

Aus der II. Medizinischen Klinik für Innere Medizin  
DRK-Kliniken Berlin Köpenick  
Akademisches Lehrkrankenhaus der Medizinischen Fakultät Charité –  
Universitätsmedizin Berlin

DISSERTATION

**Eignung eines „intelligenten“  
Dokumentationssystems für die Sonographie  
zur Qualitätskontrolle, zur Ausbildung von  
Untersuchern und zur Anlage einer  
hochwertigen Falldatenbank**

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

**Matthias Hüttig**

aus Görlitz

Gutachter: 1. Prof. Dr. med. H.-P. Buscher

2. Prof. Dr. med. S. Schewe

3. Prof. Dr. med. R. Somasundaram

Datum der Promotion: 14. Juni 2009

# Eignung eines „intelligenten“ Dokumentationssystems für die Sonographie zur Qualitätskontrolle, zur Ausbildung von Untersuchern und zur Anlage einer hochwertigen Falldatenbank

## Inhaltsverzeichnis

<b>1. Einleitung .....</b>	<b>4</b>
<b>2. Methodik .....</b>	<b>6</b>
<b>2.1. Beschreibung des Programms .....</b>	<b>6</b>
<b>2.2. Begleitende Evaluation während der Einführung des Programms .....</b>	<b>14</b>
<b>2.3. Evaluation.....</b>	<b>17</b>
2.3.1. Eingabedauer .....	17
2.3.2. Akzeptanz.....	18
2.3.3. Subjektive Einschätzung des Ausbildungseffekts bei den Untersuchern.....	19
2.3.4. Qualität der Befundberichte .....	19
2.3.5. Diagnostische Genauigkeit.....	21
2.3.6. Konkordanz von Befund und Beurteilung .....	22
<b>2.4. Statistische Auswertungen .....</b>	<b>25</b>
2.4.1. Inter-Untersucher-Variation .....	25
2.4.2. Subgruppen-Analyse zur Wissensentdeckung .....	26
<b>3. Ergebnisse .....</b>	<b>27</b>
<b>3.1. Evaluation.....</b>	<b>27</b>
3.1.1. Eingabedauer .....	27
3.1.2. Akzeptanz.....	28
3.1.3. Subjektive Einschätzung des Ausbildungseffekts bei den Untersuchern.....	33
3.1.4. Qualität der Befundberichte .....	34
3.1.5. Diagnostische Genauigkeit.....	39
3.1.6. Konkordanz von Befund und Beurteilung .....	41
<b>3.2. Statistische Auswertungen .....</b>	<b>43</b>
3.2.1. Inter-Untersucher-Variation .....	43
3.2.2. Subgruppen-Analyse zur Wissensentdeckung .....	49
<b>4. Diskussion.....</b>	<b>51</b>
<b>5. Zusammenfassung .....</b>	<b>58</b>
<b>6. Literatur .....</b>	<b>59</b>

# 1. Einleitung

Befunde technischer Untersuchungen müssen sorgfältig und nachvollziehbar dokumentiert werden, um aus ihnen die angemessenen diagnostischen, therapeutischen und prognostischen Schlüsse ziehen zu können. Eine Dokumentation im Freitextformat ist wegen der Möglichkeit einer raschen und fallorientierten Zentrierung auf das Wesentliche für viele Ärzte die beliebteste Form und ist somit weit verbreitet. Sie hat jedoch Nachteile, die sich besonders bei Ärzten, die sich noch in Ausbildung befinden, gravierend bemerkbar machen können. Dazu gehören eine fehlende Standardisierung der verwendeten Nomenklatur und Lücken in der Dokumentation. Formularbasierte Dokumentationen sollen diese Schwäche überwinden helfen, sind aber nicht flexibel genug, um die vielfachen Erscheinungen einer Krankheit adäquat erfassen zu können. Flexibilität versprechen dagegen wissensbasierte computergestützte Dokumentationssysteme, die je nach vorliegender Symptomatik oder Befundkonstellation die notwendigen Eingaben abfordern. Besonders gut sollten dazu solche Systeme geeignet sein, die aus den eingegebenen Symptomen, Befunden und Untersuchungsergebnissen selbst diagnostische Schlüsse ziehen können und aus ihnen die zusätzlich weiter zu dokumentierenden Informationen herleiten<sup>1</sup>. Sie sind dazu geeignet, dem Untersucher alle für diagnostische und therapeutische Schlussfolgerungen relevanten Fragen zu Befunden vorzulegen und ihn damit zu einer vollständigen Dokumentation und nachvollziehbaren Beurteilung anzuregen.

Diese wissensbasierten Programme werden vielfach als „Expertensysteme“ bezeichnet, da ihre zugrunde liegenden Algorithmen dem Wissen medizinischer Experten entspringen<sup>2</sup>. Sie fokussieren traditionellerweise komplexe diagnostische und therapeutische Empfehlungen<sup>3,4</sup>. Expertensysteme wurden sowohl für die gesamte Innere Medizin<sup>5</sup> als auch für Teilgebiete<sup>6,7,8,9,10,11</sup> entwickelt. Einige werden inzwischen kommerziell angeboten.

Neuere Berichte<sup>12</sup> zeigen, dass die diagnostischen und therapeutischen Empfehlungen derartiger Expertensysteme seitens der Ärzte nicht als primäres Bedürfnis gesehen werden. Sie ersetzen die Expertise erfahrener Ärzte nicht. Andere ihnen eigene Funktionen dagegen, wie die Unterstützung einer hohen Befundqualität mit einer Reduzierung von Dokumentationslücken, eine Erinnerungsfunktion und die Möglichkeit statistischer Analysen, könnten im klinischen Alltag von höherer Bedeutung sein.

Ein weiteres Potential wissensbasierter Computersysteme besteht in der Ausbildung junger Ärzte. Ein fehlerhafter Befundbericht kann nicht nur in der Untersuchung selbst, sondern auch in der diagnostischen Beurteilung, der Dokumentation oder Berichterstellung begründet sein. Eine objektive Beurteilung der Befundqualität ist im klinischen Alltag nur stichprobenartig möglich und selbst fehleranfällig. Wenn es gelingt, die internen Schlussfolgerungsalgorithmen des Programms jederzeit im Eingabedialog abrufbar und dem Arzt zur Einsicht verfügbar darzustellen, bietet es für ihn die Möglichkeit, nicht nur Lücken und Inkonsistenzen in seinem Befund zu erkennen, sondern auch die eigenen Schlussfolgerungen aus ihnen zu überprüfen. Eine weitere Programmeigenschaft, die in der Ausbildung genutzt werden kann, wäre die Abrufbarkeit von Erklärungen, beispielsweise zu Symptomen, Diagnosen oder diagnostischen und therapeutischen Schlussfolgerungen.

Ein Beispiel für eine diagnostische Methode, die außerordentlich untersucherabhängig ist und für eine verlässliche Befundung eine hohe Expertise des Untersuchers erfordert, ist die Sonographie. Für junge Untersucher sollte über längere Zeit ein Tutor zur Verfügung stehen. Die Sonographie scheint somit hervorragend geeignet als Beispiel einer diagnostischen Methode, bei der zu erwarten sein müsste, dass ein computergestütztes Dokumentationssystem eine hohe Befundqualität und eine Unterstützung bei der Ausbildung junger Untersucher gewährleistet.

Vor diesem Hintergrund wurde ein bereits existierendes und als „HepatoConsult“ in der Anwendung evaluiertes Expertensystem<sup>13</sup> auf der Basis des Diagnostik-Shell-Baukastens D3web ([<http://d3web.sourceforge.net>] Institut für Informatik der Universität Würzburg<sup>14,15</sup>), welches die Eingabe des Expertenwissens über eine graphische Benutzeroberfläche erlaubt, für die Sonographie unter der Bezeichnung „SonoConsult“ weiterentwickelt. Es ist auf die falladäquate vollständige Befunddokumentation und zudem auf eine Unterstützung bei der Ausbildung junger Untersucher ausgelegt. Es lässt die Algorithmen seiner diagnostischen Schlussfolgerungen einsehen und bietet in einer Erklärkomponente Informationen zu Symptomen, Befunden, Normwerten und Diagnosen. In der vorliegenden Arbeit werden die Akzeptanz des Programms bei den Ärzten, sein Effekt auf die Qualität der Untersuchungsbefunde und zudem die Brauchbarkeit der programmgenerierten Falldatenbank für die Erkennung medizinischer Zusammenhänge untersucht.

## 2. Methodik

### 2.1. Beschreibung des Programms

SonoConsult stellt ein PC-Programm zur Erstellung von Sonographiebefunden dar. In einzelnen Frageklassen werden die Organsysteme der Abdominalsonographie (Leber, portales Gefäßsystem, Gallenblase, Milz, Pankreas, Nieren, Nebennieren, Magen, Darm, Lymphknoten, große Gefäße, Prostata, Harnblase) systematisch dokumentiert. Weitere Eingabemöglichkeiten bestehen für extraperitoneale Lymphknoten, Schilddrüse und Thorax und duplexsonographische Befunde. Das Programm umfasst neben der Befunddokumentation eine diagnostische Komponente, in der aus den Eingaben unmittelbar auf die in Frage kommenden Diagnosen geschlossen wird. Zudem ist eine Erklärungskomponente integriert. Sie ermöglicht einerseits, die diagnostischen Schlussfolgerungen und die Gründe für die Vorlage bestimmter Fragebögen oder Nachfragen einzusehen, und stellt andererseits Erklärungen zu Befunden und Diagnosen zur Verfügung. Beides kann für Lernzwecke genutzt werden. Der generierte Fallspeicher kann für statistische Analysen und Wissensentdeckung verwendet werden.

**Die Symptomerfassung** erfolgt über Formulare mit einer oder mehreren Fragen nach Befunden, die der Untersucher aus einer vorgegebenen Menge an Alternativen auswählt. Die Formulare werden je nach Fragestellung und bereits getätigter Eingabe von Befunden so generiert, dass gezielte Nachfragen zu weiteren Befunden präsentiert werden, die zur Bestätigung oder Falsifizierung einer verdächtigten Diagnose erforderlich sind.

Ziel der Symptomerfassung ist zum einen die Erstellung eines strukturierten Befundberichtes, der über eine Schablone in ein lesbares Format überführt werden kann und anschließend im Krankenhaus-Informationssystem (KIS) zur Verfügung steht. Zum anderen entsteht ein Fallspeicher für statistische Analysen, die zu neuem Wissen führen können.

Die Symptomerfassung kann in zwei Modi erfolgen, dem Standard- und dem Expertenmodus. Der Standardmodus bedeutet eine detaillierte und systematische Symptomerfassung in Fragebögen für alle Organsysteme. Die notwendigen Fragebögen werden vom Programm durch die zu bearbeitende Fragestellung der Untersuchung, die zu Beginn zu dokumentieren ist, und im Verlauf durch die gewählten Antwortalternativen selbst bestimmt. Der Expertenmodus erfordert die Expertise des Untersuchers. Er dokumentiert primär die von ihm favorisierten Diagnosen,

anschließend präsentiert das Programm die notwendigen Eingabefelder zu deren Verifizierung. Der Expertenmodus ist deutlich weniger zeitaufwendig als der detaillierte Standardmodus. Die einzelnen Modi stellen letztlich Kompromisse zwischen einer Befunderstellung in kurzer Zeit und einer detaillierten und somit qualitativ hochwertigen Dokumentation dar.

Die Abbildung 1 zeigt einen Screenshot der Benutzeroberfläche im Standardmodus.

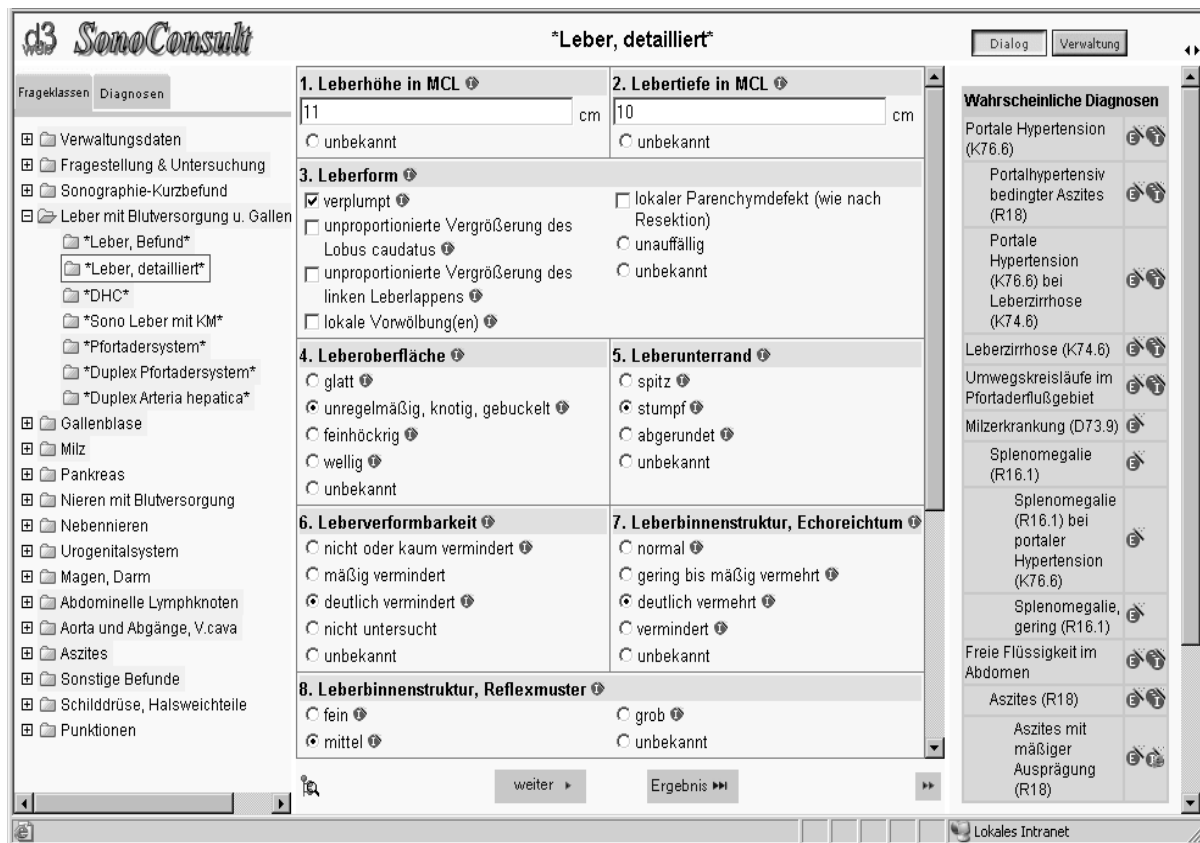


Abbildung 1: Screenshot eines Teils eines Fragebogens in SonoConsult. Die linke Leiste zeigt einen teilweise geöffneten Hierarchiebaum der Symptomklassen, die rechte Leiste die bis zum aktuellen Stand der Eingabe vom System als möglich erachteten Diagnosen.

In der Abbildung ist zentral ein Teil der Symptomerfassung der Frageklasse „Leber“ zu sehen. Der Navigationsbaum links erlaubt dem Untersucher, nach eigenem Ermessen unabhängig von der Abfolge der durch SonoConsult präsentierten Formulare in andere Frageklassen zu wechseln. Die rechte Bildleiste zeigt kontinuierlich die vom Programm bisher ermittelten Diagnosen auf Basis der Symptomerfassung. SonoConsult umfasst in der aktuellen Version 430 Fragen, von denen im Standardmodus etwa 60 beantwortet werden müssen. Neben Fragen mit vorgegebenen Antwortalternativen in Einfach- oder Mehrfachauswahl existieren Freitextfelder zur

Beschreibung sehr seltener Befunde und von Erklärungen zu Befunden. Die Terminologie ist deskriptiv und folgt der üblichen Nomenklatur in Standardwerken des Fachgebietes der Sonographie. Der Zeitaufwand für die Dokumentation einer Untersuchung ist vergleichbar mit herkömmlichen Methoden (in der Regel ein Textprogramm mit eventuell hinterlegten Textbausteinen für häufige Befunde oder Formulierungen), wie in der Evaluation des Programms gezeigt wird. Basierend auf den Antworten in den präsentierten Frageklassen wird über eine Schablone ein Freitextbefund generiert. Während ein Großteil der erfassten Symptome unmittelbar in den Befundbericht überführt werden kann, verlangt die Zusammenfassung von inhaltlich verknüpften Symptomen aus verschiedenen Frageklassen eine logische Verknüpfung.



Ein Beispiel eines Befundberichtes zeigt Abbildung 2.

<b>Sonographie</b>	
<b>Name, Vorname:</b>	<b>Mustermann, Manuel, 01.10.40</b>
<b>Fragestellung:</b>	Oberbauch-Screening; Leberzirrhose
<b><u>Befund</u></b>	vom 17.11.04; gute Untersuchungsbedingungen
<b>Leber:</b> Höhe in MCL 11 cm; Tiefe in MCL 10 cm; verplumpt; Oberfläche unregelmäßig, knotig, gebuckelt; Unterrand stumpf; Verformbarkeit deutlich vermindert; Binnenstruktur deutlich echovermehrt; mittleres Reflexmuster; Kalibersprung der Pfortaderäste intrahepatisch; Rarefizierung der Pfortaderäste intrahepatisch	
<b>D. hepatocholedochus:</b> Durchmesser 5 mm; unauffällig	
<b>Gallenblase:</b> unauffällig	
<b>Milz:</b> längs 14 cm, tief 6 cm; Parenchym unauffällig	
<b>Pfortadersystem:</b> Pfortaderdurchmesser 14 mm; keine wesentliche Zunahme des Durchmessers bei Inspiration; Milzvenendurchmesser 12 mm; Hinweis auf wiedereröffnete Nabelvene	
<b>Duplexsonographie:</b> Pfortader Fluß orthograd mit gleichmäßigem Flußprofil, Flußgeschwindigkeit 12 cm/s; Milzvene Flußgeschwindigkeit 12 cm/s; wiedereröffnete Nabelvene	
<b>Flüssigkeit im Abdomen:</b> freie Flüssigkeit im Sinne von Aszites, mäßig ausgeprägt	
<b>Abdominelle Gefäße:</b> <i>Arteria hepatica (duplexsonographisch):</i> nicht durchgeführt	
<b>Vena cava:</b> unauffällig	
<b>Lymphknoten:</b> in beurteilbaren Regionen nicht erkennbar bzw. nicht vergrößert	
<b>Pleuraerguss:</b> beidseits nicht nachweisbar	
<b>Perikarderguss:</b> nicht nachweisbar	
<b><u>Beurteilung:</u></b>	
<b>Leberzirrhose mit portalhypertensiv bedingtem Aszites und mäßiger Splenomegalie. Kontrolle nach Ausschwemmung des Aszites empfohlen.</b>	

Abbildung 2: Exemplarischer Befundbericht, der von SonoConsult aus dem Fall in Abbildung 1 generiert wurde. Bei der Formulierung seiner eigenen Beurteilung werden dem Untersucher die Diagnosen des Programmes mit ICD-10-Code angeboten. Diese können per Mausklick in die eigene Beurteilung übernommen werden. Auf dem Befundbericht erscheinen die SonoConsult-Diagnosen nicht, werden aber für spätere Auswertungen im Fallspeicher abgelegt (hier: Portale Hypertension (K76.6) bei Leberzirrhose (K74.6); portalhypertensiv bedingter Aszites (R18); Splenomegalie (R16.1) bei portaler Hypertension (K76.6)).

