analyzed into statements, that is, medically relevant information units, and subsequently categorized into the major illness script categories: enabling conditions, fault, and consequences. For example, if a patient was described as an elderly woman, this was counted as two enabling conditions statements: one about the sex of the patient and one about her age. If the patient complained about pain in the epigastric region, this counted as two consequences: the general kind of complaint (pain) and the location (in epigastrio). If the complaint was further specified (e.g., a radiating pain), this counted as an additional consequence. Similarly, if a participant mentioned reduced blood supply in coronary arteries, this counted as two fault statements: one about the afflicted organ and one about the biomedical phenomenon. Statements that did not fit into one of the three categories but seemed medically relevant were classified into a fourth category: *CDFT* items. This latter category, which coincides with the "nonmodel" items in the Hobus et al. (1989) study, eventually included four types of statements—about the expected *course* of the disease (e.g., "this patient's condition usually improves within a few days"), about *diagnostic activities* to be performed (e.g., "you can only establish a definite diagnosis by taking a biopt"), about *frequency of occurrence* (e.g., "you don't see that very often"), and about *treatment* (e.g., "it is difficult to treat")—hence, the indication CDFT category. Because the number of items in this category was too small to split up for statistical analysis, it will be dealt with as one miscellaneous category.

An expert physician, who had experience with the illness script categories, was consulted in order to establish a sound basis for the classification. Generally, not many problems were encountered during categorization. If in serious doubt, statements were omitted from the analysis; however, this was hardly ever necessary. Sometimes, categorization of a specific item was determined by the context in which it appeared: For example, *hypercholesterolemia* was classified as an en-

[3]Secondary enuresis nocturna is bed-wetting as a consequence of some disease or malfunction of the uro-genital tract. However, many participants started describing a patient with (or the clinical picture of) primary enuresis nocturna, that is, bed-wetting without pathophysiological problems. Faced with their incorrect interpretation, many participants admitted "never having heard of" bed-wetting as a consequence of a medical problem.

abling condition when it occurred in a statement such as "these are patients known with elevated serum cholesterol levels" but as a fault item when it appeared in a statement such as "this is caused by the formation of plaque at the walls of the coronary arteries, as a result of elevated cholesterol levels." Note that the correctness of a statement with respect to a particular disease was immaterial for classification.

As free production tasks usually result in large interindividual differences in total number of statements, this task being no exception, analyses were performed on the actual number of statements in each category as well as on the respective proportions. Because proportions always add to 1 and, thus, are not independent, the results of this latter analysis should be interpreted with caution. Number and proportion of statements in each of the illness script categories were analyzed by means of a 4 (levels of expertise) × 2 (probes) analysis of variance (ANOVA).

As far as the data of the patient frequency inventory concerns, the number of patients seen by each participant for each of the diseases was determined. If participants marked Categories 6–10, 11–15, or 16–20, the arithmetic mean of the respective category (i.e., 8, 13, or 18) was taken as the number of patients seen by that participant. Participants who marked Category 20+ were always assigned the value 20. If participants marked two adjacent categories, the arithmetic mean of the two categories was assigned. Pearson correlations were computed for each disease over participants between the number of patients seen and the number of statements in each of the illness script categories. Finally, a correlational analysis of the patient frequency data based on diseases, rather than participants, was performed.

RESULTS

Number and Proportion of Statements in Different Illness Script Categories

For the 17 diseases included in the analysis, participants mentioned an average of 9.97 statements per disease. Table 2 shows the total number of statements produced

TABLE 2
Mean Total Number of Statements Mentioned for Each Expertise Level and Type of Probe

	Expertise Level				
Type of Probe	*4th-Year Students*	*6th-Year Students*	*Interns*	*Family Physicians*	*M*
Prototypical patient	8.34	11.06	9.78	11.13	10.06
Clinical picture	9.12	10.85	9.33	10.25	9.88
M	8.72	10.96	9.57	10.73	9.97

for the four expertise levels and the two probes. ANOVA revealed a borderline significant effect of expertise level on the total number of statements produced, $F(3, 82) = 2.452, p < .07, MSE = 9.777$, but no significant effect of experimental condition and no significant interaction between expertise level and experimental condition. Fourth-year students and interns were somewhat less prolific than 6th-year students and family physicians. However, the data do not show an interpretable linear or monotonic trend, and there is no evidence for the existence of a so-called intermediate effect, that is, participants at the intermediate levels performing better than participants at either of the extreme ends of the expertise scale. Thus, we found no easily interpretable relation between level of expertise and verbal output, and we found that the two different types of cues did not lead to significant differences in the amount of information provided by the participants.

Number and Proportion of Enabling Conditions Mentioned

ANOVA showed a significant main effect of expertise level, $F(3, 82) = 8.831, p < .0001, MSE = 0.877$, and condition, $F(1, 82) = 24.303, p < .0001, MSE = 0.877$, on the number of enabling conditions mentioned by participants. In addition, a significant interaction between expertise level and condition, $F(3, 82) = 2.548, p < .05, MSE = 0.877$, was found. Table 3 shows that the number of enabling conditions mentioned increases with experience, at least up to the level of the interns. Furthermore, the instruction to describe a prototypical patient led to twice as many enabling conditions in the stories than the instruction to describe the clinical picture: In this latter condition, about one enabling condition was mentioned for every disease, whereas in the former, on average, more than two enabling conditions were produced. In addition, the significant interaction suggests that the effect of the

TABLE 3
Mean Number (per Disease) and Mean Percentage of Enabling Conditions Mentioned as a Function of Expertise Level and Type of Probe

	Prototypical Patient		*Clinical Picture*		*M*	
Expertise Level	*n*[a]	%[b]	*n*[a]	%[b]	*n*[a]	%[b]
4th-year students[c]	0.91	11.8	0.68	6.7	0.80	9.3
6th-year students[d]	2.17	20.5	0.98	8.6	1.58	14.6
Interns[c]	2.97	31.3	1.13	11.6	2.09	21.9
Family physicians[d]	2.47	22.2	1.57	15.5	2.06	19.2
M	2.17	21.7	1.08	10.5	1.63	16.3

[a]Mean number of statements categorized as enabling conditions per disease. [b]Percentage of total number of statements mentioned (averaged over all diseases). [c]$n = 23$. [d]$n = 22$.

probe differed between expertise levels. Table 3 shows that, although the number of enabling conditions mentioned in the prototypical patient condition seems to level off at the more advanced levels of expertise, the number of enabling conditions mentioned increases monotonically with level of experience if participants are instructed to describe the clinical picture of a disease.

