

**Aus dem Deutschen Zentrum für Herzinsuffizienz  
und der Medizinischen Klinik und Poliklinik I  
der Universität Würzburg  
Direktor: Professor Dr. med. Georg Ertl**

***„SymptomCheck“***

-

ein Programm  
zur Symptom-orientierten  
Ableitung eines  
Diagnosevorschlags

**Inaugural - Dissertation  
zur Erlangung der Doktorwürde der  
Medizinischen Fakultät  
der**

**Julius-Maximilians-Universität Würzburg**

**vorgelegt von**

**Finn Pietrowski**

**aus Schweinfurt**

**Würzburg, Februar 2018**

**Referent bzw. Referentin:**

Prof. Dr. Stefan Störk

**Korreferent bzw. Korreferentin:**

Prof. Dr. Frank Puppe

**Dekan:**

Prof. Dr. Matthias Frosch

**Tag der mündlichen Prüfung:**

27.08.2019

**Der Promovend ist Arzt.**

# Inhalt

<b>Einleitung</b> .....	<b>1</b>
<b>1 Aktueller Hintergrund</b> .....	<b>1</b>
<b>2 Begriffe und deren Stellenwert</b> .....	<b>4</b>
<b>3 Wissensbasierte Systeme in der Medizin – ein Überblick</b> .	<b>6</b>
<b>3.1 Clinical decision support systems (CDSS)</b> .....	<b>6</b>
3.1.1 Definition.....	6
3.1.2 Entwicklungen und Standards.....	7
3.1.3 Evaluation.....	9
<b>3.2 Systeme für das Patientenklintel</b> .....	<b>11</b>
3.2.1 Analyse der Beliebtheit .....	12
3.2.2 Negative Kritikpunkte .....	14
3.2.3 Positive Kritikpunkte.....	15
3.2.4 Evaluation von Selbstdiagnoseprogrammen .....	17
<b>Material und Methoden</b> .....	<b>19</b>
<b>4 Entwicklung der Wissensbasis</b> .....	<b>19</b>
4.1 Software.....	19
4.2 Diagnosen .....	19
4.3 Symptome .....	20
4.4 Entwicklung eines Bewertungssystems .....	23
4.5 Entwicklung der Startseite.....	31
<b>5 Klinische Studie zu SymptomCheck</b> .....	<b>37</b>
<b>5.1 Studienziele und Fragestellungen/Hypothesen</b> .....	<b>37</b>
5.1.1 Primäres Studienziel .....	37
5.1.2 Sekundäre Studienziele .....	38
5.1.3 Hypothesen .....	38
<b>5.2 Studiendesign</b> .....	<b>38</b>
<b>5.3 Studienpopulation</b> .....	<b>38</b>

5.4	Ein- und Ausschlusskriterien der Studienpopulation .....	39
5.5	Fragebogen zur Datenerhebung .....	40
5.6	UEQ-Fragebogen zur Benutzerfreundlichkeit .....	40
5.7	Studienablauf .....	43
5.8	Datenschutz und Ethikvotum .....	45
5.9	Fallzahlen .....	45
5.10	Nutzen-Risiko-Abwägung .....	46
5.11	Datenanalyse.....	46
<b>6</b>	<b>Vergleich von SymptomCheck mit einem Arzt .....</b>	<b>47</b>
	<b>Ergebnisse .....</b>	<b>48</b>
<b>7</b>	<b>SymptomCheck vs. Arzt .....</b>	<b>48</b>
<b>8</b>	<b>Pilotstudie I .....</b>	<b>51</b>
8.1	Rahmendaten .....	51
8.2	Benutzungsqualität .....	53
8.3	Überprüfung der Verdachtsgenerierung .....	55
<b>9</b>	<b>Online-Evaluation .....</b>	<b>59</b>
9.1	Rahmendaten .....	59
9.2	Benutzungsqualität .....	63
9.2.1	UEQ-Fragebogen .....	63
9.2.2	Bewertung spezifischer Aussagen zu SymptomCheck.....	67
<b>10</b>	<b>Pilotstudie II .....</b>	<b>71</b>
10.1	Rahmendaten .....	71
10.2	Studiendauer für die einzelnen Teilnehmer .....	74
10.3	Benutzungsqualität .....	74
10.3.1	UEQ-Fragebogen .....	74
10.3.2	Bewertung spezifischer Aussagen zu SymptomCheck.....	76
10.4	Überprüfung der Verdachtsgenerierung .....	77
	<b>Diskussion .....</b>	<b>82</b>
11	Besonderheiten des Patienten- und Probandenkollektivs.....	82