**Der Befundbericht** gliedert sich in drei Teile: den patientenbezogenen Basisdatensatz, den Befund und die Beurteilung. Während Basisdaten und Einträge über Untersuchungsbefunde direkt übernommen werden, müssen die Diagnosen vom Untersucher selbst formuliert werden. Die SonoConsult-Diagnosen werden dem Untersucher während der gesamten Eingabe und

während des Verfassens seiner eigenen Beurteilung präsentiert, jedoch nicht automatisch mit in den Befundbericht übernommen, da die Verantwortung für die Interpretation allein beim Untersucher liegen soll. Die SonoConsult-Einträge zur Befunderfassung und die vom Programm generierten Diagnosen werden in einem Fallspeicher abgelegt und stehen für spätere Aufrufe durch das Programm und für statistische Analysen zur Verfügung. Der aus den Daten generierte Befundbericht sowie die freitextlichen Zusätze werden als Textdokument ebenfalls in einem Fallspeicher abgelegt. Die Web-Oberfläche von SonoConsult ermöglicht die Kommunikation mit dem Krankenhausinformationssystem.

**Die Wissensbasis** nutzt medizinische Heuristik als Wissensquelle<sup>16</sup> und folgt den Konstruktionsprinzipien von HepatoConsult<sup>13</sup>. SonoConsult beinhaltet fünf Hauptkomponenten: Symptome (entsprechen den eingegebenen Alternativen einer Frage), Symptomklassen (Fragebögen, die einzelne Fragen zu Komplexen zusammenfassen), Symptominterpretationen (Zwischenebene zur Diagnosefindung), Diagnosen und Regeln.

- **Symptome** bestehen aus einer Eigenschaft (z. B. Milz) und einem Wert (z. B. vergrößert). Im Gang der Befunderfassung entsprechen die Eigenschaft der „Frage“ und der Wert der „Antwort“ des Untersuchers. Eine Eigenschaft kann über eine numerische Größe oder Antwortalternativen beschrieben werden. Letztere haben eine vordefinierte Auswahl (z. B. Echogenität der Leber: vermindert, normal, mäßig vermehrt, stark vermehrt) und werden in one-choice (genau eine Antwortalternative zugelassen) oder multiple-choice (mehrere Alternativen zulässig) unterschieden.
- **Symptomklassen** fassen zusammengehörige Symptome zusammen, wenn sie aufgrund ihres inhaltlichen Zusammenhanges gemeinsam erfragt werden (z. B. Charakterisierung eines Leberherdes). Innerhalb einer Symptomklasse können Regeln definiert werden, die die Reihenfolge von Fragen aufgrund von gewählten Antwortalternativen zuvor erfasster Symptome festlegen.
- **Symptominterpretationen** entstammen eingegebenen Symptomen und werden aus ihnen durch Regeln hergeleitet. Sie haben zwei wesentliche Funktionen: sie können zur Symptomzusammenfassung und als Zwischenebenendiagnose (wie Splenomegalie oder portale Hypertension) dienen.
- **Diagnosen** werden ebenfalls durch Regeln ermittelt, wobei diese auf Symptome, Symptominterpretationen und andere Diagnosen („Kriterien“) zurückgreifen.

- **Regeln** dienen (a) den Verknüpfungen innerhalb der Symptomerhebung und (b) der Herleitung von Diagnosen. Sie bestehen aus Bedingungen, Aktionen und Ausnahmen. In einer Bedingung können Kriterien mit „und“, „und nicht“, „oder“ und „n von m“ miteinander verknüpft sein. Es können Ausnahmen von einer Regel definiert werden. Sie berücksichtigen die Fälle, in denen eine Regel bei Vorliegen definierter Informationen oder Kriterien nicht angewendet werden soll. Die Regelstruktur ist je nach ihrer Aufgabe einfach oder beliebig komplex aufgebaut. Sie gehorchen prinzipiell dem Schema (WENN BEDINGUNG A – DANN AKTION B). Beispiel: WENN (SYMPTOME A UND/ODER B UND NICHT C UND `N VON M` ... – AUSSER X UND/ODER ...) DANN (BELASTE DAS PUNKTEKONTO VON DIAGNOSE D MIT P PUNKTEN) ODER (SETZE DIE SYMPTOMINTERPRETATION AUF ...) ODER (GENERIERE FORMULAR F).

Die vom Programm generierten Diagnosen sind für den Untersucher Hinweise auf Diagnosen, die in Erwägung gezogen werden sollten, und sind kritisch zu übernehmen. Ob sie zutreffen hängt nicht nur von der Kompetenz des Regelwerks, sondern auch von der Vollständigkeit der Befundeingabe und der Güte der Untersuchung selbst ab. Sie ist generell zu überprüfen.

In SonoConsult wird in Analogie zu QMR (Quick medical reference)<sup>17</sup> ein Punktesystem verwendet, welches unter der Maßgabe voneinander unabhängiger Regeln über die Vergabe von Punkten in sechs negativen und sechs positiven Kategorien das Konto einer Diagnose beschickt. Die Bewertung kann als Evidenz bezüglich einer Diagnose interpretiert werden und führt durch Definition von Schwellenwerten zu den semiquantitativ deskriptiven Kategorien „unklar oder ausgeschlossen“, „möglich“ oder „wahrscheinlich“.

**Der Verlauf der Symptomerfassung (Dialog)** ist durch das Regelwerk so strukturiert, dass er der Befunddokumentation eines Untersuchers weitgehend entspricht. Dabei werden (a) die Befunde möglichst vollständig erhoben und (b) die sich aus den ableitbaren Diagnosen ergebenden neuen Aspekte bei der Befunderhebung berücksichtigt. Das diagnostische Regelwerk von SonoConsult folgert aus den dokumentierten Daten kontinuierlich auf mögliche Diagnosen und stellt somit ein dynamisches Werkzeug dar. Die Diagnosefindung in SonoConsult erfolgt nach der „hypothesis-and-test“- und der „establish-refine“-Strategie (Abb. 3).

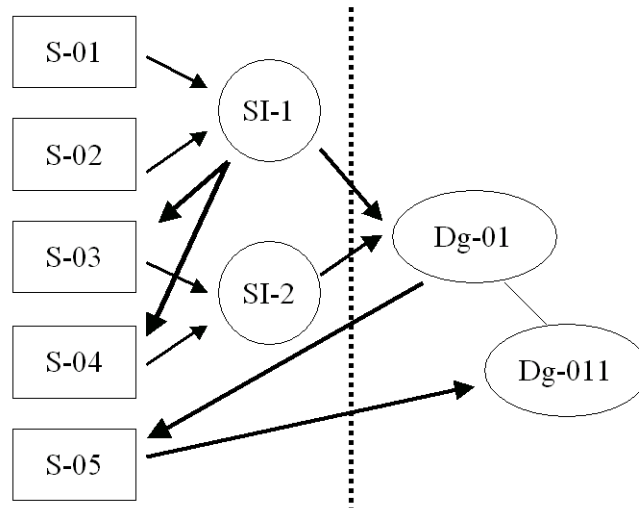


Abb. 3: Ablauf der Diagnosegenerierung in SonoConsult: Die Symptome S-01 und S-02 führen unmittelbar zu der Symptominterpretation SI-1 und von hier zur Annahme der Diagnose Dg-01. Die Symptominterpretation SI-1 kann ebenso wie eine angenommene Diagnose die Anforderung weiterer spezifizierender Symptome (S-03 und S-04) auslösen, die (ggf. über zwischengeschaltete Symptominterpretation SI-2) die Diagnose weiter festigen oder verwerfen. Im Falle der Etablierung einer Diagnose kann über weitere Symptome eine Spezifizierung angestrebt werden (Dg-011)

Die Auswahl der Fragebögen („Symptomklassen“) zu Beginn des Eingabedialogs erfolgt nach der Fragestellung, unter der die Untersuchung durchgeführt wurde (z. B. Anforderung des Stationsarztes), wobei alle praktisch wichtigen Fragestellungen als Antwortalternativen zur Auswahl angeboten werden. Von SonoConsult werden zu jedem Zeitpunkt der Erfassung die Diagnosen angezeigt, die das Programm aufgrund der bisherigen Dokumentation generiert („forward reasoning“). Während der Befundeingabe werden - getriggert durch diese Diagnosen - ergänzende Fragen präsentiert („backward reasoning“), um die Diagnosen zu spezifizieren, einzugrenzen oder auszuschließen („establish-refine“-Methode). Die Symptomerfassung wird beendet, wenn (a) der Untersucher selbständig in die abschließende Beurteilung springt, (b) alle Verdachtsdiagnosen (Status „möglich“) durch weitere Angaben in den Status „wahrscheinlich“ oder „unklar oder ausgeschlossen“ überführt sind oder (c) SonoConsult keine weiteren Fragen zur exakteren Abklärung einer Verdachtsdiagnose mehr generieren kann.

**Der Umfang von SonoConsult** wird durch das in der Wissensbasis repräsentierte medizinische Wissen bestimmt. Neben insgesamt 430 Fragen mit je 2 bis 6 Auswahlmöglichkeiten von Befunden oder Befundausprägungen enthält die Wissensbasis etwa 140 Symptominterpretationen, 230 Diagnosen und mehrere tausend Regeln verschiedener

Komplexität. Während der Dokumentation eines durchschnittlichen Falles werden aus 60 angegebenen Symptomen etwa 20 Symptominterpretationen und schließlich 3-4 diagnostische Schlussfolgerungen generiert.

**Die Datenanalyse** erfolgt in zwei Schritten. Zunächst kann der Benutzer mit einem Standardtool (Aufbau analog OLAP, On-Line Analytical Processing<sup>18</sup>) einen generellen Überblick über Daten im Fallspeicher erhalten. Das Tool ist unmittelbar in die Benutzeroberfläche von SonoConsult integriert und erlaubt eine manuelle Vorauswahl zu Basisdaten (Kategorien wie Alter, Gewicht, Geschlecht) und Diagnosen (Hierarchien wie Organe oder spezielle Organerkrankungen). Ein weiteres Subgruppen-Analyse-Tool kann zur Wissensentdeckung und Qualitätskontrolle angewandt werden. Die Technik der Subgruppenanalyse<sup>18,19,20</sup> ist für medizinische Zwecke gut anwendbar, z. B. zur Frage der Häufigkeit eines pathologischen Befundes bei gleichzeitiger Kombination anderer Symptome. Ebenfalls ausgewertet werden können Eigenschaften einzelner Untersucher, z. B. die Frage nach einer signifikant häufigeren Angabe eines Symptoms oder einer Diagnose durch einen Untersucher im Vergleich zur Gesamtmenge. In den vorliegenden Analysen wird das Tool VIKAMINE (Visual, Interactive and Knowledge-Intensive Analysis and Mining Environment) angewandt<sup>21,22</sup>, welches sowohl eine interaktive als auch eine automatische Subgruppenanalyse zulässt. Das Tool ist an die Besonderheiten medizinischer Daten, z. B. häufige fehlende Werte, angepasst, indem über intelligente Strategien zur Datensammlung die Zahl der Nachfragen minimiert wird. Weiterhin kann medizinisches Hintergrundwissen genutzt werden, um hiervon ausgehend neue, oft subtile Zusammenhänge aufzudecken. Einflussfaktoren wie Alter, Geschlecht oder Gewicht müssen bei der Interpretation berücksichtigt werden. VIKAMINE bietet zahlreiche und verschiedene heuristische Suchoptionen zur Entdeckung von Subgruppen mit Visualisierung der Ergebnisse, die dem Anwender jederzeit ein aktives Eingreifen in den Prozess auf einfache Art und Weise ermöglichen. Wenn der Nutzer bei Auswertungen mit dem Standardtool auf unerwartete, interessante oder unklare Ergebnisse trifft, können diese durch einen Wechsel zu VIKAMINE im Detail analysiert werden. Durch die Verwendung der gleichen Fallbasis wird der einfache Wechsel zwischen den Programmen möglich. Bei der Anwendung von VIKAMINE verspricht die Kombination der semi-automatischen Subgruppenanalyse mit medizinischem Hintergrundwissen durch die integrierte Visualisierungstechnik die effektivste Methode zu sein<sup>23</sup>. Dementsprechend nahmen medizinische „Experten“ unmittelbar die Datenanalyse vor.

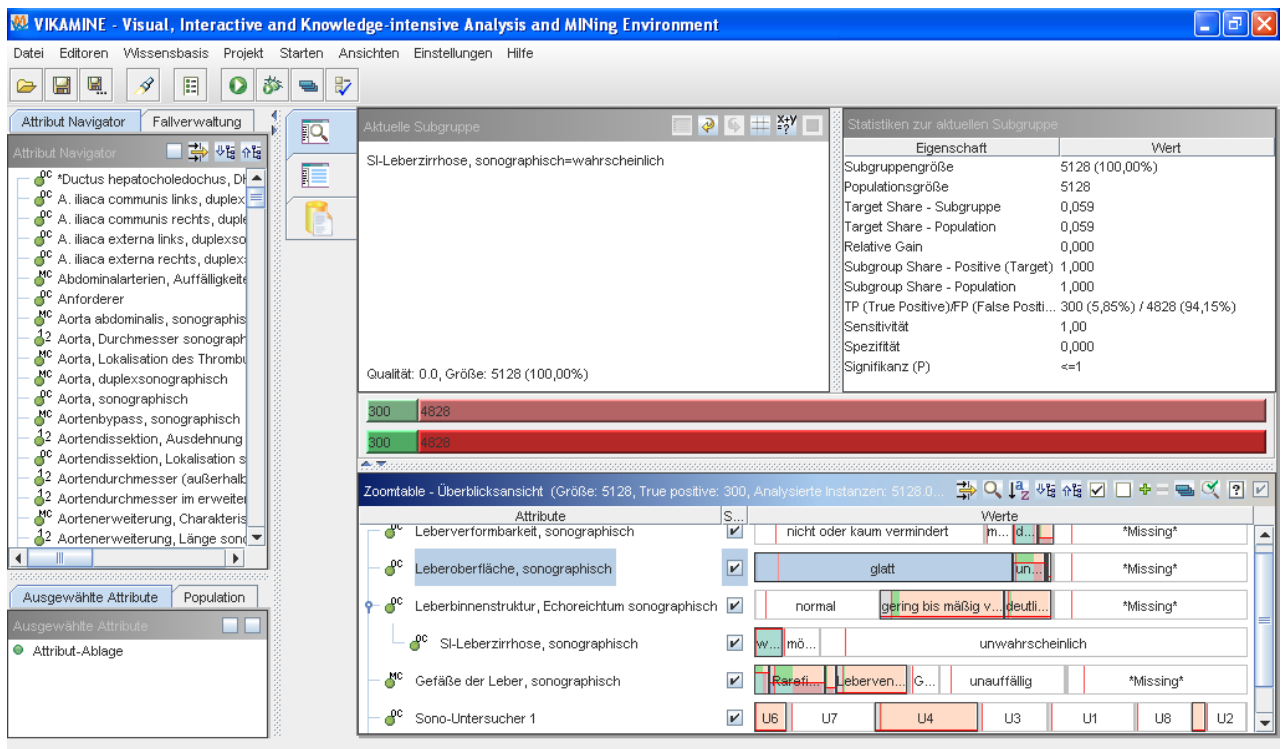


Abb. 4: Screenshot der Maske von VIKAMINE

**Die Erklärungskomponente** von SonoConsult stellt dem Anwender Wissen zur Verfügung und unterstützt somit vor allem weniger erfahrene Untersucher. Sie besteht aus drei Teilen.

- Einsicht in die aktiv gewordenen Regeln: Der Untersucher kann Folgerungen des Programms einsehen und dadurch sowohl die Herleitung der Systemdiagnosen als auch der vom Programm angebotenen Fragebögen oder Fragen nachvollziehen.
- Abrufbare Erklärungen zu Fragen und Diagnosen unmittelbar neben dem Text („I“ in Abb. 1).
- Verknüpfung mit einem Internetlehrbuch: Die meisten Diagnosen sind mit einem Internetlehrbuch verknüpft, welches weitere Fachinformationen liefern kann.

## 2.2. Begleitende Evaluation während der Einführung des Programms

Mit der Einführung einer Betaversion von SonoConsult in die Klinik erfolgte die Testung anhand realer Fälle. In dieser Phase von etwa einem Jahr wurde das Programm ständig auf einfache Fehler (z.B. Einfach-/Mehrfachauswahl in Frageklassen, Konfiguration der Ausgabeschablone), adäquate Präsentation von Symptomen, logische Abfolge der Frageklassen, korrekte Schlussfolgerungen bis hin zu den Diagnosen und deren Sicherheit sowie auf eine einheitliche Terminologie geprüft. Insgesamt wurden 299 Verbesserungen durch die Untersucher angeregt.

In diesem Zeitraum wurden 362 Veränderungen und Ergänzungen vorgenommen. Eine Aufschlüsselung zu einzelnen Frageklassen findet sich in Tabelle 1.

Organ	Vorschlag				Umsetzung		
	Inhalt wichtig	Inhalt weniger wichtig	Abfolge	Diagnose	Inhalt	Abfolge	Diagnose
Leber/DHC	10	11	11	28	40	9	9
Pfortader	2	3	6	12	4	1	6
Gallenblase	6	7	1	5	13	2	1
Pankreas	6	6	3	5	30	0	1
Milz	5	2	2	3	10	4	2
Nieren	9	8	15	20	38	10	6
Prostata	5	5	4	4	5	2	3
Harnblase	7	1	1	2	10	0	2
Gynäkologie	3	0	1	2	15	0	4
Aszites	2	0	2	2	15	0	4
Darm/Magen	2	3	2	7	15	2	3
Große Gefäße	7	2	2	4	38	1	4
Lymphknoten	2	4	2	1	2	0	1
Nebennieren	1	0	0	1	3	0	1
Schilddrüse	2	0	1	4	5	1	2
Andere	9	4	10	2	30	5	1
Insgesamt	78	56	63	102	275	37	50

Tabelle 1: Meldungen von Lücken, Fehlern und Vorschlägen zur Adaptation von SonoConsult durch Untersucher während der einjährigen Testphase, aufgeschlüsselt nach Organen

Die Befundeingabe und Diagnosegenerierung von SonoConsult wurde zunächst in der Betaversion und ein zweites Mal nach Abschluss der Weiterentwicklung zur Version 1.0 nach einem Jahr beurteilt. Dazu erfolgte die Einschätzung aller untersuchten Fälle (Betaversion n = 121 und Version 1.0 n = 186) über drei Wochen durch jeweils vier erfahrene Untersucher (mit Erfahrung von je über 1000 Untersuchungen) zeitnah zur Befundeingabe anhand eines Fragebogens.

