Quality Improvement through a Reminder System in an Outpatient Department for Thyroid Disease

Martin Reitstätter[1], Heinrich Vierhapper[2], Andrea Rappelsberger[1], Klaus-Peter Adlassnig[1]
[1]Section on Medical Expert and Knowledge-Based Systems, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria
{martin.reitstaetter, andrea.rappelsberger, klaus-peter.adlassnig}@meduniwien.ac.at
[2]Division of Endocrinology and Metabolism, Department of Medicine III, Medical University of Vienna, Währinger Gürtel 18–20, A-1090 Vienna, Austria
heinrich.vierhapper@meduniwien.ac.at

Abstract

We discuss a reminder system that automatically interprets the results of routine biochemical tests for thyroid hormones. We measured serum levels of thyroid stimulating hormone (TSH), total thyroxin (TT4), triiodothyronine (TT3), thyroxin-binding globulin (TBG), and thyrotropin-releasing hormone (TRH)-stimulated TSH (sTSH [1,2]). The developed knowledge base was used to compare and rate these results with a defined set of diagnoses. Various forms of altered thyroid and/or pituitary function [3–8] were considered. For a given combination of test results, the reminder system classified the entered diagnoses as "obligatory", "possible", or "excluded". The knowledge base is now part of a medical documentation system at the outpatient thyroid department of the Division of Endocrinology and Metabolism of the Vienna General Hospital. We report an evaluation performed after six months of operation. Although further improvement will be required, the knowledge-based reminder system in its present form is capable of improving the quality of manually entered diagnoses.

1. Introduction

There is reason to believe that the development of diagnostic and interpretative systems for the laboratory will lead to clinically useful computer applications that can be routinely implemented in laboratories [9]. The final purpose of these systems is to improve patient care.

Initially the objective of our investigation was to reduce unjustified referrals to hospitals due to misinterpretation of the laboratory test results obtained by general practitioners; the latter should have the possibility to consult an expert system concerning the interpretation of thyroid function tests. In a first step, this knowledge-based system was tested for its reliability in a thyroid outpatient clinic where diagnoses are manually entered into a documentation system based on both, laboratory data and clinical criteria. These diagnoses will then be verified by a reminder system; the existing laboratory test results will be re-assessed.

The present report describes the development of this reminder system and presents the results of its operation as recorded in log files. Log file data after six months of operation are evaluated.

2. Methods

Our aim was to develop a reminder system operating in the background and monitoring the input of diagnoses. In the event of conflicts the reminder appears on the screen and the physician is called upon to review his input. To this end it was necessary to create an adequate knowledge base [10, 11]. The system was set up to identify a set of 13 diagnoses (Table 1). Laboratory parameters used to establish the diagnoses are serum concentrations of thyroid stimulating hormone (TSH), total thyroxin (TT4), triiodothyronine (TT3) and thyroxin-binding globulin (TBG), as well as sTSH (i.e., serum concentration of TSH determined 20 minutes after intravenous administration of 400 µg thyrotropin-releasing hormone (TRH) [1,2]). Combining these two variables (diagnoses and laboratory parameters), we obtained a matrix in which each element contains the rating of each diagnosis with respect to a specific combination of qualified (interpreted) laboratory test results.

The current documentation system at the outpatient thyroid department of the Vienna General Hospital uses a database table for all diagnoses that can be entered. A maximum of three diagnoses may be stored for each

Table 1: A small section of the knowledge base.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	...	1023
TT4	○	⊥	↑	↓	○	⊥	↑	↓	○	⊥	↑	↓	○	⊥	↑	↓	○	⊥	↑	↓	○	⊥	↑	↓	...	↓
TT3	○	○	○	○	⊥	⊥	⊥	⊥	↑	↑	↑	↑	↓	↓	↓	↓	○	○	○	○	⊥	⊥	⊥	⊥	...	↓
TSH	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	↑	↑	↑	↑	↑	↑	↑	↑	...	↓↓
TBG	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	...	↓
sTSH	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	...	↓
hyperthyroidism	3	3	3	3	3	2	2	3	3	3	2	3	3	3	3	3	3	3	2	3	3	3	3	3	...	2
hyperthyroidism (subclinical)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	...	3
hyperthyroidism (subclinical, suspected)	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	...	3
hyperthyroidism (secondary)	2	2	2	2	3	2	2	2	2	2	1	2	3	3	2	3	2	2	2	2	3	2	2	2	...	3
hyperthyroidism (suspected)	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	...	2
hyperthyroidism (T3 only)	3	3	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	...	3
hypothyroidism	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	2	2	1	2	2	2	1	...	3
hypothyroidism (subclinical)	2	2	3	2	2	2	3	1	3	3	3	3	2	2	3	2	1	1	2	3	1	1	1	2	...	3
hypothyroidism (secondary)	2	2	3	2	2	2	2	2	3	3	3	3	2	2	3	2	3	3	3	3	3	3	3	3	...	2
thyroiditis	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	...	2
deficiency of TBG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	...	1
abundance of TBG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	...	3
pregnancy / exogenous estrogens	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	...	2