12	Kritikpunkte .....	82
13	Überprüfung der Hypothesen.....	84
14	Schwächen von SymptomCheck .....	85
15	Stärken von SymptomCheck.....	88
16	Anregungen für die Zukunft .....	92
<b>Zusammenfassung.....</b>		<b>93</b>
<b>Literaturverzeichnis .....</b>		<b>95</b>
<b>Tabellenverzeichnis .....</b>		<b>101</b>
<b>Abbildungsverzeichnis .....</b>		<b>104</b>
<b>Anhang.....</b>		<b>106</b>
1	Tabellen .....	106
2	Abbildungen.....	135
3	Studienprotokoll .....	137
4	Fragebögen .....	146
<b>Danksagung .....</b>		
<b>Lebenslauf .....</b>		

# Einleitung

## 1 Aktueller Hintergrund

Heutzutage ist das Internet für die Mehrzahl der Bevölkerung alltäglicher Begleiter. Dabei dienen Suchmaschinen (wie z. B. Google) oftmals auch zu medizinischen Zwecken. Ratsuchende tauschen sich auf den zahlreich verfügbaren medizinischen Portalen und Foren im Internet aus.<sup>1</sup> Die Erwartungshaltung ist, dass durch Eingabe einiger weniger Informationen (häufig: Symptome) eine auf sie zutreffende Diagnose gestellt oder ausgeschlossen werden kann. Diese Internetangebote sind jedoch nicht unumstritten: Problematisch ist, dass die meisten webbasierten Programme einzelne Symptome bewerten und keine Symptomkomplexe. Eine gewichtete Diagnosestellung ist somit nicht möglich. Die Folge sind hohe Raten an falsch positiven und falsch negativen Diagnosen. Zudem neigen Patienten, aber auch gesunde Normalpersonen, häufig dazu, sich zu „überdiagnostizieren“, sich also eine Krankheit zuzusprechen, die vermutlich gar nicht vorliegt. Die Online-Suche nach Informationen zu Symptomen oder Krankheiten kann so besonders bei gesundheitsbesorgten Personen völlig unbegründetes Unbehagen und Angst verursachen.<sup>2</sup> Forscher aus den USA haben dafür einen neuen Begriff geprägt, die Cyberchondrie. Viele Menschen neigen auch grundsätzlich zu einer hypochondrischen Einstellung ihrem Gesundheitszustand gegenüber, die durch unkoordinierte Verdachtsgenerierung (z. B. in Gesundheitsforen des Internets) oft noch verstärkt statt vermindert wird.<sup>3-5</sup>

Aus diesem Grund wurde am Deutschen Zentrum für Herzinsuffizienz in Würzburg (DZHI) gemeinsam mit dem Lehrstuhl Informatik ein Programm, *SymptomCheck*, entwickelt, das zum Ziel hatte, hier objektiver und wertneutraler zu beraten und damit möglicherweise angstvermeidend zu wirken. Das Programm *SymptomCheck* ist frei zugänglich unter dem Link: <https://go.uni-wue.de/symptomcheck>.

*SymptomCheck* ist internetbasiert und besteht aus zwei Komponenten: mit der Verdachtsgenerierungskomponente kommt man zu einem Diagnose-Vorschlag;

mittels der Verdachtsüberprüfungskomponente wird die Diagnose bekräftigt oder entkräftet.

Die vorliegende Dissertation befasst sich ausschließlich mit der Verdachtsgenerierungskomponente (Die Verdachtsüberprüfungskomponente ist Bestandteil eines komplementären Projekts und wird im Folgenden nicht behandelt).

In einer dem Projekt vorausgehenden intensiven Literaturrecherche wurde keine Studie identifiziert, in der ein Verdachtsgenerierungsprogramm direkt klinisch an Patienten und Probanden getestet wurde. Diese Studie ist somit die erste, in der versucht wurde die Benutzerfreundlichkeit und die Validität direkt im alltäglichen Gebrauch zu messen. In der vorliegenden Arbeit wird die Entwicklung von *SymptomCheck* dargestellt, sowie dessen Performance anhand von zwei Pilotstudien und eines Online-Kollektivs charakterisiert.