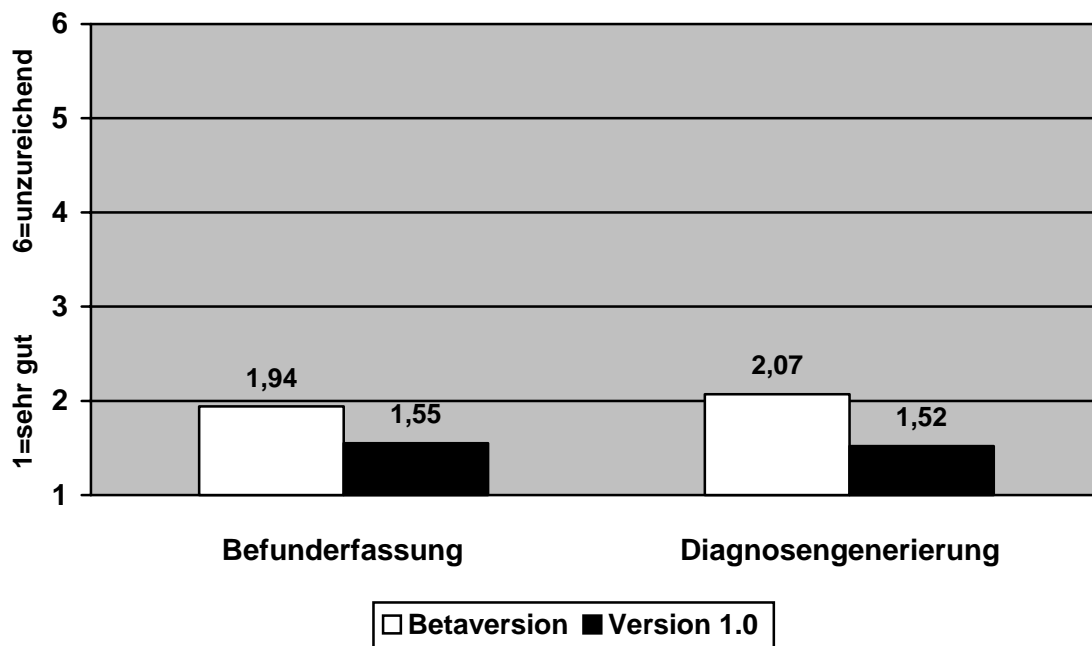


Tabelle 2: Bewertung Befundeingabe und Diagnosegenerierung

Seit 2002 wird SonoConsult als Dokumentationssystem für Ultraschalluntersuchungen in der Sonographieabteilung der II. Medizinischen Klinik an den DRK-Kliniken Berlin-Köpenick eingesetzt (etwa 400 Untersuchungen im Monat). Es befindet sich seit 2005 weiterhin im klinischen Einsatz an der Medizinischen Universitätsklinik Würzburg. Es wurde jeweils eine Anbindung an das jeweilige Krankenhaus-Informationssystem (KIS) eingerichtet. Die Befundberichte werden in der digitalen Patientenmappe gespeichert. In Berlin wird der Standardmodus zur Falldokumentation benutzt. Den Untersuchern bleibt freigestellt, die vom Programm präsentierten Fragen zu überspringen und die generierten Diagnosevorschläge in die eigene Beurteilung zu übernehmen.



## **2.3. Evaluation**

Es wurden folgende Evaluationen durchgeführt:

1. Messung der Dauer einer Befundberichterstellung (a) in freitextlicher Weise und (b) mit Hilfe von SonoConsult,
2. Erfassung der Erwartungen der Untersucher vom Einsatz eines computergestützten Befundungssystems,
3. Beurteilung von SonoConsult und der mit dem Programm erstellten Befundberichte durch die Untersucher,
4. Beurteilung der mit Hilfe von SonoConsult erstellten Befundberichte durch die Stationsärzte,
5. Evaluation von SonoConsult bezüglich der Rolle bei der Einarbeitung von Anfängern in der Sonographie,
6. Vergleich von freitextlichen und SonoConsult-Befundberichten bezüglich Befundqualität, Vollständigkeit und Richtigkeit der Diagnosen,
7. Evaluation der diagnostischen Genauigkeit von SonoConsult,
8. Evaluation der Übereinstimmung der Untersucherdiagnosen mit den durch SonoConsult hergeleiteten Diagnosen,
9. Evaluation (a) der Vollständigkeit der Befundberichte im Beurteilungsteil gemessen an den Befundeinträgen und (b) der Befundeinträge gemessen an den im freitextlichen Beurteilungsteil angegebenen Diagnosen.

### **2.3.1. Eingabedauer**

121 konsekutive Fälle wurden zur Evaluation von SonoConsult in einer ersten Fassung im klinischen Alltag von 4 erfahrenen Untersuchern (>1000 Untersuchungen) über drei Wochen erfasst (siehe Kapitel „Entwicklung der Wissensbasis und Einführung des Programms“). Gleichzeitig erfolgte die Befundschreibung in Freitext mit der Verwendung einer Standardbefund-Vorlage für die Abdomensonographie, in der Abweichungen des jeweiligen Falles ergänzt wurden. Die Dauer der Befundeingabe wurde für beide Dokumentationsformen erfasst und verglichen (Zeit vom Start der Eingabe bis zur Anzeige des Befundberichtes am Monitor). Nachdem SonoConsult über ein Jahr anhand realer Fälle kontinuierlich angepasst und überarbeitet wurde, erfolgte eine erneute Evaluation (n=186, identische vier Untersucher), bevor

SonoConsult in dieser Fassung durch die folgenden Untersuchungen evaluiert wurde. In dieser Serie wurde erneut die Dauer der Befundeingabe erfasst.

## **2.3.2. Akzeptanz**

### **2.3.2.1. Akzeptanz bei Untersuchern**

Die Akzeptanz von SonoConsult wurde in drei Kliniken (DRK-Kliniken Berlin-Köpenick und Berlin-Mitte, Universitätsklinikum Würzburg) durch einen Vergleich von Erwartungen (a) vor Einführung und eine analoge Erhebung (b) nach Einarbeitung in das Programm beurteilt. In einem anonymen und standardisierten Fragebogen erfolgte die Bewertung anhand einer Fünf-Punkte-Likert-Skala.

- (a) Der erste Fragebogen wurde bei perspektivischen Anwendern von SonoConsult vor der ersten Präsentation des Programms erhoben und von insgesamt 19 Untersuchern beantwortet, bestehend aus 8 erfahrenen Untersuchern (mehr als 400 eigenständige Untersuchungen), 6 jüngeren Untersuchern (weniger als 400 eigenständige Untersuchungen) und 5 Anfängern. Neben allgemeinen Fragen zur Qualitätssicherung in der Sonographie und dem Einsatz eines computerbasierten Programms in diesem Zusammenhang wurden Erwartungen an SonoConsult erhoben.
- (b) Der zweite Fragebogen wurde erhoben, nachdem die Untersucher das Programm für drei Monate angewandt hatten. Die Rücklaufquote dieses Fragebogens betrug 74% (14 von 19). Mehrere Fragen aus der ersten Evaluation wurden erneut gestellt, um einen Vergleich vornehmen zu können. Einzelne Programmkomponenten (Befundeingabe, Diagnosegenerierung) wurden differenzierter beurteilt.

### **2.3.2.2. Akzeptanz bei Stationsärzten**

Ein Jahr nach Einführung von SonoConsult als Dokumentationssystem in Berlin-Köpenick wurde ein anonym standardisierter Fragebogen unter den Stationsärzten verteilt, welcher eine Beurteilung der mit SonoConsult erstellten Befundberichte erfragte. Die Bewertung erfolgte anhand einer Fünf-Punkte-Likert Skala. Für eine differenzierte Auswertung wurden der Ausbildungsstand des Stationsarztes (Arzt im Praktikum, Arzt in Weiterbildung, Facharzt) sowie eigene Kenntnisse in der Abdomensonographie (mehr als 400 eigenständige Untersuchungen) erfasst. Es beteiligten sich 37 von 40 Stationsärzten (92,5%).

### **2.3.3. Subjektive Einschätzung des Ausbildungseffekts bei den Untersuchern**

In den DRK-Kliniken Köpenick existiert eine planmäßige Rotation von Ärzten in die Sonographieabteilung für 6 Monate während der Facharztausbildung. Zu Beginn wird die Einarbeitung durch einen Tutor gewährleistet, bis eigenständige Untersuchungen durchgeführt werden. Das Ende der Einarbeitung wird vom Tutor bestimmt, bei schwierigen Fragestellungen und Interpretationen werden regelhaft weiterhin erfahrende Untersucher zu Rate gezogen. Voraussetzung für einen Einsatz in der Sonographie sind die Absolvierung von Grund- und Aufbaukurs nach DEGUM (Deutsche Gesellschaft für Ultraschall in der Medizin). Insgesamt acht Untersucher, die während des Erhebungszeitraumes ihre Rotation in die Sonographieabteilung begannen, erhielten einen anonymisierten und standardisierten Fragebogen zur Evaluation des Einflusses von SonoConsult auf die sonographische Befundung und Diagnoseerhebung. Die Rückgabe der Fragebögen erfolgte am Ende der Einarbeitungszeit (zwischen 400 und 800 Untersuchungen mit Tutor).

### **2.3.4. Qualität der Befundberichte**

Es wurde untersucht, (a) ob die Verwendung von SonoConsult als Dokumentationsprogramm dazu führt, die Vollständigkeit der Befunddokumentation einer Untersuchung zu erhöhen und (b) wie hoch die Übereinstimmung der Untersucherdiagnosen und der Systemdiagnosen mit den von unabhängigen erfahrenen Untersuchern (im folgenden „Evaluatoren“) aus den Eingaben ableitbaren Diagnosen ist.

#### **2.3.4.1. Evaluation der Vollständigkeit von Untersuchungsbefunden**

Es wurde der Inhalt von 103 konsekutiven Freitextdokumentationen (7 Untersucher, mehr als 400 eigenständige Untersuchungen; vollständige Abdomensonographien) verwendet, um eine Falldokumentation mit SonoConsult durchzuführen. Dazu wurden die in diesen - vor Einführung von SonoConsult erstellten - Untersuchungsberichten enthaltenen Informationen in die Fragebögen von SonoConsult eingegeben. SonoConsult generierte Fragen nach Befunden und Daten, die nur zum Teil aus den Freitextdokumentationen beantwortet werden konnten. Zu einem Teil der SonoConsult-Fragen existierten in den Freitextbefunden keine Angaben. Die

Antwort wurde entsprechend auf „unbekannt“ gesetzt. Nach Einschätzung durch zwei Evaluatoren wurden die fehlenden Angaben als „essentiell unbekannt“ (fehlende Information ist wichtig) oder „erläßlich unbekannt“ (fehlende Information ist weniger wichtig) eingeordnet, wobei eine möglichst differenzierte sonographische Abklärung des Befundes als Ziel zugrunde gelegt wurde.

### **2.3.4.2. Evaluation der Diagnosequalität**

Die Evaluation der Diagnosequalität erfolgte gemessen an den im Dokumentationsteil enthaltenen Informationen zu Befunden und Messwerten.

- Erste Fallserie: Die zwei Evaluatoren (s. o.) verglichen die Diagnosen der Freitextbefunde mit denen von SonoConsult. Die eigenen Schlussfolgerungen der Evaluatoren basierend auf den gleichen Patientendaten im Dokumentationsteil stellten dabei den Goldstandard dar. Die beiden Evaluatoren haben sich bei unterschiedlichen Meinungen mit einem dritten Evaluator beraten und in strittigen Fällen auf eine gemeinsame Meinung geeinigt.
- Zweite Fallserie: Die alleinige Betrachtung von diagnostischen Folgerungen, die aus der Eingabe von Freitextbefunden in standardisierte Computerformulare resultieren, ist schwierig, da die Befundbeschreibung im Freitext keiner einheitlichen Nomenklatur folgt und somit bei der Überführung der Daten ein gewisser Interpretationsspielraum eingeräumt werden muss. Ein Jahr nach Einführung von SonoConsult als Dokumentationssystem wurden in einer Fallserie von 130 konsekutiven Abdomensonographien die von den Untersuchern angegebenen Diagnosen hinsichtlich ihrer Ableitbarkeit aus den eingegebenen Befunden und Messwerten und ihrer Konkordanz mit den Diagnosevorschlägen von SonoConsult untersucht. Die diagnostischen Schlussfolgerungen der zwei unabhängigen Evaluatoren dienten wieder als Goldstandard.
- Dritte Fallserie: Eine dritte Fallserie von 150 konsekutiven Abdomensonographien wurde analog der zweiten Evaluation ausgewertet mit dem Unterschied, dass das Diagnosefenster von SonoConsult in der Eingabemaske ausgeblendet wurde. Der Untersucher konnte somit bei Verfassen seiner Beurteilung nicht auf die Erinnerungsfunktion des Programms zurückgreifen. Die Systemdiagnosen standen den zwei Evaluatoren in der Auswertung jedoch zur Verfügung. Die drei beteiligten Untersucher der Fallserie äußerten in einem standardisierten anonymen Fragebogen ihre

Erwartung, welchen Einfluss das Ausblenden der Programmdiagnosen auf die Vollständigkeit der Befundung haben wird.

Um die Ergebnisse der Evaluationen vergleichend und umfassend interpretieren zu können, wurde eine Erhebung bei 7 erfahrenen Untersuchern zur Nutzung der Diagnosefunktion von SonoConsult durchgeführt. Diese erfolgte wiederum in der Form eines anonymen standardisierten Fragebogens.

### **2.3.5. Diagnostische Genauigkeit**

Um die diagnostische Kompetenz von SonoConsult zu prüfen, wurde eine prospektiv gesammelte Fallserie von 112 konsekutiven Abdomensonographien untersucht. Die Fälle wurden von drei erfahrenen Untersuchern der Sonographieabteilung mit SonoConsult dokumentiert. Drei unabhängige Evaluatoren sahen die Befundberichte einzeln durch und verglichen die vom System generierten Diagnosen mit denen der Untersucher. Der Vergleich erfolgte ausschließlich auf der Basis der dokumentierten und im schriftlichen Befund nachlesbaren Angaben.

Die diagnostischen Schlussfolgerungen des Programms wurden als richtig bzw. falsch oder problematisch angesehen, wenn sie auf Grund der Befunddaten nach Meinung von wenigstens zwei der drei Evaluatoren in dieser Weise zu interpretieren waren. Es ist zu berücksichtigen, dass SonoConsult in Freitextfelder eingegebene Befunde nicht für die Diagnosefindung verwerten kann. Zur Beurteilung der Programmgenauigkeit mussten Untersucherfehler (Eingabefehler, Informationslücken) berücksichtigt werden.

Als **Eingabefehler** wurden Angaben gewertet, die in offensichtlichem Widerspruch zu der vom Untersucher favorisierten Diagnose stehen. Beispiel: In einem Fall wurde von einem Untersucher eine Leberzirrhose diagnostiziert; bei der Befundbeschreibung wurden jedoch „Leberverformbarkeit = vermindert“ und „Echogenität = vermindert“ eingetragen. Das Programm wertet diese Angaben als gegen eine Leberzirrhose sprechend; die Schlussfolgerung des Programms auf das Nichtvorliegen einer Leberzirrhose ist damit formal richtig. Dennoch kann die Bewertung des Untersuchers korrekt sein und er eher die verkehrte Antwortalternative angekreuzt als seine Befunde falsch interpretiert haben. Die Inkongruenz zwischen dokumentierten Befunden und einer Diagnose beruhte daher nach Meinung der Evaluatoren am ehesten auf einem Eingabefehler. Als **Informationslücke** wurde gewertet, wenn aufgrund der

Eingaben vom Programm Befunde erfragt, vom Untersucher aber nicht eingegeben wurden. Beispielsweise erfordert die Diagnose „Leberzirrhose“ eine Angabe zur intrahepatischen Gefäßstruktur. Fehlte sie in der Dokumentation, wurde dies von den Evaluatoren als Informationslücke registriert. Sowohl Eingabefehler als auch Informationslücken wurden bei der Bewertung der Konkordanz zwischen den Untersucherdiagnosen und den Systemdiagnosen berücksichtigt. Eine Diskrepanz der Diagnosen, die sich auf eine mangelhafte Befundeingabe zurückführen ließ, wurde nicht als diagnostischer Programmfehler gewertet.

Bei der Evaluation der diagnostischen Kompetenz des Programms wurden einfache und komplexe Schlussfolgerungsprozesse unterschieden. Einfache Schlussfolgerungen führen beispielsweise zu den Diagnosen „Prostatavergrößerung“, „Nierenarterienstenose“, „Aortenektasie“ oder „Aortenaneurysma“, da sie auf der Interpretation nur von einem oder weniger Parameter beruhen und keine Regelinterferenzen zu berücksichtigen sind. Durch komplexe Schlussfolgerungsprozesse mit nicht-trivialen Regeln werden durch SonoConsult folgende Diagnosen hergeleitet: Leberzirrhose, Fettleber, Stauungsleber, Cholestase, portale Hypertension, Leberherd mit Dignitätsangabe, chronisch degenerative Nierenerkrankung, akute und chronische Pankreatitis. Diese Diagnosen wurden in jedem Einzelfall durch Einsicht in die Herleitungsregeln auf Richtigkeit der zu ihnen führenden Schlussfolgerungsprozesse überprüft.

### **2.3.6. Konkordanz von Befund und Beurteilung**

Die beobachteten Inkonsistenzen von in den Computerformularen dokumentierten Befunden und der freitextlichen Beurteilung des Untersuchers – nicht deren Korrektheit - wurden evaluiert, um die Verlässlichkeit der Diagnoseangaben eines Sonographiebefundes einschätzen zu können. Zur Prüfung der Konkordanz ist es einerseits notwendig festzustellen, ob alle angegebenen Diagnosen aus der Befundeingabe nachvollziehbar sind und zum zweiten, ob aus diesen Daten weitere, nicht in der Beurteilung diskutierte diagnostische Schlüsse gezogen werden können. Die Auswertung wurde in zwei Schritten durchgeführt, um diese zu standardisieren und effektiv zu gestalten:

1. Die von SonoConsult aus den Befundeingaben generierten Diagnosen wurden auf ihre Übereinstimmung mit den Diagnosen der Untersucher geprüft. Dieser Schritt wurde mit Hilfe eines erwartungsbasierten linguistischen Pre-Prozessors<sup>24</sup> automatisiert, welcher aus SonoConsult-Diagnosen und der Beurteilung des Untersuchers in Freitext eine

Tabelle erstellt, in der jeder Angabe im Freitext eine Systemdiagnose zugeordnet wird und umgekehrt (Tabelle 3).