As far as the relative contribution of enabling conditions in the protocols is concerned, ANOVA showed a significant main effect of expertise level, $F(3, 82) = 9.88$, $p < .0001$, $MSE = 69.228$, a significant main effect of condition, $F(1, 82) = 37.025$, $p < .0001$, $MSE = 69.228$, and a significant interaction between expertise level and condition, $F(3, 82) = 3.815$, $p < .05$, $MSE = 69.228$, on the proportion of statements in the enabling conditions category. Table 3 shows that the percentage of enabling conditions generated by the participants increased with experience through the level of the interns but leveled off for the experienced family physicians. Table 3 also shows that the finding of a significant interaction in this case means that, in the prototypical patient condition, an inverted-U relation between expertise level and percentage of enabling conditions seems to hold, whereas in the clinical picture condition, the percentage of enabling conditions seems to increase monotonically with experience. A separate one-tailed *t* test for the interns and family physicians in the prototype condition showed that the decline in proportion of enabling conditions at the highest level of experience indeed was significant, $t(22) = 2.578$, $p < .01$. Therefore, it can be concluded that, if instructed to describe a prototypical patient with a disease, experienced physicians mention proportionally fewer enabling conditions than beginning family physicians. If participants are instructed to describe the clinical picture of a disease, however, the proportion of enabling conditions reported increases monotonically with level of expertise.

In general, the enabling conditions data support our hypothesis that preclinical students lack certain knowledge of enabling conditions of diseases. Preclinical students are disinclined to mention patient background factors, even if probed to describe a prototypical patient. Participants at the intermediate levels of expertise, in contrast, possess knowledge of relevant patient background factors but tend to volunteer this knowledge only if the instruction directly suggests it is relevant. Finally, experienced family physicians report a relatively large number of enabling conditions, even if not directly cued toward this illness script component.

Number and Proportion of Fault Items Mentioned

ANOVA showed a significant main effect of expertise level on the number of fault items, $F(3, 82) = 7.932$, $p < .0001$, $MSE = 1.022$. However, neither a significant main effect of probe type nor a significant interaction between probe type and expertise level was found. Table 4 shows, for both experimental conditions, a monotonic decrease with expertise level in the number of fault statements men-

TABLE 4
Mean Number (per Disease) and Mean Percentage of Fault Items Mentioned as a Function of Expertise Level and Type of Probe

	Prototypical Patient		*Clinical Picture*		*M*	
Expertise Level	*n*[a]	%[b]	*n*[a]	%[b]	*n*[a]	%[b]
4th-year students[c]	1.70	18.5	1.93	19.6	1.81	19.1
6th-year students[d]	1.30	10.5	1.67	13.7	1.48	12.1
Interns[c]	0.52	4.6	0.90	9.4	0.70	6.9
Family physicians[d]	0.52	4.3	0.62	5.7	0.57	4.9
M	1.01	9.3	1.30	12.3	1.14	10.7

[a]Mean number of statements categorized as fault items per disease. [b]Percentage of total number of statements mentioned (averaged over all diseases). [c]$n = 23$. [d]$n = 22$.

tioned. These results are in line with our expectations: First, biomedical knowledge is particularly important at the lower levels of experience, and second, because neither of the two probes cued participants specifically to produce biomedical knowledge, we predicted no significant differences in the number of fault items reported.

Similarly, ANOVA of the relative contribution of fault items in the protocols showed a significant main effect of expertise level, $F(3, 82) = 17.316$, $p < .0001$, $MSE = 53.772$, on the percentage of fault items mentioned but neither a significant effect of condition nor a significant interaction between condition and expertise level (see Table 4). Thus, the more experienced participants' reporting less biomedical information is not an artifact of their providing less information in general than inexperienced participants provided.

Number and Proportion of Consequences Mentioned

ANOVA also showed a significant main effect of expertise level, $F(3, 82) = 3.253$, $p < .05$, $MSE = 3.081$, on the number of consequences mentioned but neither a significant effect of condition nor a significant interaction. From Table 5, it can be seen that the number of consequences provided by the participants displays a pattern quite similar to that for the total number of statements produced: 6th-year students and family physicians provided about seven consequences statements per disease, whereas 4th-year students and interns mentioned about six items of this type. Contrary to expectations, the probe to describe the clinical picture of a disease did not result in a significantly larger number of consequences produced; both instructions gave rise to approximately six consequences on average.