Als Vorläuferprogramm von *SymptomCheck* kann *CardioConsult* (mittlerweile vergriffen, aber gebraucht noch bei Amazon erhältlich Stand Oktober 2017) angesehen werden<sup>6</sup>. Dieses System sollte als Second-Opinion-Programm Ärzte bei der Diagnostik unterstützen und adressierte demzufolge vornehmlich Ärzte (und weniger Patienten)<sup>6</sup>. Im Gegensatz dazu richtet sich *SymptomCheck* an die Allgemeinbevölkerung, also Menschen, die allgemeinen Gesundheitsrat suchen. *SymptomCheck* soll effizienter und schneller bearbeitbar sein als herkömmliche Programme, da nicht alle Symptome abgefragt werden. Viele Fragen erscheinen nur, wenn man in vorherigen Fragen bestimmte Antworten gegeben hat. Zudem kann *SymptomCheck* nicht nur Diagnosen vorschlagen, sondern auch ausschließen. Durch die graphische Aufarbeitung soll es benutzerfreundlicher und intuitiv zu bedienen sein. Selbstverständlich ersetzt *SymptomCheck* nicht die ärztliche Konsultation, sondern soll dazu beitragen, die „Ressource Arzt“ besser zu nutzen.

Die intendierte Leistung von *SymptomCheck* lässt sich wie folgt zusammenfassen:

- Wahrscheinlichkeitsberechnung diverser Diagnosen anhand der Patientensymptomatik durch Berücksichtigung multipler Symptome
- Möglichkeit zur differenzierten Symptomabfrage trotz schlichter und auf wesentliche Symptome reduzierter Benutzeroberfläche

- Ausschluss möglicher Erkrankungen anhand spezifischer Fragen, anhand derer eine Erkrankung für eher zutreffend oder unwahrscheinlich eingestuft werden kann
- Erhalt von weiteren Informationen zu bestimmten Diagnosen und Untersuchungen

## 2 Begriffe und deren Stellenwert

### Diagnose

Diagnosen spielen in der Medizin eine entscheidende Rolle. Sie dienen als Klassifizierungswerkzeuge für diverse Krankheiten. Die nachstehende Aussage beschreibt diese hervorstechenden Eigenschaften: „*[Diagnoses work by] segmenting and ordering corporeal states, valorising some, disregarding others, and in any case, exerting an important material force.*“<sup>7</sup>

Folglich ist eine Diagnose eine medizinische Bezeichnung für eine bestimmte pathologische Symptomkonstellation, die ein Arzt nach ausreichender Diagnostik und Überlegung gestellt hat<sup>8</sup>.

Brauchte man früher noch eine Reihe an Beschreibungen der Symptome, um ein Krankheitsbild für andere nachvollziehend darzustellen, so ist heutzutage meist ein einziges Wort – „die Diagnose“ – ausreichend<sup>7</sup>. Jutel fasst die Konsequenz der Diagnosestellung gut zusammen: *“Being diagnosed gives permission to be ill. What was previously a complaint is now a disease.”*<sup>7</sup>

Wird eine Diagnose gestellt, ergeben sich daraus häufig weitreichende Konsequenzen. Es wird ein Behandlungsrahmen abgesteckt, der einen Therapieplan und eine voraussichtliche Prognose enthält. Doch neben den medizinischen Aspekten ergeben sich weitere relevante Folgen für den Patienten: Hat ein Arzt eine Diagnose gestellt, so besteht damit oftmals Anspruch auf die Kostenübernahme der Medikamente und der Behandlung.<sup>7</sup>

Dem Arzt kommt somit eine entscheidende Rolle zu: ihm obliegt die Verantwortung bei der korrekten Diagnosestellung, weshalb diese auch nur nach sorgfältiger Diagnostik erfolgen sollte.<sup>9</sup>

### Selbstdiagnose

Neben der ärztlichen Diagnostizierung existiert das Phänomen der „Selbstdiagnose“. Dieser Terminus ist in der Literatur nicht definiert. Im Folgenden soll darunter die „Diagnose“ verstanden sein, die sich der Patient selbst zugesprochen hat. Dabei kann es sich sowohl um eine bloße Vermutung, als auch um die feste Überzeugung eines Patienten handeln, von einer Erkrankung betroffen zu sein. Zudem beschreibt der Begriff der „Selbstdiagnose“ nicht,

über welches Medium der Patient zu dieser Überzeugung gekommen ist (Internet, andere Patienten, Hörensagen, Print-Medien).