2. Im Falle von Diskrepanzen wurde eine eingehendere Beurteilung des Falles durch zwei Evaluatoren vorgenommen.

<b>Teil der diagnostischen Folgerung im Freitext</b>	<b>SonoConsult-Diagnosen aus den gleichen Befunddokumentation</b>
Deutlich eingeschränkte Untersuchungsbedingungen bei Adipositas und Meteorismus	Adipositas Grad II entsprechend BMI
Fettleber, Hinweise für fibrotischen Umbau	Fettleber
Kleine Leberzyste von 1 cm in Segment IV, dem Patienten bekannt	Leberzyste(n)
Große Nierenzyste links von etwa 4 cm	Nierenzyste(n)
Pankreaslipomatose	Pankreaslipomatose
-	Prostatavergrößerung
-	Leicht vergrößerte Niere rechts
-	Leicht vergrößerte Leber

Tabelle 3: Methodik zum Aufdecken von Inkonsistenzen. Die linke Spalte enthält fragmentierte Teile der Beurteilung im Originalbefund: „Deutlich eingeschränkte Untersuchungsbedingungen bei Adipositas und Meteorismus. Fettleber mit vermutlich fibrotischem Umbau. Kleine Leberzyste von etwa 1 cm in Segment IV (Patient bekannt). Relativ große Nierenzyste von ungefähr 4 cm Durchmesser links. Pankreaslipomatose.“ Die rechte Spalte zeigt die vom Programm generierten Diagnosen. Die Allokation der Fragmente mit SonoConsult-Diagnosen erfolgte computerisiert durch heuristisches sprachliches Pre-Prozessing und wurde manuell überprüft. Die letzten drei Einträge wurden als (minor) Inkonsistenzen gezählt im Sinne von weggelassenen Diagnosen (falsch negativ)

**Erste Fallserie.** In dieser Serie wurde untersucht, ob die Häufigkeit von Dokumentationsmängeln und diagnostischen Fehlern der Untersucher bei Diagnosen verschiedener klinischer Relevanz unterschiedlich ist. Für die Evaluation wurden 250 konsekutive Fälle prospektiv gesammelt und hinsichtlich Konsistenz der eingegebenen Befunde und der von den Untersuchern gefolgerten Diagnosen geprüft. Die Fälle wurden nicht nach ihrer Fragestellung gefiltert und umfassten sowohl Sonographien des gesamten Abdomens als auch Untersuchungen einzelner Organe (zum Beispiel nur Schilddrüse oder kurzfristige Verlaufskontrolle eines Organbefundes bei kürzlich erfolgter Untersuchung des gesamten

Abdomens). Die Untersuchungen wurden von drei erfahrenen Untersuchern durchgeführt, die jedoch nicht über die geplante Auswertung der Befundberichte informiert waren. Ihre Zustimmung wurde nach Abschluss der letzten Untersuchung und vor Beginn der Auswertung eingeholt. Die Diagnosen wurden nach klinischer Bedeutung in zwei Relevanzgruppen eingeordnet, die durch die zwei an der Auswertung beteiligten Evaluatoren schematisch festgelegt wurden. Zum Beispiel bedeutet dies, dass Diagnosen wie Leberzirrhose, chronische Pankreatitis und Aszites vermutlich größere Relevanz zugeschrieben werden kann als einem Zustand nach Cholezystektomie, einer leichten Lebervergrößerung oder einer Pankreaslipomatose.

**Zweite Fallserie.** Eine prospektive Fallserie von 161 konsekutiven Fällen wurde wiederum auf die Konkordanz von eingegebenen Befunden und diagnostischen Folgerungen des Untersuchers analysiert, jedoch die Diskrepanzen in dieser Serie hinsichtlich ihrer Komplexität unterschieden. Die Fälle wurden nicht nach Durchführung einer vollständige Abdomensonographie (Angaben zur mehr als 5 Organen in 64%) gefiltert und enthielten somit zum Teil Untersuchungen einzelner Organe (16%). Eine weitere Charakterisierung der Fallserie enthält Tabelle 4.

<b>Untersuchungen gesamt</b>	<b>161</b>	<b>100%</b>
Leber	126	78,3%
Gallenwege/ Gallenblase	125	77,6%
Pfortader	113	70,2%
Milz	107	66,5%
Pankreas	77	47,8%
Niere	116	72,1%
Blase/Prostata	30	18,6%
Retroperitoneum (Aorta, Vena cava, Lymphknoten)	108	67,1%
Darm	19	11,8%
Schilddrüse	6	3,7%

Tabelle 4: Charakterisierung der Fallserie (n = 161) bezüglich untersuchter Organe

Die Auswertung dieser Fallserie differenzierte nach Diagnosen mit einfacher und mit komplexer Herleitung. Beispiele für Diagnosen einfacher Herleitung sind eine extrahepatische Gallengangektasie ohne intrahepatische Gangerweiterung, Gallenblasenhydrops oder Aszites.



Diagnosen komplexer Herleitung werden über nicht triviale Regeln generiert wie zum Beispiel Leberzirrhose, Fettleber, fokale Leberläsion, Stauungsleber, obstruktive Cholestase, portale Hypertension, Pankreaserkrankungen und chronisch degenerative Nierenerkrankung.

## **2.4. Statistische Auswertungen**

Mit Hilfe des Subgruppen-Analyse-Tool VIKAMINE wurden in Zusammenarbeit mit dem Institut für Informatik der Universität Würzburg folgende statistische Auswertungen vorgenommen:

1. Evaluation der Inter-Untersucher-Variation,
2. Exemplarische Auswertung der Falldatenbank bezüglich medizinischer Fragestellungen.

### **2.4.1. Inter-Untersucher-Variation**

Für die Auswertung wurden die über die Dauer von drei Jahren unter Anwendung von SonoConsult erstellten sonographischen Befundberichte von insgesamt 8 Untersuchern verwendet und miteinander verglichen. Da sich die Zahl der Befunde je Untersucher unterschied und eine steigende Untersuchungszahl als Ausdruck zunehmender Erfahrung verstanden werden kann, wurden in der Auswertung nur die ersten 600 konsekutiven Fälle eines Untersuchers evaluiert. Wiederholungsuntersuchungen von Patienten wurden in den Fallsammlungen nicht berücksichtigt. Ein Untersucher mit weniger als 300 Fällen wurde von weiteren Analysen ausgeschlossen. Zwei Untersucher führten etwas weniger als 600 Untersuchungen durch (574 und 526) und wurden in der Auswertung mitgeführt. Die übrigen 5 Untersucher überschritten die Zahl von 600 Untersuchungen. Aus diesen Gründen umfasste die für die Studie zur Verfügung stehende Fallsammlung 4100 Fälle von 7 Untersuchern. Die Serien der einzelnen Untersucher unterschieden sich nicht signifikant bezüglich Alter und Geschlecht. Die Fälle wurden in die ersten, zweiten und dritten 200 Fälle unterteilt, um Lernkurven feststellen zu können.

Die exemplarische Darstellung einer Inter-Untersucher-Analyse erfolgte beispielhaft an den komplex herzleitenden Diagnose „Leberzirrhose“ und „Chronisch degenerative Nierenerkrankung“. Die Experten der Wissensbasis gaben die Filter für die Charakterisierung der Subgruppe vor, indem sie die potentiell relevanten Befunde für die entsprechende Diagnose festlegten. Andere für die Diagnose uninteressante Informationen zu Normalwerten, wie zum Beispiel „Lebergefäße = normal“ für die Diagnose „Leberzirrhose“, wurden herausgefiltert.

Bestimmte Werte wurden zusammengefasst, wenn dies medizinisch sinnvoll erschien, um Klassen von unverbundenen Werten zu bilden, am Beispiel „Leberverformbarkeit = mäßig *oder* stark vermindert“. Um die Anzahl fehlender Attribute in der Auswertung zu minimieren, wurde – wo immer aus dem klinischen Kontext möglich – der Wert „normal“ angenommen. Die Subgruppen mit den gewünschten Kriterien wurden mit dem Tool VIKAMINE gebildet.

Das angewandte Prozessmodell mit Visualisierung der einzelnen Komponenten ermöglichte eine schrittweise Datenanalyse: Zuerst wurden einzelne Faktoren identifiziert, um anschließend die Subgruppen-Analyse zu verfeinern. Durch die automatische Suchmethode abgegrenzte Subgruppen konnten durch interaktive Vorgehensweise bestätigt und näher charakterisiert werden.

Die Subgruppen-Analyse der Fallserien erfolgte mit der Frage, ob Untersucher signifikant in ihrem Befundungsmuster, d. h. in ihren Angaben zu einzelnen sonographischen Diagnosen oder Befunden, von der Gesamtgruppe abweichen. Wenn solche Unterschiede festgestellt wurden, konnte vermutet werden, dass ein Untersucher mehr oder weniger häufig bestimmte Befunde dokumentiert, die diese Subgruppe charakterisieren. Diese Hypothese wurde dann gegen die Nullhypothese statistisch geprüft, die ein rein zufällig in der Fallreihe eines Untersuchers häufigeres Auftreten eines Befundes bedeuten würde (Standard-Chi-Quadrat-Test für Unabhängigkeit).

#### **2.4.2. Subgruppen-Analyse zur Wissensentdeckung**

Die Dokumentation der Sonographiebefunde mit SonoConsult führt zum Aufbau einer Falldatenbank, die mit zunehmender Größe für die Untersuchung medizinischer Fragestellungen von Interesse ist. Während die Zusammenhänge pathologischer Befunde an einem Organ meist bekannt sind, ist zu erwarten, dass mit Hilfe des standardisiert dokumentierten Datenmaterials oftmals subtile Zusammenhänge verschiedener Organsysteme aufgedeckt werden können. Am Beispiel der Diagnose „Gallensteine“ wird exemplarisch die Eignung der Datenbank von 4100 Fällen (ohne Wiederholungsuntersuchungen) für die Untersuchung medizinischer Assoziationen dargestellt.

## **3. Ergebnisse**

### **3.1. Evaluation**

#### **3.1.1. Eingabedauer**

Der Zeitbedarf für die Erstellung eines Befundberichts wurde gemessen:

1. Die Dokumentation eines Befundes und Erstellung eines Befundberichts in Freitext (unter Verwendung von Textbausteinen für häufige Formulierungen) für eine vollständige Abdomensonographie von mindestens 5 Organen (Leber mit Gallenblase und Gallenwegen, Pfortadersystem, Nieren, Pankreas, Milz) betrug zwischen 3 und 5 Minuten.
2. Der Zeitbedarf für die Dokumentation eines Befundes in SonoConsult (Beantwortung aller generierten Fragen) bis zur Befundberichtserstellung betrug für die gleichen Fälle zwischen 4 und 13 Minuten (7,5 Minuten im Mittel), wenn das Programm erstmalig verwendet wurde. Nach einer Einarbeitungszeit von wenigen Wochen (entsprechend mindestens 300 Untersuchungen) betrug der Dokumentationsaufwand bei kontinuierlicher Nutzung des Programms weniger als 5 Minuten. Nach über 1000 Untersuchungen wurde eine Eingabedauer von 2 bis 4 Minuten für eine vollständige Abdomensonographie erfasst.

## 3.1.2. Akzeptanz

### 3.1.2.1. Akzeptanz bei Untersuchern

Die Einführung eines computerbasierten Befundungsprogramms wurde von den Untersuchern skeptisch gesehen. Ein möglicher Zeitverlust durch extensive Dokumentation, fehlende Flexibilität der zu beantwortenden Fragen und geringerer Kontakt zum Patienten stellten die wesentlichen Begründungen dar.

Die Ergebnisse der Befragung von Untersuchern, die perspektivisch mit SonoConsult arbeiten sollten, zeigt Tabelle 5.

A1	Einfluss auf die Untersuchungsstrategie	3,1	1=nicht	5=sicher
A2	Möglicher zeitlicher Mehraufwand	3,6	1=nicht vertretbar	5=vertretbar
A3	Standardisierte Dateneingabe	4,3	1=unwichtig	5=sehr wichtig
A4	Anzeige unvollständige Untersuchung	3,7	1=unwichtig	5=sehr wichtig
A5	Anzeige Programm-Diagnosen	3,0	1=unwichtig	5=sehr wichtig
A6	Einfache Bedienbarkeit	4,9	1=unwichtig	5=sehr wichtig
A7	Schneller Zugriff auf Vorbefunde	4,6	1=unwichtig	5=sehr wichtig
A8	Schnelle Einarbeitung von Anfängern	3,9	1=unwichtig	5=sehr wichtig
A9	Statistische Analysen	3,8	1=unwichtig	5=sehr wichtig
A10	Erklärungsfunktion	3,8	1=unwichtig	5=sehr wichtig

Tabelle 5: Erwartungen und Anforderungen von Ärzten der Medizinischen Kliniken (n = 19) an SonoConsult vor erster Präsentation

A1, A2... = Fragekürzel

Die Erwartungen der Untersucher in Bezug auf eine programmgestützte Befunderhebung betreffen vor allem die Vereinheitlichung der Nomenklatur, einen schnellen Zugriff auf Vorbefunde und eine vollständige Befunderhebung. Als Voraussetzung der Akzeptanz eines solchen Programms werden eine einfache Handhabung und differenzierte Eingabemöglichkeiten bei adäquatem Zeitaufwand hervorgehoben. Ein zeitlicher Mehraufwand für eine bessere Befundqualität wurde von mehr als der Hälfte der Untersucher ohne oder mit geringen Einschränkungen als vertretbar angesehen. Weitere Funktionen wie zum Beispiel die Nutzung für statistische Analysen oder Erklärungskomponenten des Programms erhielten eine nahezu neutrale Bewertung.

Die Ergebnisse der Befragung nach Einarbeitung in SonoConsult zeigt Tabelle 6.

B1	Strukturierte Befunderhebung	3,8	1=störend	5=hilfreich
B2	Befundeingabe (ohne Freitextfelder)	3,7	1=unzureichend	5=zu differenziert
B3	Erinnerungsfunktion	3,8	1=störend	5=sehr hilfreich
B4	Nutzung der Erklärungsfunktion	2,5	1=niemals	5=sehr häufig
B5	Relevanz von SC-Diagnosen	2,9	1=unwichtig	5=sehr wichtig
B6	Einfluss SC auf eigene Diagnosen	2,2	1=nicht	5=wesentlich
B7	Standardisierte Nomenklatur	4,4	1=nachteilig	5=vorteilhaft
B8	Vergleichbarkeit von Befunden	4,5	1=unwichtig	5=sehr wichtig

Tabelle 6: Beurteilung der Untersucher (n=14) zu SonoConsult und dessen Funktionen nach Einarbeitung in das Programm.

SC = SonoConsult

B1, B2... = Fragekürzel

Nach Einführung von SonoConsult in die Routinedokumentation definierten die Untersucher die einheitliche Nomenklatur und damit die Vergleichbarkeit der Befunde zwischen den Untersuchern und zwischen verschiedenen Untersuchungen als wesentliche Vorteile. Weitere Eigenschaften des Programms wie die Möglichkeit zu statistischen Analysen, die Nutzung der Systemdiagnosen zur Überprüfung der eigenen Beurteilung bezüglich Richtigkeit und Vollständigkeit und die Erklärungsfunktion des Programms wurden neutral bis gering positiv bewertet. Der Einfluss der programmgenerierten Diagnosevorschläge auf die eigenen diagnostischen Schlussfolgerungen (diagnostische Hilfe) wurde als gering eingestuft. Die Differenziertheit der Eingabe schien der Mehrheit der Untersucher vor dem Hintergrund des zeitlichen Aufwandes der Dokumentation adäquat, im Verlauf wurden mehrfach Wünsche nach einem Ausbau des Programms zur Dokumentation seltener Befunde geäußert. Diese wurden bei der weiteren Programmentwicklung berücksichtigt.

Der Vergleich der Erhebungen zeigt, dass sich Erwartungen und Erfahrungen in vielen Aspekten deckten: Die standardisierte Nomenklatur erwies sich als positiv (siehe Fragen A3 / B7, B8), die Befundeingabe erschien adäquat (A2, A6 / B1, B2) und die Erinnerungsfunktion erwies sich als hilfreich (A4 / B3). Die Generierung und Mitteilung von Programm-Diagnosen wurde als weniger wichtig wahrgenommen (A5 / B5), noch geringer wird der Einfluss auf die eigene Befundung angegeben. Eine Abweichung von Erwartung und Erfahrung zeigte sich bezüglich

der Erklärungsfunktion, welche zwar als wünschenswert angegeben wurde, während der Nutzung von SonoConsult jedoch selten zur Anwendung kam (A10 / B4).

### 3.1.2.2. Akzeptanz bei Stationsärzten

Die Beurteilung der von SonoConsult generierten Befundberichte durch die Stationsärzte wird in Tabelle 7 dargestellt.

			Alle (37)	AiP (8)	AiW (23)	FA (6)	US+ (21)	US- (13)
C1	Umstellung der Sonographiebefundung	1=negativ 5=positiv	3,6	3,2	3,8	3,0	3,6	3,4
C2	Veränderter Stellenwert eines sonographischen Befundes	1=nicht 5=deutlich	2,3	2,2	2,4	2,3	2,4	2,2
C3	Erhöhter Informationsgehalt	1=keine Rolle 5=wesentliche	2,9	3,5	2,9	2,3	2,8	2,9
C4	Vereinheitlichung der Nomenklatur	1=keine Rolle 5=wesentliche	3,5	4,2	3,6	3,3	3,7	3,1
C5	Bessere Befundvergleichbarkeit	1=nicht 5=deutlich	3,6	3,3	3,9	3,3	3,7	3,3
C6	Angabe „nicht untersuchter“ Organe für Befundwertigkeit	1=keine Rolle 5=wesentliche	3,8	4,1	3,6	4,3	4,0	3,5
C7	Erfassung einzelner Organbefunde	1=nie 5=immer	3,8	3,8	3,7	4,2	4,0	3,7
C8	Erfassung der Diagnosen von SonoConsult	1=nie 5=immer	3,0	3,0	3,2	2,0	3,0	2,6
C9	Erfassung der Diagnosen des Untersuchers (Beurteilung)	1=nie 5=immer	4,9	5,0	4,9	5,0	4,9	5,0
C10	Mitausgabe der Diagnosen von SonoConsult	1=störend 5=hilfreich	2,6	3,0	2,6	1,8	2,3	2,6
C11	Diagnosen von SonoConsult	1=keine Rolle 5=wesentliche	2,1	1,9	2,3	1,7	2,1	1,9

Tabelle 7: Evaluation der Stationsärzte, die mit SonoConsult erstellte Befunde in der Betreuung von Patienten verwendeten (n=37, Fragebogen nicht von allen vollständig ausgefüllt).

AiP = Arzt im Praktikum

US+ = eigene Sonographiekenntnisse

AiW = Arzt in Weiterbildung zum Facharzt

US- = ohne eigene Sonographiekenntnisse

FA = Facharzt

C1, C2... = Fragekürzel

Die Umstellung der Sonographiebefundung wurde allgemein begrüßt (C1), auch wenn sich der Stellenwert der Untersuchung nicht wesentlich veränderte (C2). Die Ergebnisse zeigen, dass die Beurteilung des Untersuchers die wesentliche Informationsquelle für den Stationsarzt darstellt (C9). Einige Ärzte gaben an, allein diese Mitteilung zur Kenntnis zu nehmen und Befunde zu einzelnen Organen sowie von SonoConsult generierte Diagnosen zu ignorieren (C8, C10, C11). Differenzierte Angaben in der Befundbeschreibung einzelner Organe werden mit fortschreitendem Ausbildungsstand und bei eigenen Sonographiekenntnissen häufiger gelesen (C6, C7). Die vereinheitlichte Nomenklatur mit resultierend besserer Vergleichbarkeit der Befunde verschiedener Untersucher (C5) wurde vor allem von jüngeren Ärzten geschätzt (C4).



### 3.1.3. Subjektive Einschätzung des Ausbildungseffekts bei den Untersuchern

In der Ausbildung für die Sonographie können unterschiedliche Konzepte zur Anwendung kommen (Tabelle 8a, D 1 bis D4). Untersucher ohne eigene praktische Erfahrung schätzen die Wertigkeit der einzelnen Möglichkeiten mit einer klaren Präferenz für einen Tutor ein (D3).

D1	Fachbücher	4,0	1=nicht	5=sehr
D2	Kurse (z. B. DEGUM)	4,0	1=nicht	5=sehr
D3	Betreuung durch Tutor	5,0	1=nicht	5=sehr
D4	Befundsprogramme (z. B. SonoConsult)	2,8	1=nicht	5=sehr

Tabelle 8a: Einschätzung der Eignung verschiedener Konzepte für die Ausbildung durch Untersucher ohne praktische Sonographiekennntnisse (n = 8)

Der Beitrag einzelner Faktoren von SonoConsult in der Ausbildung (D5 bis D12), der von der Gesamtheit der Anwender erwartet wurde (A8 in Tabelle 5), wird in Tabelle 8b dargestellt.