ANOVA of the proportion of consequences showed a different pattern: No significant main effect of expertise level was found, but the effect of experimental

TABLE 5
Mean Number (per Disease) and Mean Percentage of Consequences Mentioned as a Function of Expertise Level and Probe

	Prototypical Patient		*Clinical Picture*		*M*	
Expertise Level	*n*[a]	%[b]	*n*[a]	%[b]	*n*[a]	%[b]
4th-year students[c]	5.47	66.3	5.74	66.6	5.60	66.4
6th-year students[d]	6.52	60.6	7.51	72.0	7.02	66.3
Interns[c]	5.38	55.7	6.54	71.9	5.94	63.5
Family physicians[d]	7.08	65.4	6.48	62.0	6.81	63.8
M	6.10	61.9	6.57	68.3	6.33	65.0

[a]Mean number of statements categorized as consequences per disease. [b]Percentage of total number of statements mentioned (averaged over all diseases). [c]$n = 23$. [d]$n = 22$.

condition was significant, $F(1, 82) = 4.768$, $p < .05$, $MSE = 173.091$, as was the interaction between expertise level and experimental condition, $F(3, 82) = 2.794$, $p < .05$, $MSE = 173.091$. Table 5 shows that participants presented with the clinical picture probe mentioned, on the average, proportionally more consequences, but this effect seems to be completely accounted for by the data of the participants at the intermediate levels of expertise, who mentioned relatively more consequences than either 4th-year students or family physicians. Thus, as far as the proportion of consequences is concerned, participants at the intermediate levels of expertise indeed appear to be most sensitive to experimental instruction.

Analysis of the CDFT Category

On average, participants in the prototypical patient and clinical picture conditions mentioned 0.82 and 0.94 CDFT items, respectively. Because these numbers are small, further analyses yielded mostly insignificant results. In addition, items in the CDFT category are of a very heterogeneous nature; therefore, the results of any analysis of the category as a whole would be difficult to interpret. Probably the only result worth mentioning is that a Kruskal–Wallis one-way ANOVA by ranks (cf. Siegel, 1956) showed a significant effect of expertise level (H corrected for ties = 9.715, $df = 3$, $p < .05$) on the number of statements mentioned in this category; as such, this result is in line with the finding of Hobus et al. (1989) that participants at higher levels of expertise provide more nonmodel information.

In summary, the data on the number and percentage of statements in the different illness script categories as a function of expertise level show a mixed picture: Some predictions were clearly born out, whereas others were not. In addition, some rather unexpected results were found. The decrease with expertise level, in both the absolute number and proportion of fault-related items, and the concomitant increase in number and proportion of items on course, diagnosis, frequency, and therapy were consistent with our expectations. Regarding the number of consequences, the predicted effects of level of experience (i.e., the intermediate or inverted-U effect) and type of probe (i.e., more consequences when asked to describe the clinical picture) did not materialize; analysis of the proportion of consequences, in contrast, revealed these expected results. The analysis of number and proportion of enabling conditions resulted in a surprising effect. Although, as expected, less experienced participants were more sensitive to type of probe than expert physicians, the finding that the interns outperformed even the experts in the number of enabling conditions mentioned when asked to describe a prototypical patient was not predicted. However, if asked to describe the clinical picture of a disease, the expected monotonic increase in the number and proportion of enabling conditions mentioned with expertise level was found.

Relation Between Actual Clinical Experience With the Diseases and Characteristics of the Descriptions

As expected, there were large differences in the frequency with which the individual diseases are encountered in daily clinical life. Diseases like dyspepsia on a nervous basis, perforated otitis media, and vaginal candidiosis have high frequencies of occurrence, whereas a threatening rupture of an aneurysm of the aortic artery, carcinoma of the head of the pancreas, and epidural hematoma are rare; even experienced physicians may never have seen a patient with one of these diseases. Appendix B contains frequency data for all diseases.

Because we had no patient frequency data for the 4th-year students, the following results are based on the three more advanced levels of expertise ($n = 67$). For these participants, Pearson correlation coefficients between the number of patients seen with a specific disease and the number of enabling conditions mentioned for that disease were, for all diseases except one, in the lower positive range: from .09 to .32, with an average value of .17. The one exception was herpes zoster (Disease 14 in Table 1), for which a correlation of −.10 was found. Because these data are definitely subject to a restriction of range effect (i.e., for a considerable number of diseases, many participants had seen few cases and mentioned few enabling conditions), we chose to perform further analyses for the group of diseases as a whole, rather than for individual illnesses. By adding the frequency values for all 17 experimental diseases for each participant,[4] an estimation of the total number of patients seen by each individual was formed. Analysis of these estimates revealed

[4]If participants had marked a category that covered more than one value (e.g., 11–15; see Figure 2), the average value of this category (e.g., 13) was taken as the number of patients seen with that disease.

that 6th-year students, interns, and experienced physicians had seen an average of 51 (±27), 126 (±38), and 219 (±39) patients, respectively, with the 17 experimental diseases included in the study. A caveat should be made before further analyses are pursued. Because participants were not required to differentiate between frequencies of patient encounters above 20 (see Figure 2), the actual number of patients they had seen with the diseases may be grossly underestimated, particularly for participants at the upper end of the expertise scale. However, although the actual values of the added frequency scores may not be very telling, there is no reason to assume that they do not accurately reflect the rank order of the participants, in terms of clinical experience with the diseases. Moreover, although the current diseases were not explicitly selected with the intention of constructing a fully representative sample of afflictions faced by family physicians in their daily practice, a case can be made for the assumption that the summed scores of the patient frequency inventory actually do reflect, at least to a certain extent, participants' clinical experience in general medicine.

Illness script theory predicts that the role of enabling conditions in disease knowledge increases with practical experience and that this is reflected by an increase in the number of enabling conditions mentioned by participants in a free production task. These data allow for two ways to test this hypothesized relation between actual experience with the diseases and the contribution of enabling conditions in participants' descriptions of prototypical patients or clinical pictures.