Während 70% der Patienten, die einen Allgemeinarzt aufsuchen, nur mit einem Leitsymptom kommen, gibt es zunehmend Patienten, die dem Arzt ihre vorüberlegte Diagnose vorstellen<sup>10</sup>. Von diesen Erkrankungen gibt es einige, wie beispielsweise eine unkomplizierte Zystitis, die, aufgrund eventueller Zystitiden in der Vorgeschichte oder einer hohen Prävalenz, leicht selbst zu diagnostizieren sind<sup>10,11</sup>. Andere komplexe oder seltene Krankheitsbilder können andererseits meist nicht sicher von Laien erkannt werden<sup>10</sup>.

Die Informationen, die zur Selbstdiagnose führen, entstammen dabei unterschiedlichsten Quellen: Patienten können auf eigene Krankheitserfahrungen oder die aus dem Bekannten- und Verwandtenkreis zurückgreifen. Zudem sind Informationen leicht über diverse Medien, wie Bücher oder das Internet verfügbar. Derzeit lässt sich ein Trend hinsichtlich der Nutzung von Selbstdiagnoseprogrammen *online* oder vermittels von sog. *Apps* feststellen.<sup>12,13</sup> Darauf wird in Kapitel 3.2 genauer eingegangen.

Die Gefahren selbst gestellter Diagnosen liegen auf der Hand: Es können schwerwiegende Erkrankungen übersehen werden, indem sie als harmlos abgehandelt werden. So kann beispielsweise ein Herzinfarkt fälschlicherweise für eine Magenentzündung gehalten werden oder eine rektale Blutung aufgrund eines kolorektalen Karzinoms für harmlose Hämorrhoiden. Doch Selbstdiagnosen können gelegentlich auch von Vorteil sein: So wurde festgestellt, dass (ansonsten nicht aufgedeckte) Diagnosehypothesen in 18% bei Hausärzten dadurch eingeleitet wurden, dass Selbstdiagnosen vorgeschlagen wurden.<sup>10</sup>

Aus diesen Ausführungen lässt sich somit durchaus ein Bedarf an fundierten, frei zugänglichen und leicht bedienbaren Selbstdiagnoseprogrammen ableiten. Durch solche wissensbasierte Systeme könnte möglicherweise die Rate an eigenen Fehlinterpretationen der Patienten oder irreführenden Informationen aus verschiedenen Medien reduziert werden.

### **3 Wissensbasierte Systeme in der Medizin – ein Überblick**

Wissensbasierte Systeme sind in der Medizin in den verschiedensten Bereichen in Gebrauch. Im folgenden Abschnitt werden die wichtigsten Entscheidungsunterstützungssysteme und deren Einsatzgebiete kurz vorgestellt. Dabei gilt es zwischen Programmen zu unterscheiden, die für den alltäglichen Gebrauch, also für ein Patientenkontingent, gedacht sind, und anderen klinischen Programmen (CDSS), die für Ärzte konzipiert werden<sup>14</sup>.

#### **3.1 Clinical decision support systems (CDSS)**

Medizinische Entscheidungsunterstützungssysteme existieren für verschiedene Zwecke. Dabei lässt sich eine Einteilung in 3 Gruppen vornehmen:

1. Systeme, die eine auf evidenzbasierter Medizin fundierte integrierte Qualitätssicherung bieten. Hierbei handelt es sich z.B. um statistische Qualitätskontrolle, intelligente Dokumentationssysteme und den Einsatz medizinischer Leitfäden einschließlich deren Umsetzung.
2. Computerized Order Entry Systeme (CPOE) zu nennen, die beispielsweise in der Lage sind, auf Medikamentenverschreibungen oder Labortestbefehle hinzuweisen.
3. Diagnostische Entscheidungsunterstützungssysteme oder Clinical decision support systems (CDSS). Diese werden nachstehend genauer behandelt, da sie für diese Dissertationsarbeit eine zentrale Rolle spielen.<sup>14</sup>