D5	Standardisierte Fragenfolge	3,0	1=nicht	5=sehr
D6	Erklärungsfunktion	2,7	1=nicht	5=sehr
D7	Diagnosenanzeige	2,5	1=nicht	5=sehr
D8	Abrufbare Herleitung der Diagnosen	3,0	1=nicht	5=sehr
D9	Erlernen standardisierter Untersuchungsgang	3,9	1=nicht	5=sehr
D10	Differenzierung pathologischer Befunde	2,9	1=nicht	5=sehr
D11	Einarbeitung in sonographische Nomenklatur	3,6	1=nicht	5=sehr
D12	Vollständigkeit der Befunderhebung	4,0	1=nicht	5=sehr

Tabelle 8b: Beurteilung von einzelnen Funktionen von SonoConsult in der Ausbildung durch Untersucher ohne praktische Sonographiekennntnisse (n = 8)

Die Einarbeitung in die Nomenklatur (D11), die standardisierte und strukturierte Dokumentation sonographischer Befunde (D5) und die Erklärungsfunktion (D6, D8) wurden als hilfreich in der Lernphase beurteilt. Der am stärksten hervorgehobene positive Effekt des Programmes bestand in der strukturierten und vollständigen Führung durch den Untersuchungsgang (D9, D12) in einer standardisierten und sinnvollen Sequenz. Die diagnostischen Fähigkeiten von SonoConsult (D7) waren in der Lernphase von untergeordneter Bedeutung.

### 3.1.4. Qualität der Befundberichte

#### 3.1.4.1. Evaluation der Vollständigkeit von Untersuchungsbefunden

Die Eingabe von Daten aus 103 Freitextbefunden in die von SonoConsult präsentierten Formulare deckte bei 542 Eingabefeldern Informationslücken auf, von denen fast ein Drittel (170) als relevant eingestuft wurden. Tabelle 9 schlüsselt die fehlenden Angaben für die einzelnen Organe auf.

Information zu	essentiell unbekannt	erläßlich unbekannt
Leber	27	56
DHC	1	41
Gallenblase	9	7
Pfortader	13	26
Milz	3	30
Pankreas	12	31
Nieren	36	71
Nebennieren	1	
Harnblase	5	13
Prostata/Gynäkologische Organe	4	9
Aorta/V. cava	28	39
Lymphknoten	8	21
Spezielle Befunde (Pleura, Aszites etc.)	20	26
Darm	3	2

Tabelle 9: Evaluation von 103 Freitextbefunden erfahrener Untersucher, deren Angaben für eine Dokumentation mit SonoConsult benutzt wurden. Informationen, die vom Programm erfragt wurden und nicht in den Freitextbefunden vermerkt waren, wurden als „unbekannt“ erfasst und von zwei Experten nach ihrer Relevanz in die Kategorien „essentiell“ und „erläßlich“ eingeordnet.

Während von den Untersuchern in ihren Freitextbefunden insgesamt 315 Diagnosen angegeben wurden, konnte SonoConsult aus den gleichen Daten auf 536 Diagnosen schließen, 325 davon im Status „wahrscheinlich“. Einige Diagnosen wurden in den Freitextbefunden differenzierter beschrieben. Durch Einzelanalyse dieser Fälle konnten Informationslücken bei der Eingabe als ursächlich festgestellt werden, ein Fehler des Programms selbst wurde nicht gefunden. Sechs

Systemdiagnosen, die von den Evaluatoren als fehlerhaft klassifiziert wurden, stellten sich in keinem Fall als gravierend heraus. Dagegen konnte mehreren fehlenden diagnostischen Folgerungen in den Freitextbefunden eine klinische Relevanz zugeschrieben werden. Im Vergleich mit SonoConsult wiesen die Freitextbefunde weiterhin 42 Inkongruenzen zwischen Befundbeschreibung und Beurteilung auf (z. B. verschiedene Seitenangaben in Befund und Diagnose).

	Freitextbefund	SC-Befund
Diagnosen insgesamt	315	536
Diagnosen, genauer als im Freitext beschrieben		18
Diagnosen, genauer als in SC beschrieben	61	
Fehlende Diagnose, nur im Freitext beschrieben		55*
Fehlende Diagnose, nur in SC beschrieben	124**	
Fehldiagnose	20	6

Tabelle 10: Diagnosen aus 103 Freitextbefunden verglichen mit SonoConsult-Diagnosen aus den gleichen Patientendaten.

SC = SonoConsult

\* 25 davon nicht ermittelt wegen Informationslücken oder fehlender Eingabemöglichkeit in strukturierte SC-Formulare. Die alternative Eingabe in Freitextfelder im Programm führt nicht zu diagnostischen Folgerungen.

\*\* 35 davon einfache Diagnose (z. B. Nierenzysten), die trotz erfasstem Befund nicht als Diagnose angegeben wurde

### 3.1.4.2. Evaluation der Diagnosequalität

Die Beurteilung der Diagnosen, die SonoConsult aus den Daten der Freitextbefunden generierte (A), wurde mit einer Serie von Fällen verglichen, die unmittelbar im Programm dokumentiert wurden (B). Eine dritte Reihe umfasst Fälle, bei deren Befundeingabe und Beurteilung die diagnostischen Folgerungen von SonoConsult für den Untersucher nicht sichtbar waren (C). Aufgrund der unterschiedlichen Größe der drei Fallserien wurden die Diagnosen in Tabelle 11 auf die Fallzahl bezogen.

		A (n=103)	B (n=130)	C (n=150)
1	Evaluatoren-Diagnosen	5,49	5,39	5,24
2	Untersucher-Diagnosen	3,06	3,83	3,37
3	SC- Diagnosen	5,21	5,10	4,62
3a	SC- Diagnosen mit Status wahrscheinlich	3,16	3,47	3,27
3b	SC- Diagnosen mit Status verdächtig	2,05	1,63	1,35
4	Fehldiagnosen des Untersuchers	0,19	0,02	0,05
5	Untersucher-Diagnosen wie SC, aber genauer	0,59	0,86	0,61
6	Untersucher-Diagnosen, nicht von SC angegeben	0,53	0,35	0,69
7	Untersucher-Diagnosen fehlen, obwohl im Befund beschrieben	0,34	0,09	0,27
8	Inkonsistenzen im Befund des Untersuchers	0,41	0,02	0,13
9	Diagnosen bei Untersucher und SC gleich	2,05	3,43	2,85
10	SC-Diagnose richtig, aber von Untersucher nicht angegeben	1,20	0,86	0,99
11	SC- Diagnosen wie Untersucher, aber genauer	0,17	0,46	0,42
12	SC-Diagnosen fehlend, da Befundeingabe in Freitextfelder in Freitextfelder	0,24	0,18	0,27
13	SC Fehldiagnosen	0,06	0,08	0,02

Tabelle 11: Vergleich der Diagnosen in drei Fallserien

A = Dateneingabe aus Freitextbefunden

B = unmittelbare Dokumentation mit SonoConsult mit Anzeige der Programmdiagnosen

C = unmittelbare Dokumentation mit SonoConsult ohne Anzeige der Programmdiagnosen

SC = SonoConsult

Mit Einführung von SonoConsult konnte die Qualität der Untersuchungsbefunde verbessert werden (4 und 8). Programminterne Fehldiagnosen aus den eingegebenen Daten sind sehr selten und erwiesen sich in der Einzelanalyse in keinem der Fälle als gravierend (13).

Die Vollständigkeit der Diagnosenangaben in den einzelnen Fallserien zeigt Tabelle 12, aufgrund der unterschiedlichen Größe der Serien bezogen auf die Fallzahl (Zeilen 14 bis 16 in Tabelle 12). Die Zahl der Diagnosen, welche die zwei Evaluatoren aus der Befundeingabe ermittelten und die in der Evaluation als Goldstandard galt (1 in Tabelle 11), gilt als Bezugsgröße in den Zeilen 17 bis 19 der Tabelle 12.

		A (n=103)	B (n=130)	C (n=150)
14	Richtige Diagnosen Untersucher (2-4)	2,87	3,81	3,32
15	Richtige Diagnosen SC (3-13)	5,15	5,02	4,60
16	Richtig mögliche Diagnosen SC (15+12)	5,39	5,30	4,87
17	Richtige Diagnosen Untersucher (14:1)	52%	71%	63%
18	Richtige Diagnosen SC (15:1)	94%	93%	88%
19	Richtige mögliche Diagnosen SC (16:1)	98%	98%	93%

Tabelle 12: Richtige Diagnosen im Vergleich von drei Fallserien (Erläuterung in Tabelle 11). Die Angabe in Klammern erläutert die Berechnung aus Zeilen der Tabellen 11 und 12.

A = Dateneingabe auf Freitextbefunden

B = unmittelbare Dokumentation mit SonoConsult mit Anzeige der Programmdiagnosen

C = unmittelbare Dokumentation mit SonoConsult ohne Anzeige der Programmdiagnosen

SC = SonoConsult

Der Einsatz von SonoConsult erhöhte die Vollständigkeit der korrekt angegebenen Diagnosen in den Befundberichten im Vergleich zu Freitextbefunden. Die Anzeige der von SonoConsult vorgeschlagenen Diagnosen spielt dabei offenbar eine Rolle, denn bei ihrer Ausblendung blieb die Zahl der Diagnosen pro Fall geringer (17). Insbesondere einfache Diagnosen, die unmittelbar aus eingegebenen Daten hergeleitet werden können (wie zum Beispiel Nierenzysten, Zustand nach Cholezystektomie, Aortenektasie, Harnstau), fehlten seltener in der Zusammenfassung der Untersuchung (7). Dies entsprach der Erwartung der teilnehmenden Untersucher, wie Tabelle 13 zeigt.

Die Ausblendung der Programmdiagnosen wird führen zu			
Abnahme der Diagnosen insgesamt	2,0	1=keinesfalls	5=in jedem Fall
Abnahme von Diagnosen mit klinischer Relevanz	1,0	1=keinesfalls	5=in jedem Fall
Abnahme von Diagnosen ohne klinische Relevanz	3,7	1=keinesfalls	5=in jedem Fall

Tabelle 13: Erwartungen der Untersucher (n = 3) bezüglich der Vollständigkeit ihrer eigenen Diagnosen bei Ausblendung der SonoConsult-Diagnosen in der Eingabemaske

Entsprechend der Evaluation muss damit gerechnet werden, dass in einem Befundbericht nur etwa drei von vier Diagnosen mitgeteilt werden (hier 71%). Die diagnostischen Folgerungen von SonoConsult erreichen dagegen eine höhere Vollständigkeit (18 in Tabelle 12). Diese wäre durch die konsequente Nutzung der standardisierten Formulare noch zu verbessern (19 in Tabelle 12), da die Dokumentation in Freitextfeldern von SonoConsult nicht für diagnostische Folgerungen des Programms verwertet werden kann.

Eine detaillierte Befragung der Untersucher zum Nutzen einer Präsentation von Diagnosevorschlägen durch das Programm in der täglichen Praxis zeigt Tabelle 14.

Genereller Nutzen der Funktion	4,0	1=überflüssig	5=sinnvoll
Hilfe bei Überprüfung der Vollständigkeit	4,0	1=nie	5=regelmäßig
Hilfe bei Diskussion von Differentialdiagnosen	1,7	1=nie	5=regelmäßig
Nutzung der automatischen Übernahme in eigene Beurteilung	1,4	1=nie	5=regelmäßig
Nutzen vor allem für jüngere Kollegen	3,7	1=nicht	5=sicher
Berücksichtigung der SonoConsult-Diagnosen bei eigener Beurteilung	3,9	1=nie	5=immer

Tabelle 14: Nutzung der Diagnosefunktion von SonoConsult bei erfahrenen Untersuchern (n = 7)

Die subjektive Einschätzung der Untersucher stellte die Unterstützung beim Verfassen eines vollständigen Befundberichtes positiv heraus (Tabelle 14). Die objektive Analyse der Fallserien (Tabellen 11 und 12) deckte jedoch deutliche Defizite in dieser Hinsicht auf. Die automatische Übernahme der Programmdiagnosen in die eigene Beurteilung könnte dem entgegenwirken, wird aber von den Untersuchern offenbar kaum genutzt. In frei ergänzbaren Bemerkungen auf den Umfragebögen wurde angegeben, dass

- die Beschreibung nie exakt dem vom Untersucher präferierten Wortlaut entspricht (2x),
- die Diagnosen keine Wertungen beinhalten (1x),
- die Häufigkeit einer Diagnose in den Schlussfolgerungen des Programmes nicht berücksichtigt wird (1x).

### 3.1.5. Diagnostische Genauigkeit

In 112 konsekutiven Falldokumentationen fanden sich insgesamt 412 von SonoConsult hergeleitete und 332 von den Untersuchern angegebene Diagnosen. Von den Untersuchern wurden 27 Diagnosen angegeben, die von SonoConsult nicht hergeleitet wurden (Tabelle 15). Grund war in allen Fällen eine Lücke oder ein Fehler bei der Befundeingabe. Von SonoConsult wurde 107 mal Diagnosen angegeben, die der Untersucher nicht erwähnt und deren zugrunde liegende Befunde auch nicht anders interpretiert hat. Darunter waren 32 mal die Diagnose einer Organvergrößerung oder –verkleinerung (wie Hepatomegalie, Prostatavergrößerung oder Nierenverkleinerung) und vereinzelt andere Diagnosen wie Prostatitis, Cholezystitis, Nebenniere oder Blasenentleerungsstörung. Die Übereinstimmung zwischen SonoConsult und Untersuchern bezüglich aller SonoConsult-Diagnosen lag bei 74%.

	SonoConsult	Untersucher
Angegebene Diagnosen	412	332
Von SC angegebene Diagnosen, die aus den beschriebenen Befunden hervorgingen, aber vom Untersucher in seiner Beurteilung nicht erwähnt wurden	107	
Vom Untersucher angegebene Diagnosen, die von SC nicht angegeben wurden		27
Übereinstimmung von ... mit ... (%)	74	92

Tabelle 15: Auswertung der Diagnosen von 112 konsekutiven sonographischen Untersuchungen auf fehlende Diagnoseerwähnung in der Befundbeurteilung  
SC = SonoConsult

Unter den von SonoConsult präsentierten Diagnosevorschlägen fanden sich insgesamt 145 komplex herzuleitende Diagnosen folgender Spezifizierung:

14 Fälle mit Leberzirrhose,

21 Fälle mit Fettleber,

21 Fälle mit Stauungsleber,

5 Fälle mit Cholestase,

10 Fälle mit gut- oder bösartigen Leberherden,

8 Fälle mit portaler Hypertension,

44 Fälle mit chronisch degenerativer Nierenerkrankung,

22 Fälle mit Pankreaserkrankungen (akute und chronische Pankreatitis, Tumor, Zyste)

Die Aufschlüsselung der Übereinstimmung zwischen den durch das Programm generierten Vorschlägen komplex herzleitender Diagnosen und denen der Untersucher wird in Tabelle 16 gezeigt.

Diagnoseangabe	SonoConsult richtig, Untersucher richtig	SonoConsult richtig, Untersucher nicht	Untersucher richtig, SonoConsult nicht
Leberzirrhose	7	7	0
Fettleber	15	6	0
Stauungsleber	15	6	0
Portale Hypertension	7	1	0
Cholestase, obstruktiv	3	2	0
Leberherd	10	0	0
Chronisch degenerative Nierenerkrankung	15	29	0
Pankreaserkrankung	16	6	0
Gesamt	88	57	0

Tabelle 16: Auswertung der komplex herzleitenden Diagnosen von SonoConsult und vom Untersucher auf der Basis der Befundberichte von 112 Fällen. „Richtig“ bedeutet: Herleitbar aus der Befundeingabe. Eine fehlende Übereinstimmung in Diagnosen beruht zum einen Teil auf einer Befund-Diagnose-Inkonsistenz, zum anderen Teil auf einer fehlenden Angabe des Untersuchers bezüglich einer aus den Befunden ableitbaren Diagnose.

Alle von SonoConsult genannten komplexen Diagnosen (SC-Diagnosen, n = 145) wurden nach Datenlage als richtig eingeordnet; von ihnen sind 88 auch vom Untersucher genannt worden. Die Übereinstimmung ist ca. 61 %. Von den 57 SC-Diagnosen, die vom Untersucher diskrepant bewertet wurden, waren 42 nicht erwähnt und 15 zu den eingegebenen Daten inkonsistent. Der Anteil fehlender Diagnosen des Untersuchers im Vergleich zu SonoConsult-Diagnosen beträgt somit 39%.

In 12 Fällen komplexer Diagnosen wurden Eingabefehler und in weiteren 8 Fällen wesentliche Eingabelücken registriert. In 2 Fällen wurden durch Eingabefehler Diagnosen generiert, die wesentlich von den Diagnosen des Untersuchers abwichen. Eingabefehler kamen meistens durch offensichtliches Verklicken (z.B. „rechts“ statt „links“) (6 mal) oder dadurch zustande, dass Befunde in Freitextfelder statt in die präsentierten Formulare eingetragen wurden (4 mal).



Die Abweichungen zwischen SonoConsult und der Beurteilung des Untersuchers sind bezüglich der klinischen Richtigkeit nicht zu beurteilen, da die Evaluation anhand der Befundeingabe, jedoch nicht mit sonographischem Bildmaterial durchgeführt wurde. Aus diesem Grunde ist grundsätzlich nicht zwischen inkorrektter Symptombeschreibung und falscher diagnostischer Folgerung zu unterscheiden, obwohl der hohe Anteil fehlender Diagnosen einfacher Herleitung (50 von 267; 19%) auf Fehler in der Dokumentation hinweist.