First, Pearson correlation coefficients between participants' clinical experience in general, as assessed by an estimation of the total number of patients seen with the 17 experimental diseases and the number of statements they mentioned for the different illness script categories, were computed, as reported in Table 6. Two correlations are significant, namely the correlation between clinical experience and the number of enabling conditions mentioned by 6th-year students, and the correlation between clinical experience and the number of fault items mentioned over all partici-

TABLE 6
Pearson Correlation Coefficients Between Clinical Experience and Number of Statements in Respective Illness Script Categories

	No. of Statements in Illness Script Categories				
Level of Expertise	*Enabling Conditions*	*Fault*	*Consequences*	*CDFT*	*Total*[a]
6th-year students[b]	.63*	–.11	.09	.14	.26
Interns[c]	.10	–.07	–.28	–.02	–.13
Family physicians[b]	.01	.02	–.08	.05	–.03
All participants[d]	.24	–.37**	–.07	.18	–.01

[a]Total number of statements mentioned. [b]$n = 22$. [c]$n = 23$. [d]$n = 67$.
*$p < .01$, one-tailed. **$p < .005$, one-tailed.

pants. Apparently, for 6th-year students, a positive relation exists between the amount of clinical experience in general and the number of enabling conditions in the protocols, whereas for the interns and family physicians, no such relation can be discerned. Although theoretically the effect may be spurious (i.e., due to a third, yet unknown factor), this is unlikely because the 6th-year students are a homogeneous group in terms of formal medical experience, all having 4 years of preclinical training and almost 2 years of clerkships. In fact, in our view, the correlation found (.63) is remarkably high, particularly because the two variables involved (clinical experience and amount of patient background information volunteered in an interview) are highly disparate and can only be distantly related. Hence, the most likely explanation for the association is that participants are faced with real patients and, as a consequence, include salient background features of these patients in their memory representations. Assuming that relatively few patient encounters are sufficient for this effect to occur, it can explain the correlation drop at higher levels of expertise. If you have already seen quite a number of patients with a disease, the benefit of additional ones will be limited, at least in constructing an illness script (but not necessarily in strengthening or tuning an already existing illness script). In contrast, the drop in the correlations at the higher end of the expertise scale between number of enabling conditions mentioned and patient experience may be exaggerated because of the inability of our measure of clinical experience to discriminate effectively between participants at these more expert levels.

The negative relation between clinical experience and number of fault items mentioned over all participants is consistent with the hypothesis that the importance of biomedical knowledge decreases with experience. Although the actual correlation (–.37) is moderate, again, in light of the disparate character of the two variables and the low discriminating power of the measure of clinical experience, it suggests that clinical experience in general plays a role in the decrease of importance of biomedical information in the mental representations of patients.

Second, a convergent but slightly different approach was taken. For each disease, the average number of cases participants at a particular expertise level had seen and the average number of statements they mentioned in specific illness script categories were calculated. Correlation coefficients were computed, this time with diseases rather than participants as origin.[5] Positive correlations between the mean frequency of patient encounters and the mean number of enabling conditions mentioned by participants at a particular level of experience would support the hypoth-

[5]This procedure was performed as a substitute for computation of individual participant-based correlations between number of patients seen and number of statements in different illness script categories for each disease; because many participants had seen few patients with (some of) the diseases and, in addition, mentioned few statements in some illness script categories, serious restriction of range effects would prevent possibly existing relations between participants' practical experience and the nature of the information they provided from being revealed.

TABLE 7
Pearson Correlation Coefficients Between the Mean Number of Patients Seen With Each Disease and the Mean Number of Statements Mentioned in Each Illness Script Category (Computed Over 17 Diseases)

	Mean No. of Statements in Illness Script Categories				
Level of Expertise	*Enabling Conditions*	*Fault*	*Consequences*	*CDFT*	*Total*[a]
6th-year students	.44*	.12	.29	.22	.40
Interns	.50**	–.05	.13	.43*	.33
Family physicians	.35	–.24	.21	.29	.34
All participants	.44*	–.15	.24	.29	.34

[a]Total number of statements mentioned.
*$p < .1$, one tailed. **$p < .05$, one-tailed.

esis that the integration of enabling conditions in knowledge structures is a consequence of actual experience with diseases. In other words, participants' knowledge representations of diseases that are more frequently encountered in practical situations will generally include more enabling conditions too. Table 7 gives the results.

Although statistically significant only for the interns, at all three levels of expertise, a positive relation appears to exist between the instances of a particular disease encountered and the number of enabling conditions mentioned. The data in Table 7 also suggest that, for diseases encountered more frequently, participants tend to include more consequences statements and CDFT statements in their descriptions ($r = .13–.43$). In contrast, the relation between practical experience with a disease and number of fault items mentioned seems absent or negative.

In summary, whereas the previous analysis (Table 6) showed, at least for 6th-year students, that clinical experience in general is associated with the inclusion of enabling conditions into the descriptions, the correlations in Table 7 support the hypothesis that actual experience with a specific disease is positively related to inclusion of enabling conditions knowledge into the illness script structure for that disease. In other words, having seen more patients in general is associated, at least for relatively inexperienced participants, with mentioning more enabling conditions overall; having seen more patients with a specific disease is associated with mentioning more enabling conditions for that disease, probably for participants at all expertise levels.

DISCUSSION

The inclusion of two additional expertise levels and an additional experimental probe into a replication of the Hobus et al. (1989) study yielded a considerably more complex picture than the original study. By including four expertise levels, we were able to investigate the developmental trend of illness scripts at a more fine-grained level and to discover whether the development of the different illness scripts components showed monotonically increasing, monotonically decreasing, intermediate, other, or no relations with expertise level. By including two types of probes, it was possible to find out whether participants are sensitive to differences in type directions given to report on a disease. In addition, the simultaneous inclusion of expertise level and type of probe as experimental variables enabled us to detect interactions between these factors; by investigating the effects of this interaction for four different illness script components, an even more detailed picture could be given. Finally, asking participants how much practical experience they had with the different diseases included in the study enabled us to investigate relations between illness script development and experience with specific diseases, rather than merely between illness script development and general expertise level.

The most salient difference between these results and those of Hobus et al. (1989) is the failure of this study to replicate the increase, with expertise level, of the number of enabling conditions mentioned when participants are asked to describe a prototypical patient with a particular disease. The inclusion of interns in our study suggests that the number and proportion of enabling conditions mentioned in this task either remains approximately constant for participants from about 18 months clinical experience onward, as the data on the absolute number of enabling conditions suggest, or shows an inverted-U effect, with a maximum for the interns, as the data on the proportion of enabling conditions mentioned seem to indicate. When participants were asked to describe the clinical picture associated with a particular disease, however, we found a continuous increase with expertise in both the absolute number and the proportion of enabling conditions mentioned.