##### **3.1.1 Definition**

Diagnostische Entscheidungsunterstützungssysteme sollen Ärzte dabei unterstützen, im Klinikalltag die korrekte Entscheidung bezüglich der Diagnose und der damit verbundenen Therapie zu treffen. Dies läuft folgendermaßen ab: Der Arzt gibt die Symptome und/oder diverse diagnostische Werte (z. B. Laborwerte) in das System ein. Dieses CDSS basiert auf einer Wissensbasis, die Informationen bezüglich zahlreicher Diagnosen und deren zugehörigen Symptome

verknüpft. Diese Wissensbasis wurde, wie die meisten gegenwärtigen wissensbasierten medizinischen Systeme, anhand Wenn-Dann-Regeln mit Gewichtungen, extrahiertem Wissen aus medizinischer Fachliteratur oder Leitlinien erstellt. Heutzutage können Diagnosen auch auf der Basis der Auswertung großer Fallsammlungen bewertet werden. Somit ist es vielen Programmen möglich, aus Fällen zu lernen, wenn genügend Fälle vorhanden sind. Mögliche Verfahren hierzu sind fallbasiertes Schließen (Similarity learning; k-nearest neighbor), statistisches Schließen (Theorem von Bayes und Bayessche Netzwerke), subsymbolische Lernverfahren (z.B. Support Vektor Machines, deep neural networks).<sup>15-17</sup> Eine gute Übersicht mit weiteren Literaturangaben findet sich auf [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning) und im Werk „Data Mining“<sup>18</sup>. Nach Dateneingabe berechnet das System regelbasiert die Übereinstimmung der Symptome und Werte mit den Diagnosen und generiert eine Liste mit Verdachtsdiagnosen, die es dem Arzt aufzeigt. Daraufhin kann der Arzt diese Ergebnisse mit seiner eigens aufgestellten Verdachtsdiagnose vergleichen, kann in seiner Diagnose bestätigt werden und wird zudem auf mögliche andere zutreffende Erkrankungen aufmerksam gemacht. Zudem kann er auch einige Verdachtsdiagnosen aufgrund von zusätzlichen Patienteninformationen, die nicht vom System erfasst wurden, leicht ausschließen.<sup>19</sup>

### **3.1.2 Entwicklungen und Standards**

Der Beginn der Entwicklung medizinischer Entscheidungssysteme findet sich in den 70er Jahren. Damals zeigte sich, dass es möglich war, komplexe CDSS zu entwickeln. Um einen Einblick in die ersten Systeme zu erhalten, werden nun zwei Entscheidungsunterstützungssysteme vorgestellt: Das Leeds Abdominal Pain Programm und MYCIN. Hier muss jedoch darauf hingewiesen werden, dass beide bis heute keinen Einsatz im Klinikalltag fanden, obwohl sie in Evaluationen gute Ergebnisse erzielten.<sup>14,20</sup>

#### **Leeds Abdominal Pain Programm**

Im Jahr 1971 wurde in einer Studie<sup>21</sup> überprüft, ob ein Programm, welches speziell auf abdominellen Schmerz ausgerichtet war, eine höhere Rate an

korrekten Diagnosen stellen konnte, als Ärzte in der Klinik. Man wollte damit eine Antwort auf diese Fragen erhalten: „*[D]o clinicians actually need such a system? [A]nd can it offer any measurable advantage in terms of diagnostic accuracy and reliability over more conventional methods?*“<sup>21</sup> Dabei wurden die Enddiagnosen von 304 Patienten mit akuter abdomineller Symptomatik mit den Diagnosen des Leeds Abdominal Pain Programms verglichen, die auf Aufnahme-Symptomen des Patienten basieren, sowie den Initialdiagnosen der Ärzte. Dabei stellte sich heraus, dass die diagnostische Treffergenauigkeit beim Computersystem (92%) signifikant höher war, als die der am meisten erfahrensten Ärzte der Klinik (80%). Daraus leitete man ab, dass die Implementierung dieses Programms auch einen Nutzen im Klinikalltag ergeben würde.<sup>21</sup> Diese Vermutung wurde 1986 auch durch eine multizentrische Studie bestätigt<sup>22</sup>. Dennoch weisen de Dombal und seine Kollegen strikt darauf hin, dass dieses System nicht dafür gedacht ist, eine Diagnose selbstständig zu stellen. Vielmehr soll es lediglich die Ärzte bei deren Arbeit unterstützen.<sup>21</sup>

Ergänzend ist anzumerken, dass dieses Programm nie in den klinischen Einsatz kam<sup>14</sup>. Die Gründe hierfür könnten ähnlich sein wie beim Programm MYCIN<sup>23</sup>, das nachstehend charakterisiert wird.