### 3.1.6. Konkordanz von Befund und Beurteilung

**In der ersten Fallserie** von 250 Fällen wurden von den Untersuchern 934 Diagnosen (3,7 Diagnosen je Fall) gestellt; in 71,1% stimmten sie mit den Diagnosevorschlägen von SonoConsult überein und waren konsistent mit den dokumentierten Befunden (Tabelle 17). 24,4% der Diagnosen, die aus den Eingabedaten hätten gefolgert werden können, wurden von den Untersuchern nicht erwähnt (falsch negativ), jedoch von SonoConsult angegeben, während 4,5% der von den Untersuchern beschriebenen Diagnosen nicht durch die dokumentierten Befunde gestützt wurden (falsch positiv). Diese Diagnosen konnten von SonoConsult nicht abgeleitet werden. 40,4% der falsch negativen Diagnosen (etwa 10% aller Diagnosen) wurden von den Evaluatoren als medizinisch relevant eingestuft, so zum Beispiel chronische Pankreatitis, Pleuraergüsse, Gallensteine, Polypen der Gallenblase, verdickte Harnblasenwand oder vergrößerte intraabdominelle Lymphknoten. Den verbleibenden 59,6% wurde von den Evaluatoren eine geringe klinische Relevanz beigemessen.

	n	pro Fall	%	Weitere Angabe
Diagnosen gesamt	934	3,7	100	
Untersucherdiagnosen konsistent zu Befunden	664	2,7	71,1	
Falsch positiv (zusätzliche Diagnosen)	42	0,2	4,5	
Falsch negativ (nicht angegebene Diagnosen)	228	0,9	24,4	
- höhere Relevanz	92	0,4	9,9	40,4% von 228
- geringere Relevanz	136	0,5	14,6	59,6% von 228

Tabelle 17: Evaluation von 250 Fällen bezüglich der Übereinstimmung von dokumentierten Befunden und Diagnosen des Untersuchers

**Die zweite Fallserie** mit 161 Fällen (3,1 Diagnosen pro Fall) zeigte eine Konkordanz von 61,1% der Untersucher-Diagnosen bezogen auf die dokumentierten Befunde. 36,1% der Diagnosen

wurden als falsch negativ, 2,8 % als falsch positiv bewertet (Tabelle 18). Die Differenzierung nach der Komplexität der Herleitung aus den Befunden ist in Tabelle 19 dargestellt und weist einen höheren Anteil falsch negativer Diagnosen für Diagnosen komplexer Herleitung auf.

	n	pro Fall	%
Diagnosen gesamt	501	3,1	100
Untersucherdiagnosen konsistent zu Befunden	306	1,9	61,1
Exakte Übereinstimmung SC und Untersucher	225	1,4	44,9
Falsch positiv (zusätzliche Diagnosen)	14	0,1	2,8
Falsch negativ (nicht angegebene Diagnosen)	181	1,1	36,1

Tabelle 18: Evaluation von 161 Fällen bezüglich der Übereinstimmung von dokumentierten Befunden und Diagnosen des Untersuchers  
SC = SonoConsult

	Komplexe Diagnosen	%	Einfache Diagnosen	%
Diagnosen gesamt	152	100	349	100
Untersucherdiagnosen konsistent zu Befunden	80	52,6	226	64,7
Falsch positiv (zusätzliche Diagnosen)	5	3,3	9	2,6
Falsch negativ (nicht angegebene Diagnosen)	67	44,1	114	32,7

Tabelle 19: Differenzierung der Diagnosen von 161 Fällen nach einfacher (Folgerung aus einer Befundeingabe) und komplexer (Folgerung aus einer Kombination von Befunden) Herleitung

## 3.2. Statistische Auswertungen

### 3.2.1. Inter-Untersucher-Variation

4100 Fälle von 7 Untersuchern standen in einer Fallsammlung zur Verfügung. In einem ersten Schritt wurde die Verteilung aller Diagnosen über die Untersucher analysiert. Zur Aufdeckung eines möglichen Lerneffektes wurde eine weitere Aufschlüsselung in die ersten, zweiten und dritten 200 Fälle jedes Untersuchers vorgenommen. Es ließen sich für mehrere Diagnosen unterschiedliche Häufigkeiten zwischen den Untersuchern feststellen, die am Beispiel zweier komplex herzleitender Diagnosen („Leberzirrhose“ und „Chronisch degenerative Nierenerkrankung“) dargestellt werden (Tabelle 20).

	E1			E2			E3			E4			E6			E7			E8			all			1400	1400	1300					
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III					
CRF I	73	26	18	29	60	21	20	19	26	12	9	5	58	7	23	28	11	6	2	3	51	15	10	26	51	16	15	20	330	103	97	130
CRF ?	174	82	48	44	91	27	28	36	23	15	2	6	81	24	38	19	10	8	0	2	13	3	7	3	84	37	28	19	476	196	151	129
CRF	247	108	66	73	151	48	48	55	49	27	11	11	139	31	61	47	21	14	2	5	64	18	17	29	135	53	43	39	806	299	248	259
LC I	27	13	12	2	23	12	7	4	25	7	10	8	32	12	11	9	24	10	11	3	38	7	13	18	12	3	9	0	181	64	73	44
LC ?	15	3	3	9	20	4	2	14	107	28	40	39	20	3	7	10	17	4	12	1	4	1	0	3	20	3	5	12	203	46	69	88
LC	42	16	15	11	43	16	9	18	132	35	50	47	52	15	18	19	41	14	23	4	42	8	13	21	32	6	14	12	384	110	142	132

Tabelle 20: Diagnosehäufigkeit von zwei Diagnosen: Chronisch degenerative Nierenerkrankung (chronic renal failure = CRF) und Leberzirrhose (liver cirrhosis = LC), unterteilt nach SonoConsult- Status „wahrscheinlich“ (!) und „möglich“ (?) für 7 Untersucher (E1, E2, E3, E4, E6, E7, E8). differenziert nach den jeweils ersten, zweiten und dritten 200 Fällen („I“, „II“, „III“). Untersucher E2 nur 574 und E6 nur 526 Fälle. Die markierten Felder zeigen besonders starke Abweichungen vom Durchschnitt.

### 3.2.1.1. Entwicklung der Diagnosehäufigkeiten bei einzelnen Untersuchern

Einzelne Untersucher zeigten während ihrer Einarbeitungszeit eine Entwicklung. Beispielsweise hat E1 in den ersten 200 Fällen 13 mal Befunde so eingegeben, dass SonoConsult daraus eine Leberzirrhose als „wahrscheinlich“ ableitete. In den zweiten 200 Fällen wurde die Leberzirrhose 12 mal und in den dritten 200 Fällen 2 mal als „wahrscheinlich“ gefolgert. Umgekehrt wurden in den ersten 200 Fällen die Befunde so eingegeben, dass SonoConsult eine Leberzirrhose in drei Fällen als „möglich“ einstufte, in den nächsten 200 Fällen ebenfalls 3 mal, in den dritten 200 Fällen dagegen 9 mal. Es deutet sich damit eine Verschiebung der Befundinterpretationen von „wahrscheinlich“ (weniger) zu „möglich“ (mehr) an. Ein ähnlicher Trend ist weniger ausgeprägt bei E2, E4 und E8 zu erkennen, nicht dagegen bei E3, E6 und E7.

### 3.2.1.2. Befundungsmuster verschiedener Untersucher

Auffällig ist es, dass ein Untersucher (E3) seine Untersuchungsbefunde in einer Weise dokumentierte, dass SonoConsult eine „Leberzirrhose“ deutlich häufiger als Diagnose generierte als bei allen anderen Untersuchern. In ähnlicher Weise stechen die Untersucher E1 und E6 bezüglich der Diagnose „Chronisch degenerative Nierenerkrankung“ mit über- und unterdurchschnittlicher Häufigkeit hervor.

Die Abweichungen zwischen verschiedenen Untersuchern wurden weiter differenziert. Die folgenden Tabellen zeigen die Ergebnisse für einige der markierten Felder in Tabelle 20: Untersucher E1 und E6 für „Chronisch degenerative Nierenerkrankung“ (CRF, Tabelle 21) und Untersucher E3, E7 und E8 für „Leberzirrhose“ (LC, Tabelle 22).

#	A1		A2		A3		A4					A5				Parameters								
	V1	V2	V3	V4	V5	V7	V9	V10	V11	V12	V13	V20	V21	V22	V23	V24	Size	Pop.	p0	p	Rel. Gain	Sig.	TP	FP
1	X	X															806	3707	0,15	0,31	1,22	+++	247	559
2			X														505	3332	0,16	0,31	1,12	+++	159	346
3				X													98	3332	0,16	0,01	-1,12	+++	1	97
4					X												578	3369	0,16	0,3	0,99	+++	173	405
5						X											130	3369	0,16	0,03	-0,97	+++	4	126
6		X		X													276	3249	0,16	0,31	1,04	+++	85	191
7							X	X									411	3621	0,15	0,18	0,26	0,05	76	335
8									X	X	X						407	3621	0,15	0,13	-0,15	0,25	54	353
9												X	X				404	3673	0,15	0,19	0,27	0,04	75	329
10														X	X	X	349	3673	0,15	0,15	-0,04	0,77	51	298
11		X					X	X									208	3313	0,16	0,22	0,43	0,02	46	162
12				X								X	X				213	3361	0,16	0,23	0,51	+++	50	163
13	X			X													328	3369	0,16	0,37	1,47	+++	120	208
14	X	X															271	3332	0,16	0,38	1,6	+++	103	168

E6		A1			A2		A3		A4				A5				Parameters							
#	V1	V2	V3	V4	V5	V7	V9	V10	V11	V12	V13	V20	V21	V22	V23	V24	Size	Pop.	p0	p	Rel. Gain	Sig.	TP	FP
1	X	X															806	3707	0,13	0,03	-0,92	+++	21	785
2			X														505	3332	0,14	0,03	-0,9	+++	16	489
3				X													98	3332	0,14	0,09	-0,39	0,17	9	89
4					X												578	3369	0,14	0,02	-1,03	+++	9	569
5						X											130	3369	0,14	0,08	-0,52	0,04	10	120
6		X		X													276	3249	0,14	0,01	-1,04	+++	4	272
7							X	X									411	3621	0,13	0,1	-0,22	0,12	43	368
8									X	X	X						407	3621	0,13	0,09	-0,36	0,01	36	371
9												X	X				404	3673	0,13	0,06	-0,62	+++	24	380
10														X	X	X	349	3673	0,13	0,13	0,06	0,72	47	302
11		X					X	X									208	3313	0,14	0,04	-0,84	+++	8	200
12					X							X	X				213	3361	0,14	0,03	-0,93	+++	6	207
13	X			X													328	3369	0,14	0,01	-1,08	+++	3	325
14	X	X															271	3332	0,14	0,03	-0,95	+++	7	264

Tabelle 21: Genauere Analyse der Untersucher E1 (oben) und E6 (unten) als Zielvariable für die Diagnose CRF (Chronisch degenerative Nierenerkrankung) durch Subgruppenanalyse in einem Subgruppentableau.

Folgende Attribute wurden in der Subgruppenanalyse verwendet:

A1 = CRF mit den Werten V1 = möglich, V2 = wahrscheinlich

A2 = linke Niere: Nierenparenchym mit den Werten V3 = verschmälert und V4 = verbreitert

A3 = rechte Niere: Nierenparenchym mit den Werten V5 = verschmälert und V7 = verbreitert

A4 = linke Nierengröße mit den Werten V9 = stark verkleinert, V10 = moderat verkleinert,

V11 = leicht vergrößert, V12 = moderat vergrößert und V13 = stark vergrößert

A5 = rechte Nierengröße mit den Werten V20 = stark verkleinert, V21 = moderat verkleinert,

V22 = leicht vergrößert, V23 = moderat vergrößert und V24 = stark vergrößert

Jede Zeile des Tableaus beschreibt eine Subgruppe mit ihren Parametern „Size“ (Subgruppengröße), „Pop.“ (Population ohne Fälle, in denen mindestens eine Angabe zu einer Subgruppenvariable fehlt), der durch die Zielvariablen charakterisierte Anteil in der Population „p0“ und in der Subgruppe „p“ bzw. „Rel. Gain“ (Relative Zunahme), „Sig“ (Signifikanz) mit dem Wert „+++“ für sehr hohe Signifikanz ( $< 10^6$ ), „TP“ (richtig positiv, Zielvariable positiv in der Subgruppe), „FP“ (falsch positiv, Zielvariable negativ in der Subgruppe). Folgende Beziehungen unter diesen Parametern gelten:  $TP+FP=Size$ ;  $TP/Size=p$ ;  $(p-p0)/(p0*(1-p0))=Rel. Gain$ . Die Signifikanz wurde mit dem Chi-Quadrat-Test für Unabhängigkeit unter Bildung einer Vierfeldertafel ermittelt (FN und TN als Komplementärgruppe, dabei  $FN=Pop*p0-TP$  und  $TN=Pop*(1-p0)-FP$ ).

Zum Beispiel bedeutet Zeile #6 im oberen Tableau: Die Subgruppe „A2=V3“ und „A3=V5“ (Nierenparenchym links und rechts verschmälert) wurde in 31% der Fälle von E1 gewählt ( $p=0,31$ ), dies ist mehr als doppelt so häufig (Rel. Gain=1,04) als der Anteil seiner Fälle an der Gesamtpopulation ( $p0=16\%$ ). Obwohl die Fallsammlung 4100 Fälle enthält, war die relevante Population für diese Subgruppe 3249, da in 851 Fällen die Angabe zum Nierenparenchym auf wenigstens einer Seite fehlte. Die Zunahme ist hochsignifikant (+++).

Die Tabelle 21 zeigte für Untersucher E1 und E6 signifikante Unterschiede im Befundungsmuster. Untersucher E1 diagnostizierte in seiner Fallserie eine „Chronisch degenerative Nierenerkrankung“ als „möglich“ oder „wahrscheinlich“ mehr als zehnmals häufiger wie Untersucher E6 und zweimal so häufiger wie im Mittel aller Untersucher (Tabelle 20).

Die detaillierte Analyse von E1 (oberes Tableau in Tabelle 21) deckte auf, dass nur ein

Parameter für diese Abweichung verantwortlich ist: Eine „Parenchymverschmälerung rechts oder links“ wurde sehr viel häufiger dokumentiert, während die „Nierengröße“ nicht signifikant im Vergleich zur Gesamtheit differierte ( $p > 0,01$ ). Eine „Parenchymverbreiterung“ wurde andererseits von E1 signifikant seltener als im Durchschnitt angegeben, de facto wurde dieser Wert fast nie festgestellt. Die detaillierte Analyse von E6 (unteres Tableau in Tabelle 21) deckte auf, dass E6 sowohl ein „verschmälertes Nierenparenchym“ als auch eine „Nierenverkleinerung“ extrem selten in seiner Fallserie vermerkte und konsekutiv die auf diesen beiden Angaben basierende Diagnose „Chronisch degenerative Nierenerkrankung“ von SonoConsult nicht generiert werden konnte.

E3																		
#	A1		A2		A3		A4		A5		Parameters							
	V0	V1	V2	V3	V10	V11	V20	V21	V25	V27	Size	Pop.	p0	p	Rel. Gain	Sig.	TP	FP
1	X	X									384	4100	0,15	0,34	1,58	+++	132	252
2		X									203	4100	0,15	0,53	3,05	+++	107	96
3					X						127	2163	0,14	0,02	-0,96	+++	3	124
4						X					645	2163	0,14	0,23	0,77	+++	147	498
5			X								962	2164	0,14	0,15	0,08	0,24	141	821
6				X							421	2164	0,14	0,08	-0,49	+++	33	388
7									X		215	2164	0,14	0,06	-0,69	+++	12	203
8							X				161	2164	0,14	0,72	4,94	+++	116	45
9								X			163	2164	0,14	0,1	-0,33	0,14	16	147
10										X	20	2164	0,14	1	7,31	+++	20	0
11		X					X				118	2164	0,14	0,77	5,37	+++	91	27
12					X	X					104	2163	0,14	0,9	6,49	+++	94	10
13		X			X	X					98	2163	0,14	0,92	6,62	+++	90	8

E7																		
#	A1		A2		A3		A4		A5		Parameters							
	V0	V1	V2	V3	V10	V11	V20	V21	V25	Size	Pop.	p0	p	Rel. Gain	Sig.	TP	FP	
1	X										181	4100	0,15	0,21	0,51	0,013	38	143
2		X									203	4100	0,15	0,02	-1,01	+++	4	199
3					X						127	2163	0,14	0,2	0,5	0,042	26	101
4						X					645	2163	0,14	0,18	0,31	0,001	117	528
5			X								962	2164	0,14	0,19	0,34	+++	178	784
6				X							421	2164	0,14	0,15	0,07	0,567	64	357
7									X		215	2164	0,14	0,18	0,27	0,14	38	177
8							X				161	2164	0,14	0	-1,17	+++	0	161
9								X			163	2164	0,14	0,21	0,53	0,013	34	129
10							X	X			324	2164	0,14	0,1	-0,31	0,033	34	290
11	X							X			149	2164	0,14	0,22	0,64	0,005	33	116
12		X						X			10	2164	0,14	0,1	-0,36	0,399	1	9

E8																	
#	A1		A2		A3		A4		A5	Parameters							
	V0	V1	V2	V3	V10	V11	V20	V21	V25	Size	Pop.	p0	p	Rel. Gain	Sig.	TP	FP
1	X									181	4100	0,15	0,07	-0,64	0,002	12	169
2		X								203	4100	0,15	0,1	-0,38	0,048	20	183
3					X					127	2163	0,15	0,05	-0,8	0,001	6	121
4						X				645	2163	0,15	0,07	-0,62	+++	45	600
5			X							962	2164	0,15	0,13	-0,17	0,014	122	840
6				X						421	2164	0,15	0,12	-0,23	0,061	50	371
7								X		215	2164	0,15	0,09	-0,44	0,017	20	195
8							X			161	2164	0,15	0,16	0,11	0,613	26	135
9								X		163	2164	0,15	0,06	-0,69	0,001	10	153
10				X	X					664	2163	0,15	0,07	-0,62	+++	46	618
11					X		X			138	2163	0,15	0,07	-0,66	0,005	9	129
12	X					X	X			137	2163	0,15	0,07	-0,65	0,005	9	128

Tabelle 22: Ergebnisse der Subgruppen-Analyse für Untersucher (Zielvariable) E3 (oben), E7 (Mitte) und E8 (unten) bezüglich Leberzirrhose (vgl. Legende Tab. 18). Die folgenden Attribute waren relevant:

A1 = Leberzirrhose mit den Werten V0 = wahrscheinlich, V1 = möglich

A2 = Leberechogenität mit den Werten V2 = moderat vermehrt, V3 = stark vermehrt

A3 = Lebergefäße mit den Werten V10 = Kalibersprung der Portalgefäße, V11 = Rarefizierung der Portalgefäße

A4 = Leberverformbarkeit mit den Werten V20 = mäßig vermindert, V21 = stark vermindert

A5 = Leberoberfläche mit den Werte V25 = gebuckelt/ knotig, V27 = wellig

Zum Beispiel bedeutet Zeile #13 von E3 (oben): Die Subgruppe "A1 = V1" und "A3 = V11" und "A4 = V20" (Leberzirrhose möglich und Rarefizierung der Portalgefäße und Leberverformbarkeit mäßig vermindert) trifft für E3 auf 92% seiner Fälle zu ( $p = 0.92$ ; also 90 von 98), dies ist mehr als 7-mal (relative Zunahme = 6.62) so häufig wie sein Anteil über alle Fälle ( $p_0 = 14\%$ ). Während die Gesamtpopulation 4100 Fälle umfasst, in 1937 Fällen aber wenigstens eines der in der Subgruppe relevanten Attribute fehlte, ist die Bezugspopulation für die Subgruppe  $4100 - 1937 = 2163$ . Der Anstieg ist hoch signifikant ( $+++ = < 10^{-6}$ ).

Untersucher E3 dokumentierte seine Befunde in der Frageklasse „Leber“ in der Weise, dass SonoConsult eine „Leberzirrhose“ viermal häufiger als „möglich“ oder „wahrscheinlich“ schlussfolgerte als E8 und doppelt so häufig wie im Mittel bei allen Untersuchern. Die Abweichung ist allein auf die häufigere Generierung der Diagnose im Status „möglich“ zurückzuführen (nicht „wahrscheinlich“). Eine genauere Betrachtung hinsichtlich dieser Diagnose zeigte, dass die Befundkombination „Rarefizierung der Portalgefäße“ und „Leberverformbarkeit moderat vermindert“ (Zeile 12 und 13 im oberen Tableau der Tabelle 22) die häufigere Angabe einer „Leberzirrhose“ bei E3 begründen, da E3 einen Anteil von 90% an dieser Befundkombination aufweist (bei 14% Anteil an allen Fällen). Die genannten Variablen generieren in Kombination nach den Regeln der Wissensbasis von SonoConsult die Diagnose „Leberzirrhose = möglich“. Zur weiteren Etablierung der Diagnose bis zum Status

„wahrscheinlich“ benötigt das Programm zusätzliche hinweisende Befunde (wie „Lebergefäße = Kalibersprung der Portalgefäße“ in Zeile 3, „Leberoberfläche = gebuckelt/knotig“ in Zeile 7), die E3 jedoch sehr selten dokumentierte. Dagegen nutzte er als einziger Untersucher die Antwortalternative „Leberoberfläche = wellig“ (Zeile 10).