In addition, in contrast with the Hobus et al. (1989) study, but in line with the illness script theory, we found a decrease in the number and proportion of fault statements mentioned as expertise increased. The absence of a significant interaction between expertise level and experimental condition for fault items indicates that this decrease holds for both types of probes. This finding was also expected because neither the instruction to describe a prototypical patient nor the instruction to describe the clinical picture specifically tunes participants toward mentioning biomedical knowledge.

Furthermore, the average number of consequences mentioned in this study was considerably larger than in the Hobus et al. (1989) study. In our study, 6th-year students and experienced physicians mentioned 6.52 and 7.08 consequences (cf. Table 5), respectively, whereas the corresponding values in the Hobus et al. study amounted to 4.44 and 2.45 consequences, respectively. Because Hobus et al. provided participants with the main complaint of the patient to be described, whereas we did not, the finding that participants volunteered more consequences did not come as a surprise: The main complaint of a disease is obviously a major consequence. More important, however, we did not find the expected decrease in number of consequences mentioned by participants at the highest level of expertise. In fact, the effects of both expertise level and type of instruction on number and proportion of consequences in the protocols are difficult to interpret. It is hard to conceive of a theory that can explain the peculiar, N-shaped relation between level of expertise and number of consequences mentioned. Comparing the data in Table 2 and Table 5 suggests that the number of consequences mentioned parallels the total number of items mentioned over all four categories. Perhaps participants who want to expand on their patient or disease descriptions are inclined to add consequences to their descriptions, rather than items in any of the other illness script categories. Thus, the number of consequences mentioned may be determined to a large extent by quantitative differences in verbal output. It is also salient that the differences in proportion of consequences, averaged over the two types of instruction, are extremely small; at all levels of expertise, participants' descriptions seem to consist of about two thirds of consequences.

Because neither of the instructions specifically oriented participants toward including CDFT items in their descriptions, it is not surprising that participants mentioned few of these items. The increase in the number of CDFT (or nonmodel) items with expertise level reported by Hobus et al. (1989), however, was replicated in our experiment. Although this result suggests that experienced participants may include some knowledge about course, diagnosis, frequency, and treatment into their illness script structures, the heterogeneous nature of the category suggests that it is not, as a whole, a natural part of the script structure. Some CDFT items may actually be part of one of the other script components. For example, knowledge of the course of a disease may be regarded as knowledge of possible future consequences and may be included in this illness script component. Knowledge of frequency of occurrence, to the extent that it plays a role early in the diagnostic process, may be captured in the illness script model by the role of enabling conditions. For example, a physician will be more inclined to activate hypotheses about malignancies for older patients than for younger patients, even if frequencies are small in any age group.

In general, the results of this study support our current view of the development of illness scripts. This perspective holds that, in the early stages of medical expertise development, the basic sciences play an important role in constructing scripts for diseases. With increasing expertise, the role of this biomedical knowledge lessens, whereas the role of contextual factors of diseases simultaneously becomes more important. This developmental trend is reflected by the "stories" told by our participants. Whereas biomedical and pathophysiological information forms a substantial part of the stories told by 4th-year students, interns and family physicians provide hardly any information of this kind. This finding is also in line with a broader theoretical framework concerning the development of medical expertise that basically encompasses illness script theory (Schmidt et al., 1990). The absence of an interaction between expertise level and type of instruction on number of fault items mentioned (i.e., under both instructions, a monotonic decrease with expertise level was found) attests to the pervasiveness of this developmental trend. Neither of the two instructions encourages participants to provide biomedical knowledge; hence, as they become more experienced, they are increasingly less inclined to mention it because it is generally not included in their automatically activated illness scripts. It cannot be concluded from these data, however, that expert physicians do not possess the relevant biomedical knowledge for the diseases used in the study, only that biomedical knowledge occupies a less prominent place in their practically oriented knowledge structured as illness scripts.

Whereas the data on the fault category are completely in line with our expectations, the results concerning enabling conditions force us to modify previous conclusions. Remember that Hobus, Hofstra, Boshuizen, and Schmidt (1988), Hobus et al. (1990), and Hobus et al. (1987) found experienced physicians' diagnostic performances to increase considerably when contextual information about a patient was provided, compared to a situation in which this information was absent, whereas 6th-year students were unable to profit from this additional source of information. Thus, Hobus et al. (1988) concluded that the illness scripts of experienced physicians are enriched with an enabling conditions component, with 6th-year students still lacking much of this knowledge. However, because we found that 6th-year students, compared to experts, do not lag behind in the number of enabling conditions produced if asked to describe a prototypical patient, it seems unlikely that a simple lack of knowledge is responsible for the Hobus et al. (1990), Hobus et al. (1988), and Hobus et al. (1987) results. Rather, knowledge accessibility appears to be the critical aspect. Sixth-year students know quite a lot about contextual information, but they only access this knowledge when it is explicitly cued, as with the instruction to describe a prototypical patient with a disease. In other circumstances, such as the diagnostic situation employed by Hobus et al. (1990), Hobus et al. (1988), and Hobus et al. (1987) or our request to describe the clinical picture of a disease, enabling conditions are apparently activated to a lesser extent by nonexperts. In this study, this conclusion is bolstered by the interaction between expertise level and instruction: The prototypical patient instruction triggered enabling conditions, even at less advanced levels of expertise, whereas in the clinical picture condition, the number of enabling conditions volunteered increased monotonically with expertise level. In this condition, the importance of contextual information is less conspicuous, and as a consequence, this type of knowledge is less readily accessed by nonexpert participants. Thus, on basis of our results and in light of the findings of Hobus et al. (1988) and Hobus et al. (1989), we conclude that, although knowledge of contextual factors is already present at lower levels of expertise—a proportion of 30% enabling conditions mentioned by interns in the prototypical patient condition buttresses this conclusion—integration of this knowledge into illness script structures is a relatively slow process, extending well beyond graduation.