## **MYCIN**

MYCIN wurde an der Universität von Stanford entwickelt<sup>23</sup>. Mark Musen beschreibt es folgendermaßen: „*[MYCIN is] the first convincing demonstration of the power of the rule-based approach in the development of robust clinical decision-support systems.*“<sup>23</sup> Anders als das Leeds Abdominal Pain Programm, erstreckt sich sein (geplantes) Einsatzgebiet auf die Diagnostik und Therapie bakterieller Infektionskrankheiten. Das heuristische System enthält Regeln von formuliertem Fachwissen. Es soll ebenso wenig wie das Leeds Abdominal Pain Programm die Diagnose eigenständig festlegen, sondern Ärzten bei der korrekten Diagnostizierung von Mikroorganismen und der Wahl einer adäquaten antimikrobiellen Therapie behilflich sein.<sup>20,24</sup>

Da von Beginn einer Blutkulturentnahme bis hin zum Ergebnis der Blutkultur gewöhnlich zwischen 24 bis 48 Stunden vergehen, müssen sich die Ärzte für ein

wahrscheinlich zutreffendes Erregerspektrum entscheiden und mit einer antibiotischen Breitbandtherapie beginnen. Erst nach Ergebnis der Blutkultur und einer Resistenztestung können sie sich dann auf eine spezielle antibiotische Therapie festlegen.<sup>25-27</sup> Mittels leicht zugänglichen Expertenwissens soll MYCIN die therapeutische Unsicherheit verbessern. Das Programm lässt sich in drei Unterprogramme unterteilen: a) ein Konsultationssystem, das durch die Abfrage von Patienteninformationen einen Medikamentenvorschlag für ein voraussichtliches Erregerspektrum vorschlägt, falls MYCIN eine Behandlung überhaupt für notwendig hält; b) ein Erklärungssystem, das dem Benutzer die Hintergründe der vom Programm getroffenen Entscheidung erläutert; c) ein Programm zur Regelübernahme oder -korrektur durch Experten. Somit ist das MYCIN fähig, neue Regeln zu erlernen. Zudem ist MYCIN nicht auf eine vollständige Informations-eingabe angewiesen und kann auch auf der Grundlage recht weniger Informationen eine relativ exakte Empfehlung abgeben.<sup>24,28</sup>

Ein wichtiger limitierender Faktor für den Erfolg dieses Programms war jedoch die mangelnde Initiative der Ärzteschaft, MYCIN regelmäßig zu nutzen. Außerdem waren aber auch viele Computer der Kliniken für die Softwarefunktion nicht leistungsstark genug und deren Präsenz und Wichtigkeit im Alltag in den siebziger Jahren wesentlich geringer als heute. So hat MYCIN, obwohl es ein ausgeklügeltes, medizinisches Entscheidungsunterstützungssystem darstellt, keinen breiten Anklang bei der Ärzteschaft gefunden.<sup>24,28</sup>

### 3.1.3 Evaluation

In den folgenden Jahren wurden die diagnostischen Entscheidungsunterstützungssysteme weiterentwickelt. Diese waren nun nicht mehr nur auf ein bestimmtes medizinisches Gebiet, wie beispielsweise Infektionskrankheiten, beschränkt, sondern umfassten ein breites Spektrum an Diagnosen.<sup>14</sup>

Im Jahr 1994 wurde in einem Artikel im New England Journal of Medicine die Ergebnisse einer ersten umfänglichen Evaluation von vier kommerziell verfügbaren CDSS durchgeführt<sup>29</sup>. Dazu wurden die Programme Dxplain<sup>30</sup>, Iliad<sup>31</sup>, Meditel<sup>32</sup> und QMR<sup>33</sup> zur Differentialdiagnose für den Bereich der Inneren Medizin anhand von 105 Testfällen im Hinblick auf ihre Sensitivität geprüft<sup>29</sup>.