Untersucher E7 dokumentierte seine Befunde zur Leber in der Weise, dass SonoConsult die Diagnose „Leberzirrhose“ fast ausschließlich als „wahrscheinlich“ herleitete, dagegen fast nie als „möglich“. Die Ursache liegt entsprechend der Subgruppenanalyse darin begründet, dass E7 nie „Leberverformbarkeit = moderat vermindert“ (Zeile 8 im mittleren Tableau der Tabelle 22) dokumentierte, jedoch signifikant häufiger als im Mittel aller Untersucher „Leberverformbarkeit = stark vermindert“ (Zeilen 9 und 11).

Untersucher E8 diagnostizierte eine „Leberzirrhose“ nur halb so oft wie die anderen Untersucher. Durch Subgruppenanalyse konnte dies auf die seltenere Angabe „Lebergefäße = Rarefizierung der Portalgefäße“ zurückgeführt werden (Zeile 4 im unteren Tableau der Tabelle 22). Die Abweichung wurde verstärkt durch den Befund „Lebergefäße = Kalibersprung der Portalgefäße“, der bei E8 nur ein Drittel der Häufigkeit bei allen Untersuchern erreicht, insgesamt aber deutlich seltener diagnostiziert wird als der vorgenannte Befund. Die Kombination beider Befunde schließlich (Zeile 11 und 12) korrelierte wiederum stark mit der Diagnose „Leberzirrhose“ und zeigte bei E8 somit ein faktisch entgegengesetztes Befundungsmuster im Vergleich zu Untersucher E3.



### 3.2.2. Subgruppen-Analyse zur Wissensentdeckung

Zur Darstellung medizinischer Fragestellungen, die mit Hilfe der Datenbank ausgewertet werden könne, wurden Subgruppen gebildet und untersucht, denen die Diagnose „Gallensteine“ (= Zielvariable) gemeinsam ist. Einige Ergebnisse des Analyseprozesses, der mit Hilfe von VIKAMINE durchgeführt wurde, zeigt Tabelle 23.

Target Variable: Gallstones																					
#	Age			Sex		Liver size						Aorta sclerosis									
	1	2	3	m	f	1	2	3	4	5	6	n	c	Size	TP	FP	Pop.	p0	p	RG	Bin. QF
1			X		X				X	X	X		X	89	37	52	3171	0.172	0.416	1.71	6.17
2			X		X				X	X	X	X	X	119	46	73	3171	0.172	0.387	1.5	6.31
3	X	X			X				X	X	X		X	132	51	81	3171	0.172	0.386	1.5	6.66
4					X		X	X	X	X			X	190	68	122	3177	0.172	0.358	1.3	6.99
5			X		X				X	X	X			207	72	135	3171	0.172	0.348	1.23	6.92
6	X	X			X					X	X	X	X	64	22	42	3171	0.172	0.344	1.2	3.67
7			X											1651	414	1237	3743	0.177	0.251	0.51	10.57
8	X	X												3019	620	2399	3743	0.177	0.205	0.2	9.43
9												X	X	1334	310	1024	3749	0.177	0.232	0.38	6.66
10					X									1776	408	1368	3749	0.177	0.23	0.37	8.1
11									X	X	X			894	178	716	3177	0.172	0.199	0.19	2.52
12										X	X			316	65	251	3177	0.172	0.206	0.24	1.66
13										X				95	16	79	3177	0.172	0.168	-0.02	-0.09

**Age:** 1 = <50, 2 = 50-69, 3 = >=70  
**Sex:** m = male, f = female  
**Liver size:** 1 = smaller than normal, 2 = normal, 3 = marginally increased, 4 = slightly increased, 5 = moderately increased, 6 = highly increased  
**Aorta sclerosis:** n = not calcified, c = calcified

Tabelle 23: Beispiele interessanter entdeckter Subgruppen in den Zeilen 1-6 und Referenz-Wahrscheinlichkeiten in den Zeilen 7-13. Angegebene Subgruppen-Parameter in den Spalten: „Size“ (Größe der Subgruppe), „TP“ (richtig positiv), „FP“ (falsch positiv), „Pop.“ (Größe der Bezugs-Population), „p0“ (Zielgröße in der Bezugs-Population), p (Zielgröße in der Subgruppe), RG (relative Zunahme in der Subgruppe) und binominale Qualitätsfunktion BQF (Bin. QF), welche die Subgruppengröße und Zielgröße kombiniert. Die folgenden Relationen gelten: TP+FP = Größe; p = TP/(TP+FP), RG = (p - p0) / (p0\*(1-p0)) und

$$BQF = \frac{(p - p0)}{\sqrt{p0(1 - p0)}} \sqrt{Size} \sqrt{(Pop. - Size) / Pop.}$$

Die einzelnen Subgruppen werden in den Reihen der Tabelle gezeigt, z. B. beschreibt die erste Zeile die Subgruppe (89 Fälle) mit ihren Charakteristika (Alter >= 70 und Geschlecht = weiblich und Lebergröße = leicht oder moderat oder stark vergrößert und Aortensklerose = kalzifizierend) mit der Zielgröße (Gallensteine) von 41,6% (p) im Vergleich zu 17,2% (p0) in der Gesamtpopulation mit einer relativen Zunahme von 171% (RG).

Die allgemeine Wahrscheinlichkeit eines Gallensteinleidens in der Fallserie von 4100 Fällen (ohne Wiederholungsuntersuchungen) lag bei 17%. Bekannte Risikofaktoren wie Alter (hier >=

70 Jahre) erhöhten die Prävalenz (25%, Zeile 7). Das gleichzeitige Vorliegen von „Alter  $\geq$  70 Jahre“, „Geschlecht = weiblich“ und „Lebergröße = leicht, moderat *oder* stark vergrößert“ (Zeile 5) ging mit einer Wahrscheinlichkeit von etwa 35% einher. Die Prävalenz stieg weiter an auf 39%, wenn außerdem eine „Aortensklerose“ dokumentiert wurde (Zeile 2), sowie auf 42% bei zusätzlicher Angabe von „Aortensklerose, kalzifizierend“ (Zeile 1). Die Zeilen 1 bis 6 der Tabelle 23 zeigen verschiedene Kombinationen von Alter, Geschlecht, Lebergröße und Aortensklerose mit der Zielgröße „Gallensteine“, wobei für alle eine Signifikanzniveau von wenigstens  $10^{-6}$  errechnet wurde. Die Zeilen 7 bis 13 stellt den Einfluss der genannten Parameter im Einzelnen dar.

## 4. Diskussion

Computerbasiert erstellte Untersuchungsprotokolle in der Medizin unterliegen den gleichen Anforderungen an Genauigkeit, Konsistenz, Lesbarkeit und Vollständigkeit in möglichst einfacher Darstellung wie konventionell verfasste Untersuchungsberichte<sup>25</sup>, um als Basis medizinischer Entscheidungen dienen zu können. SonoConsult erstellt aus problem- und fallbezogen akquirierten Informationen unter Anwendung einer „intelligenten“ Befundabfrage in nachvollziehbarer Abfolge einen Befundbericht von sonographischen Untersuchungen und stellt damit das einzige Programm dar, welches Expertenwissen in dieser Art einsetzt. Eine inhärente diagnostische Fähigkeit ist erforderlich, um Folgefragen in der Konsequenz bereits dokumentierter Daten oder vermuteter Diagnosen zu generieren. Programme mit derartigen Eigenschaften können das Untersuchungsverhalten von Ärzten beeinflussen<sup>4</sup>, befinden sich derzeit aber selten im klinischen Einsatz. Vermutlich spielen dabei sowohl Unzulänglichkeiten der Programme<sup>26</sup> als auch Bedenken der Anwender eine Rolle. Die vorliegenden Evaluationen bescheinigen SonoConsult eine hohe diagnostische Leistungsfähigkeit. Trotzdem wurde die Diagnosefunktion nicht von allen Untersuchern vorbehaltlos begrüßt.

Die Mehrzahl der perspektivischen Nutzer von SonoConsult gab eine einfache Bedienbarkeit, die standardisierte Befundeingabe und den schnellen Zugriff auf Vorbefunde für Befundvergleiche – nicht jedoch eine diagnostische Funktionalität - als die am meisten erwünschten Eigenschaften eines Computerprogramms an, mit dessen Hilfe strukturierte Befundberichte erstellt werden können. Im Gegensatz zur initialen Forderung, den Eingabedialog so kurz wie möglich zu gestalten, wurde später bei der Arbeit mit SonoConsult mehrfach der Wunsch nach einer Erweiterung der Wissensbasis geäußert, um eine möglichst differenzierte und umfassende Eingabe in standardisierte Befundformulare vornehmen zu können. In der Konsequenz wurden wiederholt Ergänzungen der Wissensbasis vorgenommen, da offenbar eine schnelle Dokumentation im Vergleich zur Vollständigkeit als sekundär eingeschätzt wird. Die Eingabedauer für konventionelle Befunde und mit SonoConsult erstellte Berichte ist vergleichbar. Wird zusätzlich die Befundqualität berücksichtigt, ergibt sich sogar ein Vorteil für Untersuchungsprotokolle, die mit dem Programm verfasst wurden.

SonoConsult wurde von den Untersuchern vorrangig als Werkzeug zur strukturierten Erfassung sonographischer Befunde gesehen, welches eine Standardisierung der Nomenklatur und daraus folgend eine verbesserte Befundvergleichbarkeit bewirkt. Es wurde bereits früher berichtet, dass

derartige Programme – auch ohne diagnostische Funktionalität – die Qualität von Befunden bezüglich Terminologie und Vollständigkeit verbessern<sup>27,28</sup>. Die Möglichkeit von SonoConsult, der diagnostischen Komponente folgend Daten „intelligent“ abzufragen, wurde von den Anwendern nicht in besonderer Art und Weise hervorgehoben. Unabhängig davon enthalten die Befundberichte durch diese Funktionalität sehr differenzierte Angaben zu erhobenen Befunden, insbesondere im Vergleich zur vor Einführung des Programms angewandten Befunderstellung in Freitext. Die Ärzte auf Station, die als Adressaten des Untersuchungsberichtes verstanden werden können, verwerten die detaillierten Angaben und diagnostischen Folgerungen von SonoConsult wiederum kaum. Von herausragender Bedeutung blieb für sie die Wertung des Untersuchers. Allerdings ist diese Tatsache problematisch, da mit einem solchen Verhalten ein erheblicher Informationsverlust einhergehen kann, wie die Evaluation der Konkordanz von Befundeinträgen und Untersucherdiagnosen zeigte. Trotzdem wurden die vom Programm generierten Diagnosen aufgrund des Ergebnisses der Befragung unter den Stationsärzten nicht weiter auf dem Befundbericht ausgegeben, da die Prüfung von Diskrepanzen zwischen Beurteilung des Untersuchers und Systemdiagnosen in der Verantwortung des Untersuchers – und nicht des Stationsarztes – liegen sollte.

Die Diagnosefunktion des Programms wurde von den Untersuchern mit mittlerer Bedeutung bewertet. Obwohl die Erinnerungsfunktion als prinzipiell hilfreich angegeben wurde, sahen sich die Untersucher von der Anzeige der Systemdiagnosen kaum beeinflusst. Diese Einschätzung scheint jedoch subjektiv und wird durch die Ergebnisse der Evaluation nicht in vollem Umfang gestützt. Seit Einführung von SonoConsult als alleinigem Dokumentationssystem im klinischen Alltag sind die Vollständigkeit und Verlässlichkeit der Untersucherdiagnosen eindeutig verbessert, erwartungsgemäß konnte dieser Effekt vor allem bei jüngeren Untersuchern beobachtet werden. Als ursächlich konnte neben einer strukturierten Befundeingabe auch die Erinnerungsfunktion herausgearbeitet werden, da eine Verblindung der SonoConsult-Diagnosen in einer vergleichenden Fallserie eine deutliche Abnahme der Diagnosequalität bewirkte. Offenbar finden diese Vorgänge unterbewusst statt, wie aus der differierenden Eigeneinschätzung der Untersucher gefolgert werden muss. Die höhere diagnostische Ausbeute des Programms wurde vor allem durch eine vollständige Aufführung von Diagnosen erzielt, die zum Untersuchungszeitpunkt nicht im Mittelpunkt des Interesses standen. Die Untersucher dagegen tendierten dazu, die Ergebnisse des zentralen Problems herauszustellen und Diagnosen von – aus ihrer Sicht – geringerer Relevanz unerwähnt zu lassen, obwohl diese auch beim Verfassen der eigenen Beurteilung in der SonoConsult-Oberfläche angezeigt werden. Zeitmangel

und Arbeitsaufwand behindern mutmaßlich den Willen zur Korrektur zusätzlich. Es ist anzunehmen, dass in der Mehrzahl der Fälle das Fehlen von Diagnosen geringerer Relevanz ohne Konsequenz für die weitere Behandlung des Patienten bleibt. Allerdings kann eine Filterung der Information durch den Sonographie-Untersucher nicht gewollt sein und läge eher in den Händen des Stationsarztes. Weiterhin kann das Hinzutreten neuer Aspekte im Krankheitsverlauf bisher nebensächliche Befunde in den Blickpunkt des Interesses rücken. Aus den genannten Gründen stand die Qualität der Bewertung des Untersuchers im Fokus weiterer Untersuchungen.

In vereinzelt Studien konnte gezeigt werden, dass der Einsatz wissensbasierter Systeme in der Medizin die diagnostische Qualität der Untersucher positiv beeinflussen kann<sup>4</sup>. Unbekannt ist dagegen, in welchem Maße Beurteilungen sonographischer Befunde als verlässlich gelten können, wenn die Kongruenz der dokumentierten Befunde und die daraus ableitbaren Diagnosen als Kriterium zugrunde gelegt werden. Da sich die von SonoConsult generierten Diagnosen als hochgradig konsistent zu den Untersuchungsbefunden charakterisieren ließen, wurden diese mit der Beurteilung des Untersuchers verglichen, um die Konkordanz zu prüfen.

Die Ergebnisse der Evaluation lassen erkennen, dass die im klinischen Alltag einer Medizinischen Klinik verfassten Befundberichte sonographischer Untersuchungen nicht fehlerfrei sind. In einem erstaunlichen Umfang stimmen die Ergebnisse der Untersucher nicht mit den diagnostischen Folgerungen von SonoConsult überein. Die folgenden Gründe können angenommen werden: (a) fehlende oder falsche Dokumentation von Befunden, die zur Klärung einer Diagnosen beitragen, (b) fehlende diagnostische Schlussfolgerungen des Untersuchers und (c) unterschiedliche Interpretation von Befunden durch Untersucher und bewertende Evaluatoren infolge des subjektiven Charakters der Interpretation sonographischer Befunde.

Spätere Diskussionen mit den Untersuchern ließen erkennen, dass (c) keine wesentliche Rolle spielt. Es ist letztlich nicht möglich, eine scharfe Trennung der beiden anderen Ursachen (a) und (b) vorzunehmen (zum Beispiel inkorrekt oder nicht erfolgter Eingabe von Befunden oder Diagnosen als Ursache von Diskrepanzen), da die Auswertung auf Basis der dokumentierten Befunde und nicht des abgespeicherten Bildmaterials einer Untersuchung erfolgte. Prinzipiell wäre es dem Untersucher beim Verfassen seiner Beurteilung möglich, falsche Befundeinträge zu erkennen, wenn abweichende Systemdiagnosen vorliegen. Wie bereits mehrfach angeführt, fließen die Diagnosen von SonoConsult jedoch nicht bewusst in die Befundbeurteilung ein. Eine

Diskussion mit den Untersuchern und das Ergebnis von gelegentlichen Nachuntersuchungen lassen vermuten, dass es sich in der Mehrzahl der Fälle um Eingabefehler oder -lücken handelt. Der relativ hohe Anteil fehlender Diagnosen einfacher Herleitung kann als hinweisend gelten. Fehlende diagnostische Schlussfolgerungen dürften dagegen nur bei jüngeren Untersuchern eine Rolle spielen und dann vor allem Diagnosen komplexer Herleitung betreffen. Unabhängig davon muss es als Ziel gelten, den Anteil von fast 10% fehlender relevanter Diagnosen in der Beurteilung zu reduzieren. Da die einfache Anzeige der Programmdiagnosen in einem separaten Fenster die Aufmerksamkeit des Untersuchers nicht ausreichend auf sich zieht, um seine eigene Freitextbeurteilung zu beeinflussen, müssen andere Mechanismen gefunden werden, die im Falle von Diskrepanzen eine einfache Korrektur ermöglichen, ohne die Akzeptanz des Programms negativ zu beeinflussen.

Untersucher wissen meist nicht, ob ihre Diagnosen korrekt angegeben sind<sup>29</sup>. Eine Programmkomponente, die automatisch die Diagnosen von SonoConsult und die der Untersucher vergleicht und deren klinische Relevanz untersucht, könnte perspektivisch zum Einsatz kommen. Im Falle von nennenswerten Diskrepanzen könnten die entsprechenden Frageformulare zur Korrektur der Eingabedaten oder die resultierenden Diagnosen zur Änderung angeboten werden. Gleichzeitig ermöglicht die Komponente ein Aufdecken von etwaigen Regelfehlern der Wissensbasis. Die Vorbedingung für eine solche Prozedur ist die linguistische Analyse der Freitexteingabe des Untersuchers und deren Vergleich mit den Programmdiagnosen<sup>30</sup>. Der Ausbau eines solchen Programms zu einem Kritiksystem<sup>31,32</sup> ist ein realistischer Ansatz, die Qualität computerbasiert erstellter medizinischer Untersuchungsberichte und ihrer inhärenten diagnostischen Schlussfolgerungen zu verbessern. Davon unberührt bleibt die hohe Untersuchungsqualität des Arztes Vorbedingung für einen verlässlichen Befund.

Vor Einführung von SonoConsult in Würzburg wurde aufgrund der Erfahrungen in Berlin ein Expertenmodus zur Befundeingabe integriert. Der Untersucher dokumentiert zuerst seine erhobenen Diagnosen, im Anschluss bietet das Programm die wesentlichen Frageformulare an, die zur Verifizierung der Diagnosen erforderlich sind. Die Konsistenzprüfung durch SonoConsult ist in diesem Modus in geringerem Maße möglich als bei Anwendung des Standardmodus, andererseits ist dieses Vorgehen zeitsparend. Mutmaßlich kann die Konkordanz von Befunden und Diagnosen erhöht werden, allerdings setzt der Modus ein hohes Maß sonographischer Kenntnisse voraus und stellt daher nur eine Alternative für sehr erfahrene Untersucher dar. Die Einführung von SonoConsult in Würzburg zeigte wie bereits die

Untersuchung der Genauigkeit der Programmdiagnosen durch Evaluatoren der zwei Kliniken, dass bezüglich einiger verwendeter Elemente der Wissensbasis (zum Beispiel Grenzen normaler Organgrößen) verschiedene Referenzen verwendet werden. Es ist notwendig, diese Tatsache bei der Auswertung von Daten verschiedener Kliniken zu berücksichtigen und gegebenenfalls Anpassungen der Wissensbasis vor Einführung des Programms in anderen Einrichtungen vorzunehmen.