patient. From this table of diagnoses, we selected those which were important for the reminder system. Some clinically defined forms of hyper- and hypofunction of the thyroid gland and the pituitary appeared too specific and were therefore subsumed into classes; (e.g., the diagnoses "*iatrogenic hyperthyroidism*", "*hyperthyroidism induced by contrast medium*", and "*hyperthyroidism*" were combined into the class "*hyperthyroidism*" because currently the system only assesses the presence of hyperthyroidism, but not its cause). This resulted in six defined forms of hyperthyroidism, three forms of hypothyroidism, two diagnoses related to TBG, and the diagnosis of thyroiditis. The diagnosis "euthyroidism" was omitted because it is obligatory when no hyper- or hypothyroidism is present.

The diagnoses were evaluated with regard to the parameters TSH, TT4, TT3, TBG, and TRH-stimulated TSH (sTSH). Each of these parameters may yield one of four (different) possible qualitative results. The qualitative test results of TT4, TT3, TBG, and sTSH are *not tested* (○), *normal* (⊥), *increased* (↑), or *decreased* (↓). Similarly, the qualitative test results of TSH may be *normal* (⊥), *increased* (↑), *decreased* (↓), or *highly decreased* (↓↓). Thus, there may be 1,024 combinations of test results. Since there are logical combinations which do not naturally occur, a virtual class named "*inconclusive or erroneous data*" was created. Drawing up the matrix with the 13 diagnoses and the 1,024 test result combinations, we had to rate 13,312 fields.

Table 1 shows a small part of the knowledge base with the respective rating in each table cell. A diagnosis may be obligatory (1, green), possible (2, blue), or excluded (3, red) with regard to a given combination of laboratory test results.

The knowledge base was verified retrospectively against the laboratory test results and the corresponding diagnoses stored in the database of the outpatient thyroid department. In the event of differing results, the case in question was copied into a conflict list with the appropriate patient ID and the laboratory test results. This conflict list was then supplemented by the diagnoses which—with reference to the underlying new knowledge base—occurred, but were to be excluded or were missing despite being mandatory. The medical expert then had to establish the origin of the respective conflict—an erroneous knowledge base, an incorrect inference process, or a wrong diagnosis. Consequently, either a revision of the knowledge base and/or the inference process followed, or the wrong diagnosis was marked. The iterative procedure of verification was restarted and repeated until no further conflicts occurred.

The results after five cycles of review were as follows: of 17,053 total datasets we considered 15,839. The remaining 1,214 datasets were ignored because they did not contain the appropriate laboratory data (n=328). Evaluating the knowledge base using these 15,839 entries, the system revealed 628 conflicts. This amounts to about 4% of the evaluated datasets. 548 diagnoses were mandatory but missing and 44 were interpreted as false; 19 cases included both wrong and missing diagnoses, and in 17 cases the laboratory test results indicated "*inconclusive or erroneous data*". Inconclusive data were obtained when the test results fell slightly outside the anticipated ranges; logically inconsistent test results indicated erroneous data.

For the future elimination of the error "*inconclusive and erroneous data*", an extension of the knowledge base will be necessary. For instance, if the serum concentration of TBG is increased the diagnosis *abundance of TBG* is to be entered obligatorily, provided that there is space left in the three fields available for the input of diagnoses. A review showed that, in the past, in 100 cases of increased TBG this diagnosis had not been entered despite the availability of space. In the future this will be mandatory. Furthermore, classification of the test results into normal (⊥), increased (↑), and decreased (↓) was not optimal for clinical purposes, especially when the values were at the upper or the lower end of the normal range. For example, a test result within the normal range but close to its lower end might have been clinically interpreted as 'decreased' because the physician made the diagnosis on the basis of all available data for the patient. This problem remains unresolved at the present time.

After the final evaluation the knowledge base was integrated into the documentation system of the outpatient thyroid department. The applied inference process is the same as that used in Hepaxpert, a knowledge-based system for the interpretation of hepatitis serology test results. A description can be found in [12]. If obligatory diagnoses are missing, a reminder in a message box will inform the physician of these (see Figure 1). He then may either adjust his input or override the system.