Diese Programme waren als Unterstützung der Ärzte im Klinikalltag gedacht<sup>14</sup>. Hierbei tauchte die korrekte Diagnose in 52-71% der Fälle in der Liste möglicher Diagnosen auf. Wurden nur die 63 Diagnosen berücksichtigt, die alle vier Programme erkannten, verbesserte sich diese Quote auf 71-89%. Im Durchschnitt erschienen die korrekten Diagnosen in der Liste der vom Programm für möglich gehaltenen Diagnosen an sechster bis dreizehnter Stelle. Darüber hinaus wurde die jeweils korrekte Diagnose zu etwa 50% unter den besten 20 Rängen und zu etwa 10-20% auf Rang 1 platziert. Von den jeweils 20 wahrscheinlichsten Verdachtsdiagnosen wurden 19-37% als relevant eingestuft. Die Autoren folgerten, dass diese Programme für Ärzte von Nutzen sein können, wenn sie unter den präsentierten Informationen relevante von irrelevanten unterscheiden können.<sup>14,29</sup>

Heute existieren weiterentwickelte Entscheidungsunterstützungssysteme, die auf dem oben erwähnten MYCIN beruhen. Beispiele hierfür sind ISABEL, ein pädiatrisches Differentialdiagnoseprogramm, oder GIDEON, ein Entscheidungsunterstützungssystem für Infektionskrankheiten. Diese wurden ebenfalls, ähnlich zur oben erläuterten Studie, getestet und als nützlich eingestuft.<sup>34-37</sup> Dabei lag der Fokus der Evaluation auf der Platzierung der korrekten Diagnose. Darüber hinaus existieren weitere Studien, die sich mit dem Einfluss solcher Programme auf Ärzte befassen.<sup>38,39</sup>

Als Schlussfolgerung des Artikels im New England Journal of Medicine erschien folgende Aussage der Autoren: „*The programs should be used by physicians who can identify and use the relevant information and ignore the irrelevant information that can be produced.*”<sup>29</sup> Somit sollen diese Programme Ärzten nur bei der korrekten Diagnosestellung behilflich sein, die Ärzte selbst jedoch nicht durch eigenständige Diagnosestellung ersetzen. Obwohl klinische Entscheidungsunterstützungssysteme laut Studien gute Ergebnisse erzielen und somit wortwörtlich eine Unterstützung für Ärzte darstellen können, stehen auch heutzutage viele Ärzte den Programmen (noch immer) mit Skepsis gegenüber. Zusätzlich zu den bereits beschriebenen Ursachen könnte ein weiterer Grund sein, dass die Ärzte den tatsächlichen Nutzen dieser Programme zum Teil deutlich unterschätzen.<sup>14</sup>

### 3.2 Systeme für das Patientenklintel

Doch wissensbasierte medizinische Systeme sollten nicht nur auf den Klinik-einsatz und damit lediglich auf den Gebrauch durch Ärzte beschränkt sein. Ein wachsender Anteil der Bevölkerung nutzt das zunehmend ubiquitär verfügbare Internet als Informationsquelle. Doch dabei soll es nicht bleiben: Seit Einführung von Smartphones wurde ein regelrechter *App-Boom* verzeichnet. Bei den Anbietern Google Play und dem App Store von Apple wurden Mitte 2015 jeweils über eine Million *Apps* zum Download angeboten. Dabei besitzen Medizin- und Gesundheitsapps einen bedeutenden Marktanteil und zeugen von hoher Popularität.<sup>13,40-42</sup> Doch wie kommt es, dass gerade diese Art von Apps so gefragt ist?

Dafür sollte man sich vor Augen führen, dass sich das Arzt-Patienten-Verhältnis in den letzten Jahrzehnten stark gewandelt hat. Seit sich ein zunehmendes Gesundheitsdenken in den 1970er Jahren entwickelte, wurde den Patienten mehr und mehr zu einer gesünderen und bewussten Lebensweise geraten. Die Patienten sollen sich aktiv mit ihrem Gesundheitszustand befassen, indem sie sich beispielsweise eigenständig Informationen beschaffen, und es nicht nur bei der bloßen Annahme von ärztlichen Ratschlägen belassen. Folglich wird der Patient aktiv zu mehr Verantwortungsbewusstsein seinem Gesundheitszustand gegenüber aufgefordert. Das Arzt-Patienten-Verhältnis veränderte sich somit von einem ärztlich dominierten hin zu einem ausgeglicheneren Verhältnis. Dabei sollen sowohl Arzt als auch Patient gleichermaßen in den Entscheidungsprozess mit eingebunden werden.<sup>7,43,44</sup>