Die Verlässlichkeit medizinischer Befunde durch unterschiedliche Bewertung der Untersucher ist ein wichtiges Problem im klinischen Alltag. Studien zu dieser Fragestellung umfassen meist die wiederholte Untersuchung eines Patienten mit anschließendem Vergleich der Ergebnisse<sup>33,34</sup>. Um die Ursachen von Inkonsistenzen in der sonographischen Befundung zu differenzieren, ist in der Regel die Wiederholung der Untersuchung durch einen unabhängigen Experten notwendig. Der Experte kann im Vergleich der Befunde zwischen Problemen in der Untersuchungstechnik, bei der Befundeingabe oder im diagnostischen Schluss unterscheiden. Die alleinige Beurteilung des Bildmaterials der Erstuntersuchung durch einen Experten ist zu diesem Zwecke nicht ausreichend, da sie selbst fehlerbehaftet ist. Da Wiederholungsuntersuchungen für verschiedene Analysen und Untersucher mehrfach durchgeführt werden müssten und somit sehr zeit- und kostenintensiv wären, wurde auf den Einsatz einer statistischen Qualitätskontrolle fokussiert, deren Anwendbarkeit gezeigt werden konnte.

In den vorgenommenen Analysen zur Inter-Untersucher-Variation lassen sich Unterschiede im Befundungsverhalten auf eine oder zwei in der Häufigkeit von der Gesamtheit abweichende Variablen zurückführen. Dadurch ergibt sich die Möglichkeit, im klinischen Alltag gezielt Unterstützung oder Übungen zur Interpretation einzelner Befunde anzubieten. Ein weiteres Instrument bietet die standardisierte und formularbasierte Dateneingabe in SonoConsult, in der zu jeder Frage oder Antwortalternative eine Information hinterlegt werden kann („I“ in Abb. 1), welche erklärendes Wissen in Form von Text oder Bildern - zum Beispiel zur Festlegung einheitlicher Messpunkte - beinhaltet. Für die meisten kritischen Befundalternativen existieren solche Angaben und sind während der Dokumentation für den Untersucher durch einfaches Anklicken verfügbar, jedoch werden sie bisher wenig genutzt. Es erscheint sinnvoll, den Untersuchern die Ergebnisse der Auswertungen zu ihrem Befundungsprofil zur Verfügung zu stellen, um sie über Abweichungen in der Häufigkeit von Diagnosen, Befunden oder Befundkombinationen im Vergleich zur Gesamtheit zu informieren. Dieses Bewusstsein könnte in der Folge eine häufigere Nutzung der hinterlegten Informationen bei Befundeingaben, die als

abweichend herausgearbeitet werden konnten, bewirken. Einen zweiten Ansatz besteht im bereits erwähnten Vergleich der diagnostischen Schlussfolgerungen des Programms mit denen des Untersuchers. Dieser Weg setzt eine entsprechende Präsentation inkonsistenter Befundeingaben nach automatischer linguistischer Analyse voraus. Es ist denkbar, mit Hilfe von SonoConsult und der Auswertung zu Untersucherprofilen individuelle Tests mit mehrfachen Antwortalternativen („multiple choice“) zu entwerfen, welche zum Beispiel von weniger erfahrenen Untersuchern zu Trainingszwecken auf der Basis realer Fälle beantwortet werden. Die Programmierung von SonoConsult (Java) lässt die Anwendung in einem Netzwerk zu. Der Effekt eines derartig konzipierten Trainings ist durch erneute Subgruppen-Analyse von Fällen des Untersuchers unter Verwendung der gleichen Variablen quantifizierbar.

Insgesamt zeigt der Vergleich der Untersucher deutliche Unterschiede im Dokumentations- und Interpretationsverhalten. Die mit Hilfe der Subgruppen-Analyse vorgenommene statistische Qualitätskontrolle weist darauf hin, dass die diagnostische Zuverlässigkeit einzelner Untersucher variiert. Diese Beobachtung stimmt überein mit der Tatsache, dass die dokumentierten Befunde und angegebenen Diagnosen bis zu etwa einem Drittel inkonsistent sind. Eine Möglichkeit der Interpretation besteht darin, dass mit der Sonographie eine Untersuchungsmethode evaluiert wurde, die sehr von den Fertigkeiten und Fähigkeiten des Untersuchers abhängt, zum Beispiel der Erfahrung in der praktischen Durchführung selbst als auch der Exaktheit der Dokumentation. Die Annahme, dass die dokumentierten Befunde weniger verlässlich sind als die diagnostischen Folgerungen der Untersucher, fokussiert die weitere Entwicklung auf Methoden zur Kontrolle der Datenqualität. Es muss beachtet werden, dass Abweichungen zwischen den Untersuchern auch durch Inhomogenitäten der Fallgruppen bedingt sein können. Dieser Faktor gewinnt an Einfluss, wenn die Zahl der analysierten Fälle eines Untersuchers klein gewählt wird, um frühzeitig – vor allem bei jüngeren Untersuchern – Unterschiede im Befundungsverhalten feststellen und reagieren zu können, was im Sinne der Qualitätskontrolle andererseits wünschenswert ist.

Zur Einschätzung der Datenqualität sonographischer Befunde kann der Vergleich auf andere patientenbezogene Angaben ausgeweitet werden, die zum Ende eines Krankenhausaufenthaltes beispielsweise im Entlassungsbrief durch die Stationsärzte angegeben oder im Krankenhausinformationssystem codiert werden. Die Berücksichtigung weiterer Untersuchungsergebnisse aus Labor, Computertomographie und Kernspintomographie ist denkbar. Die Auswertung dieser Daten setzt jedoch – mit Ausnahme der ICD-Diagnosen – eine



Standardisierung der Diagnosen in diesen Berichten voraus, die ohne automatisierte PC-basierte Unterstützung nicht möglich ist<sup>30</sup>. Die codierten ICD-Diagnosen wären einem Vergleich mit sonographisch generierten Diagnosen zwar zugänglich, bilden allerdings die Entlassdiagnosen des Patienten unter medizinischen Gesichtspunkten oft unzureichend ab, da ihre Verwendung primär auf Abrechnungszwecke ausgerichtet ist.

Exemplarisch für eine medizinische Auswertung wurde die Assoziation von Gallensteinen mit verschiedenen sonographischen Untersuchungsbefunden analysiert. Das Interesse an einer Subgruppe entsteht zum einen durch die Zunahme der Wahrscheinlichkeiten („Relative Gain“ = RG) sowie durch die Größe der Subgruppe. Es ist weiterhin notwendig, für klinische Fragen einfache und nachvollziehbare Subgruppen zu definieren. Aus diesem Grunde ist die Einbindung von Experten mit medizinischen Hintergrundwissen in die Auswertung unabdingbar.

Die Korrelation der Gallensteinhäufigkeit mit dem Alter und Geschlecht entspricht im Wesentlichen der Literatur, wobei Diskrepanzen in der Häufigkeit wahrscheinlich der Selektion des Patientengutes in einer gastroenterologisch spezialisierten internistischen Klinik geschuldet sind. Die Aufschlüsselung weiterer mit dem Gallensteinleiden assoziierter Untersuchungsbefunde zeigt die Möglichkeiten einer hochpotenten und sehr differenzierten Wissenssuche exemplarisch auf. Aufgrund der in dieser Arbeit festgestellten Inter-Untersucher-Variation ist die Qualität der Falldatenbank jedoch weniger hochwertig als erwartet anzunehmen, so dass die Voraussetzung für die Untersuchung mancher Fragestellungen kritisch bewertet werden muss. Die Reduktion der Eingabefehler und -lücken sowie der Inkonsistenzen zwischen Eingaben und Beurteilungen, wie sie zum Beispiel durch eine Kritikkomponente zu erwarten ist, kann die Inter-Untersucher-Variation minimieren und eine wichtige Maßnahme zur Generierung einer für wissenschaftliche Zwecke der Datenanalyse hochwertigen Falldatenbank sein.

## 5. Zusammenfassung

**Hintergrund:** Die Qualität medizinischer Untersuchungsbefunde hat einen unmittelbaren Einfluss auf die medizinische Versorgung. Freitextbefunde können hinsichtlich ihrer Vollständigkeit problematisch sein, während eine Befunderfassung über Masken in einem PC-Programm wenig Flexibilität bietet. Beide Sachverhalte können die Qualität der Dokumentation negativ beeinflussen.

**Methoden:** SonoConsult, ein medizinisches Expertensystem für die Sonographie, bietet die Möglichkeit einer strukturierten und fallbezogenen Befunderfassung und verfügt über eine Diagnosefunktion und Erklärungsfunktion. Das Programm wird seit fünf Jahren in zunächst einer Ultraschallabteilung einer internistischen Klinik verwendet. Es erfolgte die Evaluation über Fragebögen, die den Anwendern vor und nach Einführung offeriert wurden. Eine gleichartige Befragung fand unter den auf Station tätigen Ärzten als Adressaten der Befunde statt. Die Befundqualität wurde im Vergleich zu den vorher verwendeten Freitextbefunden eingeschätzt, der Wert und die Validität der diagnostischen Kompetenz des Programmes in weiteren Fallserien erhoben, wobei in einfache und komplex herzleitende Diagnosen sowie Diagnosen mit weniger und deutlicher Relevanz für den Patienten unterschieden wurde. Die entstehende Falldatenbank wurde zur Auswertung untersucherspezifischer Befundungsprofile und zu medizinischen Fragestellungen benutzt.

**Ergebnisse:** SonoConsult wurde von den Untersuchern und Stationsärzten gut akzeptiert und kann die Einarbeitung von Anfängern in der Sonographie unterstützen. Der diagnostischen Funktionalität des Programmes wurde nur untergeordnete Bedeutung beigemessen, obwohl ein positiver Einfluss auf die Vollständigkeit der Untersucherdiagnosen gezeigt werden konnte. Fehlende Diagnosen der Untersucher betrafen in 40% relevante Befunde. Im Vergleich zu Freitextbefunden führt die Anwendung des Programmes zu deutlich detaillierteren und besser vergleichbaren Befunden. Die diagnostische Validität insbesondere für komplexe Diagnosen ist hoch. Die entstehende Falldatenbank kann medizinische Zusammenhänge aufdecken und für die Qualitätsverbesserung über die Erstellung von Untersucherprofilen verwendet werden.

**Schlussfolgerung:** Die Anwendung von SonoConsult kann die Befundqualität in der Sonographie ohne zeitlichen Mehraufwand steigern. Selbst bei Einsatz eines Expertensystems muss mit Diskrepanzen zwischen den formularbasiert dokumentierten Befunden und den freitextlich diagnostischen Beurteilungen der Untersucher gerechnet werden. Durch einen automatisierten Vergleich mit den Programmdiagnosen, der eine Analyse des Freitextes voraussetzt, könnte dieses Problem überwunden werden.

## 6. Literatur

---

- <sup>1</sup> Kuhn K, Zemmler T, Reichert M, et al. An integrated knowledge-based system to guide the physician during structured reporting. *Meth Infom Med* 1994; 33: 417-422.
- <sup>2</sup> Schwartz W. Medicine and the Computer: The promise and problems of change. *N Engl J Med* 1970; 283: 1257-1264.
- <sup>3</sup> Darmoni SJ, Poynard T. Computer-aided decision support in hepatology. *Scand J Gastroenterol* 1992; 27: 889-896.
- <sup>4</sup> Berner ES, Maisiak RS, Cobbs CG, Taunton OD. Effects of a decision support system on physicians' diagnostic performance. *J Am Med Inform Assoc* 1999; 6: 420-427.
- <sup>5</sup> Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer based diagnostic systems. *N Engl J Med* 1994; 330: 1792-1796.
- <sup>6</sup> Adlassnig KP, Horak W. Development and retrospective evaluation of Hepaxpert-I: A routinely-used expert system for interpretive analysis of hepatitis A and B serologic findings. *Artif Intell Med* 1995; 7: 1-24.
- <sup>7</sup> Puppe B, Riecker G. *CardioConsult: Kardiologisches Second-Opinion-Programm*. Ullstein-Medical, 1998.
- <sup>8</sup> Schewe S, Schreiber MA. Stepwise development of a clinical expert system in rheumatology. *Clinical Investigator* 1993; 71: 139-144.
- <sup>9</sup> Lucas PJ, Segaar RW, Janssens AR. HEPAR: An expert system for the diagnosis of disorders of the liver and biliary tract. *Liver* 1989; 9: 266-275.
- <sup>10</sup> Bamberger S, Puppe F. Kooperierende Diagnoseagenten. *Informatik: Forschung und Entwicklung* 1999; 14: 135-144.
- <sup>11</sup> Lucas PJ. Refinement of the HEPAR Expert System: Tools and Techniques. *Artif Intell Med* 1994; 6: 175-188.
- <sup>12</sup> Berner ES (ed.). *Clinical Decision Support Systems*. Springer 2007, 2<sup>nd</sup> edition.
- <sup>13</sup> Buscher HP, Engler C, Führer A, Kirschke S, Puppe F. HepatoConsult: A knowledge-based second opinion and documentation system. *Artif Intell Med* 2002; 24: 205-216.
- <sup>14</sup> Puppe F, Gappa U, Poeck K, Bamberger S. *Wissensbasierte Diagnose- und Informationssysteme*. Springer 1996.

- 
- <sup>15</sup> Puppe F. Knowledge reuse among diagnostic problem-solving methods in the Shell-Kit D3. *Int J Human-Computer Studies* 1998; 49: 627-649.
- <sup>16</sup> McDonald CJ. Medical Heuristics: The silent adjudicators of clinical practice. *Ann Intern Med* 1996; 124: 56-62.
- <sup>17</sup> Miller R, Masarie FE, Myers JD. Quick medical reference (QMR) for diagnostic assistance. *Medical Computing* 1986; 5: 34-38.
- <sup>18</sup> Han J, Kamber M. *Data mining, concepts and techniques*. Morgan Kaufmann Publishers Inc. 2006, 2<sup>nd</sup> edition.
- <sup>19</sup> Klösgen W. *Handbook of data mining and knowledge discovery*. Oxford University Press 2002; Chapter 16.3: Subgroup Discovery.
- <sup>20</sup> Lamberger D, Lavrac N. Expert-guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research* 2002; 17: 501-527.
- <sup>21</sup> Atzmüller M. *Knowledge-intensive subgroup mining – Techniques for automatic and interactive discovery*. AKA Berlin, DISKI 307, 2007.
- <sup>22</sup> Atzmüller M, Puppe F, Buscher HP: Exploiting background knowledge for knowledge-intensive subgroup discovery. *Proc. 19<sup>th</sup> International Joint Conference on Artificial Intelligence (IJCAI-05)*, 2005.
- <sup>23</sup> Atzmüller M, Puppe F. Semi-automatic visual subgroup mining using VIKAMINE. *Journal of Universal Computer Science (JUCS), Special Issue on Visual Data Mining* 2005; 11: 1752-1765.
- <sup>24</sup> Dressler A. *Diagnosekodierung medizinischer Freitexte – am Beispiel sonographischer Befundberichte*. Diplomarbeit, Universität Würzburg, Institut für Informatik, 2005.
- <sup>25</sup> Bradbury A. Computerized medical reports: The need for a standard. *J Am Med Rec Assoc* 1990; 61: 25-37.
- <sup>26</sup> McIntyre N. Computer-aided diagnosis in jaundice and liver disease. *J Hepatol* 1986; 3: 269-272.
- <sup>27</sup> Morris AH. Developing and implementing computerized protocols for standardization of clinical decisions. *Ann Intern Med* 2000; 132: 373-383.
- <sup>28</sup> Kuhn K, Gaus W, Wechsler JG, et al. Structured reporting of medical findings: Evaluation of a system in gastroenterology. *Methods Inf Med* 1992; 31: 268-274.
- <sup>29</sup> Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *J Gen Intern Med* 2005; 20: 334-339.

- 
- <sup>30</sup> Klügl P, Schuhmann M., Puppe F, Buscher HP. Expectation driven information extraction in incomplete sentences (German). Proc. of LWA in Halle 2007, Germany: 237-243.
- <sup>31</sup> Guerlain S, Smith J, Rudmann S, et al. Interactive critiquing as a form of decision support: An empirical evaluation. *Human Factors* 1999; 41: 72-89.
- <sup>32</sup> Puppe F. Meta knowledge for extending diagnostic consultation to critiquing systems. Proc. of EKAW-99 11<sup>th</sup> European Workshop on Knowledge Acquisition, Modeling and Management. Springer 1999: LNAI 1621: 367-272.
- <sup>33</sup> Close RJ, Sachs CJ, Dyne PL. Reliability of bimanual pelvic examinations performed in emergency departments. *West J Med* 2001; 175: 240-244.
- <sup>34</sup> Najm WI, Seffinger MA, Mishra SI, et al. Content validity of manual spinal palpatory exams – A systematic review. *BMC Complement Altern Med* 2003; 3: 1.

## **Danksagung**

Mein Dank gilt Herrn Professor Hans-Peter Buscher für die Themenstellung und jahrelange unermüdliche Unterstützung bis zur Fertigstellung der vorliegenden Promotion. Seine stets diesem Ziel verpflichtete Betreuung vermittelte mir Optimismus und Orientierung vor allem in schwierigen Phasen der Weiterentwicklung des Themas.

Herrn Professor Frank Puppe und Dr. Martin Atzmüller vom Institut für Informatik und künstliche Intelligenz der Universität Würzburg danke ich für die konstruktiven Diskussionen und statistischen Auswertungen.

In den vergangenen sieben Jahren haben viele ärztliche Kollegen in der Sonographieabteilung unserer Klinik mitgearbeitet und durch ihre Bereitschaft zur Teilnahme an Evaluation und Befragungen zum Ergebnis der Arbeit beigetragen. Stellvertretend für alle sollen an dieser Stelle Dr. Eckhard Kopke, Dr. Cordula Hagen und Dr. Daniel Merkel genannt werden.

## Erklärung

Ich, Matthias Hüttig, erkläre, dass ich die vorgelegte Dissertation mit dem Thema:

*Eignung eines „intelligenten“ Dokumentationssystems für die Sonographie zur Qualitätskontrolle, zur Ausbildung von Untersuchern und zur Anlage einer hochwertigen Falldatenbank*

selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.“

Berlin, den 02. Dezember 2008

Matthias Hüttig

## **Publikationen**

Huettig M, Buscher G, Menzel T, Scheppach W, Puppe F, Buscher HP. A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. *Med Klin* 2004; 99: 117-122.

Puppe F, Buscher G, Atzmueller M, Hüttig M, Buscher HP. Clinical Experiences with a Knowledge - Based System in Sonography (SonoConsult). In K.-D. Althoff et al. (Eds.): *WM 2005*, Springer-Verlag Berlin-Heidelberg 2005, LNAI 3782: 319-329.

Huettig M, Buscher G, Puppe F, Buscher HP. Checking Concordance between Findings and Diagnoses in Sonographic Reports by a Knowledge-Based Documentation System. *Ultraschall in Med* 2008; 29: 289-293.



## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.