Although we have no direct evidence, the data may be partly explained by the assumption that interns are more apt to invoke the image of a specific, recently seen patient with a disease than are experienced physicians, who tend to rely more on general structures (i.e., illness scripts), with irrelevant information—including irrelevant enabling conditions—filtered out. This hypothesis directly opposes current views of medical expertise development, in which memories of actual patients are the ultimate knowledge structures (e.g., Brooks, Norman, & Allen, 1991; Schmidt et al., 1990). This hypothesis would be supported if it could be shown that nonexperts provide more enabling conditions that are not in a medically meaningful way related to the disease in question but that are possibly based on a single highly atypical but exceptionally well remembered case. Although this explanation—that expert physicians' enabling conditions are qualitatively, rather than quantitatively, different from those of 6th-year students—cannot be definitely excluded, these data do not seem to favor it: Age, sex, and medical history were by far the most frequently mentioned enabling conditions at all expertise levels, and there are no indications of qualitative differences in content of these enabling conditions between participants at different levels of experience.

This study also provides evidence that, at least for participants with only a few years of practical clinical experience, the integration of enabling conditions into illness scripts is a consequence of experience with real patients, rather than a general effect of becoming more knowledgeable in medicine. For these participants, moderately positive correlations were found between clinical experience in general, as measured by total experience with the diseases used in this study, and the number of enabling conditions mentioned. In addition, based on particular diseases, positive correlations were found between the number of patients seen by participants at each of the expertise levels (although these correlations were significant only for the two intermediate levels) and the number of enabling conditions mentioned. Thus, if a disease is quite often seen by participants, relatively many enabling conditions will be mentioned when they are asked to describe a prototypical patient or the clinical picture of that disease, whereas no such relation was found between actual experience and number of items in the fault or consequences categories. It should be emphasized that we think that these results are meaningful, not because the size of the relation is impressive but because the variables that are related are so disparate: rough estimations of number of patients seen with a specific disease and amount of patient background information given in a laboratory interview task. Illness script theory can explain why this specific relation was found and not, for example, a more general relation between expertise level and total amount of information included in patient or disease descriptions.

In summary, our findings support the hypothesis that illness script development at the intermediate levels consists of the gradual integration of enabling conditions into knowledge structures, a process that is fostered by experience with actual patients, both in general and as far as specific diseases are concerned. As such, our results are in line with those of Weber, Böckenholt, Hilton, and Wallace (1993), who also found no general effect of expertise level on activation of diagnostic hypotheses but, instead, found a specific effect (of expertise level) on availability of diagnosis, which may be considered a direct effect of experience with diseases. Apart from this outstanding exception, surprisingly little attention has been paid in the literature to the nature of diagnostic hypothesis generation, either in the more practically oriented educational work or in more experimental studies. After Elstein et al.'s (1978) conclusion that it is not the quantity but the quality of generated hypotheses that distinguishes expert from nonexpert physicians, the issue of hypothesis generation has received, in our view, too little attention, the aforementioned studies by Hobus et al. (1990), Hobus et al. (1987), and Weber et al. (1993) notwithstanding. There is evidence that patient background factors, in relation to the complaint, constitute main triggers for pertinent hypotheses. However, there is still little information on the nature of the mistakes nonexperts make if they fail to activate the correct hypothesis. Some studies (e.g., Wagenaar, 1987) suggest that people can live with hypotheses that fly in the face of at least some of the information in the problem description, but the importance of this in a medical educational context is as yet unclear. A thorough investigation might reveal the "traps" nonexperts are likely to fall into or might point to knowledge deficiencies that have a certain amount of generality. If there is a relation between type of information in a case (enabling conditions, fault, and consequences) and nonexperts' diagnostic errors, illness script theory may provide valuable clues as to the correction or even prevention of these errors. From an educational point of view, this might be an interesting avenue to explore.

These results also have implications for instructional practice. First, they support the educational approach of providing students in the clerkship stage, or even before, with as much practical clinical experience as possible. As far as the development of illness scripts is concerned, this practical experience should emphasize, apart from symptoms, signs, and complaints, the patient background and contextual factors, rather than an elaborate explanation of biomedical and pathological processes occurring in the body of the patient. In addition, for a widely applicable base of illness scripts to develop, students in the early stages should be exposed especially to patients with frequently occurring, preferably nonimpressive diseases. Although practical objections may of course prevent faculties from large-scale implementations of practical clinical instruction in the early stages of medical education, the argument that less advanced students are unable to benefit from such experiences does not seem to be valid, at least not in light of these results.

Second, an early introduction of practical clinical experiences should not be interpreted as a plea for accepting or fostering "shallow" knowledge development by de-emphasizing the importance of the biomedical sciences or decreasing the amount of time spent on these subjects. Rather, we are arguing that biomedical knowledge development and clinical experiences should be linked as much as possible to help students understand that underlying fault and perceptible consequences are often, but not always, closely connected. For example, it might be easier for students to acquire expertise in cardiology if they are allowed to take an oral history and to listen to heart murmurs and other physical signs of actual cardiac patients in addition to studying and constructing explanations of the pathophysiological mechanisms behind the cardiac problems. An optimal alignment of biomedical knowledge and practical experience might accelerate the process of encapsulation (cf. Schmidt & Boshuizen, 1993a, 1993b) necessary for the development of the appropriate illness scripts. Such alignment may also increase advanced students' awareness that, in some cases, especially the more difficult ones, it is necessary to construct a complete biomedical explanation, encompassing all findings, whereas in other cases, a more "shallow" approach (i.e., diagnostic labeling) might be sufficient. Providing preclinical students on a small scale with carefully selected, preferably prototypical, cases devoid of irrelevant information might favor their learning process. This idea is not new, of course; for example, it was already advocated by Norman (1988). Similarly, Van Rossum, Bender, and Meinders (1991) argued that teachers should be cautious in embellishing cases used for educational purposes with too salient a context. If context is important for the activation of illness scripts and diagnostic hypotheses, then it will be surely advantageous if this context, in cases used for educational purposes, is relevant for the disease in question. Although it may be particularly the occasional instances of rare diseases and peculiar patients that make a clinician's job interesting and challenging, effective and efficient routine diagnosis may be the most important characteristic of an outstanding expert physician.