The reminder system is now in operation since 22 April 2005 and each step is recorded in a log file.

3. Results

The implemented knowledge-based system is currently being tested at the outpatient thyroid department of the Vienna General Hospital as part of its documentation system. It is used to verify the diagnoses entered by physicians. A reminder appears in case of contradictions or missing obligatory diagnoses with respect to the given laboratory test results.

A preliminary evaluation of the log file was performed after 173 days of operation. During this time, 1,630 diagnoses for 1,180 patients had been entered. The reminder system rejected the entered diagnoses in 9 cases because the diagnoses were ruled out by the obtained laboratory test results. In 5 of the 9 cases the system had initially been overruled by the physician, but this was eventually proven wrong in 4 cases. In 51 cases the system expected obligatory diagnoses that were missing. In 37 of these cases the decision to overrule the system by omitting these diagnoses was maintained on the basis of clinical criteria, whereas in 14 cases the additional diagnoses were appended. During the data selection process [13] we also found 3 data inconsistencies. Once it was caused by first clicking the FIX button of the reminder box, followed by a click of the SAVE button without correcting the diagnosis in the meantime. Finally, the OVERRIDE button of the reminder was clicked. In another instance the reminder system was obviously shut down directly after it appeared, and the input procedure remained incomplete. This happened twice. In summary, during 173 days of operation 18 diagnoses were corrected by the assistance of the knowledge-based reminder system alone.

Furthermore, at the beginning of our study we realized that the classification of diagnoses into obligatory, possible, and excluded was too strict. Because the knowledge acquisition process took a long time and

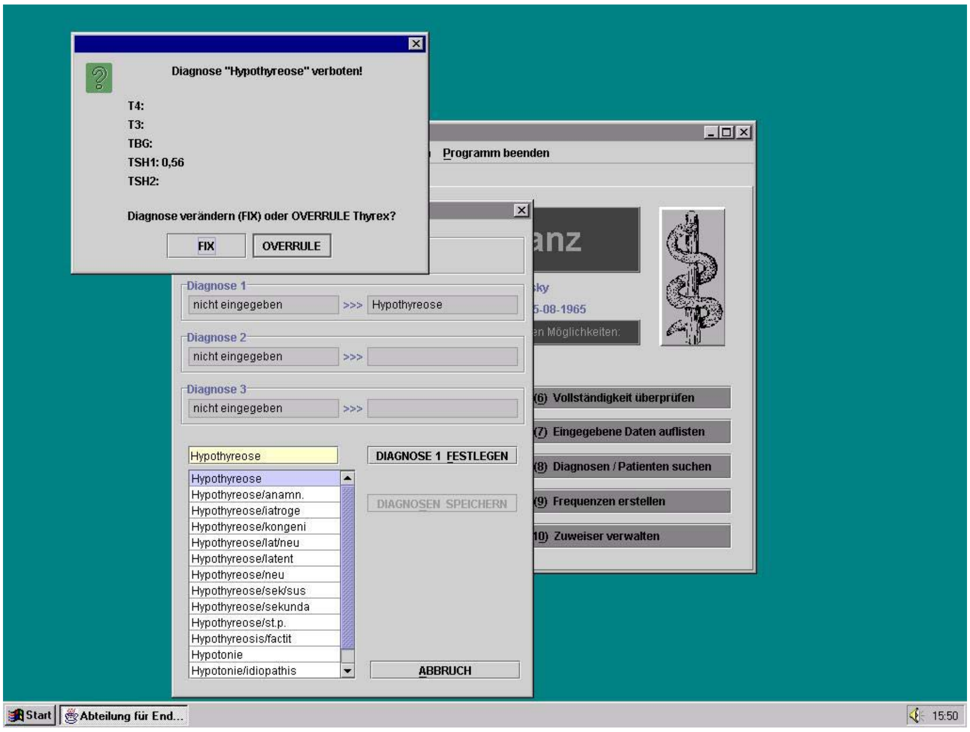

Figure 1: Documentation system with reminder message box.

required several meetings, the classification was modified over time. The medical expert and the knowledge engineer became increasingly familiar with the system from one meeting to the next. For example, a diagnosis that was considered *possible* at the first meeting was rated *possibly obligatory* at one of the subsequent meetings. The system would no longer accept very rare diagnoses as “possible” without issuing a reminder, given their very low pre-test probability. Overall this resulted in a refinement of the interpretation of laboratory test results, progressively reducing the number of *possible* diagnoses. The knowledge engineer should strive to keep this process in a straightforward time frame in order to avoid inconsistent interpretations over time.