Medical diagnosis, as well as medical practice in general, has a number of features that make it a unique area of expertise, although these individual features can be found in other domains as well. These features include (a) an underlying basic model; (b) a complex "interface" that mediates between the expert and the model (i.e., the patient features at the clinical level); (c) probabilistic relations between aspects of the model and features of the interface; (d) the necessity of the expert making decisions, often under time pressure; (e) highly limited abilities to manipulate; (f) a setting that often incites negotiating behavior; and (g) the involvement of personal and social variables, such as emotions and values. In all domains in which these aspects come into play, they should be learned; sometimes, but not always, they can be effectively taught.

This article addressed two aspects of diagnosis: (a) the importance of context, or background factors, in diagnosis; and (b) the relation between knowledge of the underlying model and of the interface—the outwardly perceptible aspects of a system or organism. These aspects are important in medical diagnosis but also in other professional domains, such as auditing (cf. Vaatstra, 1996) or troubleshooting in complex automated systems (cf. Schaafstal, 1991), and in many everyday situations. In medicine, illness scripts are the knowledge structures presumably relevant for coping with diagnostic problems. Hence, in other domains, similar script-like structures may be responsible for solving diagnostic problems at the expert level. For example, in human-created systems, it is often easily assumed that the working of the system as a whole can be construed as the sum of the workings of the individual parts. Although this may be true in principle, many of these systems may be too complicated for a single person to fully comprehend. People who are in charge of controlling such a system and detecting possible errors may not be able to explain its behavior in terms of the basic underlying model and may have to rely on clinical symptoms, possibly even including contextual features. In general, this type of knowledge will emerge with increasing expertise; no formal instruction may be involved. It cannot be excluded, however, that explicit instruction involving repeated application of model knowledge and increased sensitivity to context may accelerate the development of well-tuned knowledge structures. Similarly, in a highly analytical domain such as auditing, in which basically everything is determined by explicit rules embodied by an underlying model, experts may gradually develop scripted structures. "When I have to audit a firm I am not familiar with, I first take a walk around the building, try to form an impression, have a peek at the cars on the parking lot, especially those of the managers and the board of directors," an expert auditor once said (R. Vaatstra, personal communication, April 22, 1993). This may be comparable to the first impression an expert physician forms from a patient, one that he is able to fully exploit for diagnostic purposes. In most complex professional domains, except medicine, the development of relevant knowledge structures has not been systematically studied, let alone the question of whether development of such structures can be fostered by some form of instruction. Although the content of these professional domains may differ widely, we would not be surprised if the processes underlying knowledge development and the structure of the knowledge representation would be similar.

ACKNOWLEDGMENTS

Parts of this article were presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1992, as well as in Boshuizen, Hobus, Custers, and Schmidt (1992) and Boshuizen, Schmidt, Custers, and Van de Wiel (1995).

REFERENCES

Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical & Investigative Medicine, 5,* 49–55.

Boshuizen, H. P. A., Hobus, P. P. M., Custers, E. J. F. M., & Schmidt, H. G. (1992). Cognitive effects of practical experience. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 337–348). Berlin, Germany: Springer-Verlag.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16,* 153–184.

Boshuizen, H. P. A., Schmidt, H. G., Custers, E. J. F. M., & Van de Wiel, M. W. (1995). Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction, 5,* 269–289.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

Clancey, W. J. (1984). Methodology for building an intelligent tutoring system. In W. Kintsch, J. R. Miller, & P. G. Polson (Eds.), *Method and tactics in cognitive science* (pp. 51–83). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Custers, E. J. F. M. (1995). *The development and function of illness scripts: Studies on the structure of medical diagnostic knowledge* (Doctoral dissertation, Maastricht University). Maastricht, The Netherlands: Datawyse Universitaire Pers.

Custers, E. J. F. M., Boshuizen, H. P. A., & Schmidt, H. G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & Cognition, 24,* 384–399.

de Groot, A. D. (1946). *Het denken van den schaker* [The chess player's thinking process]. The Hague, The Netherlands: North-Holland.

de Groot, A. D. (1965). *Thought and choice in chess.* The Hague, The Netherlands: Mouton.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. de Volder (Eds.), *Tutorials in problem-based learning: New directions in training for the health professions* (pp. 128–142). Assen, The Netherlands: Van Gorcum.

Feltovich, P. J., Coulson, R. L., Spiro, R. J., & Dawson-Saunders, B. K. (1992). Knowledge application and transfer for complex tasks in ill-structured domains: Implications for instruction and testing in biomedicine. In D. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 213–244). Berlin, Germany: Springer-Verlag.

Hobus, P. P. M., Boshuizen, H. P. A., & Schmidt, H. G. (1989, July). *Mental representation of prototypical patients: Expert–novice differences.* Paper presented at the First European Congress of Psychology, Amsterdam, The Netherlands.

Hobus, P. P. M., Boshuizen, H. P. A., & Schmidt, H. G. (1990, April). *Expert–novice differences in the role of contextual factors in early medical diagnosis.* Paper presented at the annual meeting of the American Educational Research Association, Boston.