Based on this experience, we started to rebuild the entire system from scratch using a finer weighting system for the diagnoses. The categories are now named “obligatory”, “likely”, “possible”, “rather impossible”, and “excluded”. The reminder system issues an alert every time a diagnosis is not absolutely proven or an obligatory diagnosis is missing. For testing purposes, the sensitivity of the reminder can be adjusted (e.g., accepting likely diagnoses without generating an additional message).

4. Conclusions

The developed knowledge base is essentially a table of ratings. It can be easily translated into plain text for uses other than the local thyroid documentation system.

For application in a pure knowledge-based interpretive system that is able to interpret the entered laboratory test results and provide a written textual interpretation, the knowledge base was translated into Arden Syntax [13, 14]. The resulting package of Arden Medical Logic Modules can be integrated into any medical information system that is equipped with an Arden interpreter/compiler.

For the purpose of testing and possibly as a future decision support tool for medical practitioners, a web version was designed. Once quantitative or qualitative laboratory test results have been entered, this version generates the appropriate diagnosis.

Improvements in the knowledge base and the inference process are being considered. With regard to the laboratory parameters, there will be a change from TT4 to fT4, i.e., the free portion of T4 which is not bound to TBG. The current laboratory analyzers support fT4 rather than TT4; therefore currently both fT4 and TT4 are accepted.

Last but not least, classifications such as *normal*, *increased*, and *decreased* can be easily fuzzified [15, 16]. This might be the logical next step. However, we first have to evaluate the most recent version of the system.

Acknowledgement

The authors are indebted to P. Grösser, MSc., for his programming work during this project.

References

[1] L. Thomas. *Labor und Diagnose – Indikation und Bewertung von Laborbefunden für die medizinische Diagnostik.* TH-Books Verlagsgesellschaft mbH, Frankfurt/Main, 5th edition, 1027–1044, 2000.

[2] F.-W. Tiller and B. Stein. *Das klinische Labor*. ecomed medizin Verlag, Landsberg, Germany, 373–378, 404, 437, 2003.

[3] A. Waldeyer and A. Mayet. *Anatomie des Menschen 2*. deGruyter Verlag, Berlin, 15th edition, 151–155, 1986.

[4] G. Herold. *Innere Medizin – Eine vorlesungsorientierte Darstellung*. Gerd Herold, Köln, 593, 1999.

[5] H.-G. Schwarzacher and W. Schnedl. *Histologie*. Facultas Universitätsverlag, Wien, 4th edition, 389–391, 1998.

[6] G. Thews and P. Vaupel. *Vegetative Physiologie*. Springer Verlag, Berlin, 2nd edition, 420–421, 1990.

[7] R. Schmidt and G. Thews. *Physiologie des Menschen*. Springer Verlag, Berlin, 26th edition, 390–392, 1995.

[8] U.-N. Riede and H.-E. Schäfer. *Allgemeine und spezielle Pathologie*. Georg Thieme Verlag, Stuttgart, 4th edition, 1003, 1995.

[9] D-P. Conelly. Embedding Expert Systems in Laboratory Information Systems. *American Journal of Clinical Pathology* Suppl. 1:7–14, 1990.

[10] M. Schnabel. *Expertensysteme in der Medizin*, Gustav Fischer Verlag, Stuttgart, 182–194, 1996.

[11] F. Puppe. *Einführung in Expertensysteme*. Springer Verlag, Berlin, 2nd edition, 114–131, 1991.

[12] K.-P. Adlassnig and W. Horak. Development and Retrospective Evaluation of Hepaxpert-I: a Routinely-Used Expert System for Interpretive Analysis of Hepatitis A and B Serologic Findings. *Artificial Intelligence in Medicine* 7:1–24, 1995.

[13] G. Hripcsak. The Arden Syntax for Medical Logic Modules: Introduction. *Computer in Biology and Medicine* 24(5):329–330, 1994.

[14] G. Hripcsak. Writing Arden Syntax for Medical Logic Modules. *Computer in Biology and Medicine* 24(5):331–363, 1994.

[15] L.A. Zadeh. Fuzzy Sets. *Information and Control* 8:338–353, 1965.

[16] K. Boegl, K.-P. Adlassnig, Y. Hayashi, T.E. Rothenfluh, and H. Leitich. Knowledge Acquisition in the Fuzzy Knowledge Representation Framework of a Medical Consultation System. *Artificial Intelligence in Medicine* 30, 1–26, 2004.