Hobus, P. P. M., Hofstra, M. L., Boshuizen, H. P. A., & Schmidt, H. G. (1988). De context van de klacht als diagnosticum [The context of the complaint as a diagnostic tool]. *Huisarts en Wetenschap, 31,* 261–267.

Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert–novice differences. *Medical Education, 21,* 471–476.

Johnson, P. E., Durán, A. S., Hassebrock, F., Möller, J., & Prietula, M. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5,* 235–283.

Lesgold, A. M., Rubinson, H., Feltovich, P. J., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education, 15,* 315–322.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychologica l Review, 65,* 151–166.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, G. R. (1988). Problem-solving skills, solving problems and problem-based learning. *Medical Education, 22,* 279–286.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, Cognition, 15,* 1166–1174.

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 53–112). Cambridge, MA: MIT Press.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10,* 91–116.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition, 18,* 394–406.

Schaafstal, A. M. (1991). *Diagnostic skill in process operation: A comparison between experts and novices* (Doctoral dissertation, Rijksuniversiteit Groningen). Soesterberg, The Netherlands: Author.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review, 5,* 205–221

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21,* 338–351.

Schmidt, H. G., Boshuizen, H. P. A., & Norman, G. R. (1992). Reflections on the nature of expertise in medicine. In E. Keravnou (Ed.), *Deep models for medical knowledge engineering* (pp. 231–248). Amsterdam: Elsevier.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65,* 611–621.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Vaatstra, R. (1996). *Expertise in accountancy: Empirisch onderzoek naar de kennisontwikkeling van student tot ervaren accountant* [Expertise in auditing: Emperical studies on the development of knowledge from student to expert auditor] (Doctoral dissertation, Maastricht University). Maastricht, The Netherlands: Datawyse Universitaire Pers.

Van Rossum, H. J. M., Bender, W., & Meinders, A. E. (1991). De invloed van biografische details in casuïstische mededelingen op het diagnostisch oordeel [The influence of biographical details in case descriptions on diagnostic judgment]. *Nederlands Tijdschrift voor de Geneeskunde, 135,* 802–805.

Wagenaar, W. A. (1987). Logisch voorwaarts en intelligent achterwaarts: Modellen voor het stellen van diagnoses in de hulpverlening [Logically forward and intelligently backward reasoning: Diagnostic models for psychological assistance]. In J. D. van der Ploeg & P. M. van den Bergh (Eds.), *Besluitvorming en jeugdhulpverlening* (pp. 9–18). Leuven, Belgium: Acco.

Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1151–1164.

APPENDIX A
Examples of Protocols (the Target Disease Was Herpes Zoster)

Enabling Conditions Centered

Eh, also an elderly man or woman, an elderly man or woman, also in a somewhat—what bad shape, a bit emaciated, perhaps; probably, but not necessarily, known with underlying afflictions,— eh— who gets a rash, and pain, in a certain segment ...

Fault Centered

It's the chicken-pox virus, you've had chicken-pox in your youth, and the virus settles down in your nerves, —and you get vesicles on your skin, and these are very contagious, because the fluid contains the virus, and you'll never get rid of it—; and it can cause an infection of the eyes, and that might be very dangerous ...

Consequences Centered

Pain, unilateral, in the shape of a belt, and it coincides with a skin nerve, and eh— not always in the shape of a belt—, it might be, but not when it's on your head, or on your legs— ; anyway, it coincides with a skin nerve, —redness, a swelling, an elevated swelling, with vesicles, and eh— hypersensitivity, and often a general feeling of weak health— a classical picture.

CDFT Centered

In the elderly often a problem, because they visit the physician in a stage when it's already too late to treat it properly— nowadays with Zovirax-pills; eh— very nasty disease, many problems with so-called post-herpetical pains, these may continue to trouble the patient for years and are very difficult to combat; eh— in the beginning stages often difficult to diagnose— in case of vague, unexplainable pains I'm always on guard and ask them to return the next day.

APPENDIX B

Results of the Patient Frequency Inventory (Diseases Ordered According to Mean Frequency Over the Three Expertise Levels)

	Mean No. of Patients Seen by:				No. of Expert Physicians[a] Indicating:	
Disease	*6th-Year Students*	*Interns*	*Family Physician*	*M*	*Never Seen*	*20+ Times*
Meningitis as complication of mumps	0.00	0.00	2.32	0.82	11	1
Epidural hematoma	1.52	0.78	1.57	1.28	6	0
Rupture of aortic aneurysm	1.77	2.85	4.34	2.99	5	0
Cancer of the head of the pancreas	1.96	2.63	6.52	3.69	2	1
Digitalis intoxication	1.18	3.04	7.16	3.78	3	2
Urosepsis	2.91	2.09	6.61	3.84	2	2
Pediculosis pubis	0.23	3.30	11.59	5.02	0	4
Hepatitis A	0.98	2.04	13.05	5.31	0	7
Dermatitis peri-oralis	2.71	6.07	14.50	7.73	1	11
Metastases in lungs (sigmoid cancer)	7.27	7.83	13.23	9.42	0	10
Monilia of the mouth	2.59	8.15	18.05	9.58	0	17
Stomatitis aftosa	2.43	11.93	18.73	11.03	0	18
Herpes zoster	5.00	9.96	18.50	11.20	0	15
Kidney stones colic	2.09	13.17	18.77	11.37	0	16
Perforated otitis media	2.41	12.63	19.50	11.53	0	19
Pre-infarct syndrome	6.41	10.50	17.75	11.54	0	16
Dyspepsia on a nervous basis	6.18	16.39	19.27	13.99	0	19
Vaginal candidiosis	6.32	18.57	20+	15.02	0	22
Nervous abdominal pain	10.09	18.70	20+	16.30	0	22
Secondary enuresis nocturna[b]	—	—	—	—	—	—

[a]$n = 22$. [b]No relevant data available.