Aus dieser Aufforderung zu mehr Mitwirkung resultiert, dass immer mehr Menschen das Internet zur Informationsbeschaffung über medizinische Sachverhalte nutzen. Auf sehr einfache Weise kann man eine Fülle an Informationen über diverse Webseiten oder Internetforen erhalten, indem man lediglich einige Stichwörter in eine Suchmaschine eingibt. Früher noch wurde stets der Arzt für medizinische Fragen regelmäßig konsultiert und auf dessen Ratschlag gehört. Heutzutage hingegen wird oftmals das Internet als erste Instanz verwendet, um beispielsweise eine Erklärung für aktuelle Beschwerden zu erhalten. Haben

Patienten dann eine mögliche, auf sie zutreffende Diagnose im Internet gefunden, dient der Arzt im Anschluss dazu, diese zu be- oder entkräftigen.<sup>13</sup>

Fox und Duggan fanden heraus, dass 35% der US-Amerikaner das Internet benutzen, um eine Erklärung für ihre Beschwerden zu erhalten. Von diesen 35% suchen 50% danach ärztlichen Rat auf.<sup>12</sup>

Aus diesem Trend heraus resultieren zahlreiche Webseiten und Apps für das Smartphone, welche dem Patienten suggerieren, sich selbstdiagnostizieren zu können. Lupton und Jutel veröffentlichten einen Beitrag im Social Science & Medicine zum Thema Selbstdiagnose-Apps. Dabei stellte sich u.a. in einer Analyse solcher Apps heraus, dass populäre Programme, wie „WebMD“<sup>45</sup> und „iTriage Health“<sup>46</sup>, von Millionen Smartphone Benutzern heruntergeladen werden.<sup>13</sup>

Der folgende Abschnitt liefert einen Überblick über aktuelle Apps zur Selbstdiagnose, untersucht deren Wirkung auf die Benutzer und soll sich kritisch mit den Programmen auseinandersetzen.

### **3.2.1 Analyse der Beliebtheit**

Warum sind Apps bei der Bevölkerung so beliebt? Eine mögliche Erklärung ist, dass die meisten Smartphonebesitzer ständig über eine Internetverbindung verfügen und somit Plattformen wie der Apple App Store allzeit zugänglich sind. Selbstdiagnose-Apps können bei Bedarf direkt heruntergeladen und genutzt werden. Zudem können diese Programme mit nur einer Fingerberührung gesteuert werden, wodurch sie im Vergleich zu Online-Programmen oder Internetforen noch leichter zugänglich sind.

In dem oben genannten Paper von Lupton und Jutel analysieren diese auch die Gründe der großen Beliebtheit der Selbstdiagnose-Apps<sup>13</sup>: Betrachtet man alleine die Namen dieser Apps – Pocket Doctor, WebMD, Virtual Doctor etc. – fallen schnell Schlagworte medizinverbundener Termini wie „Doctor“ auf. Die daraus hervorgehende Werbebotschaft liegt auf der Hand: Die Suggestion, es handele sich um professionelle und medizinfundierte Hilfestellungen bei der Diagnosestellung.<sup>13,47</sup>

Bei der genaueren Betrachtung der einzelnen Apps, fällt zudem eine bestimmte Werbestrategie auf: Dem Kunden wird durch die Nutzung solcher Apps Folgendes versprochen: „[The Apps] will save lay people's time and money (by potentially allowing them to avoid a visit to the doctor), allay their anxieties, improve their health by allowing them to diagnose a medical condition and then seek treatment, educate them by enhancing their medical knowledge and support patient empowerment by bestowing information about diagnoses.“<sup>13</sup>

Zudem suggerieren viele Werbeslogans dem Kunden, dass dieser sich auf die Ergebnisse und Ratschläge verlassen könne, da führende Ärzte und Softwareentwickler an dem Schaffensprozess der Apps beteiligt waren. Es wird damit geworben, dass die Programme auf der Basis der besten medizinischen und technischen Kenntnisse konzipiert wurden.<sup>13</sup>

So wird der „Isabel Symptom Checker“<sup>48</sup> folgendermaßen beschrieben: “Using the latest searching technologies, the system can take a pattern of symptoms in everyday language and instantly compute from our vast database of 6,000 diseases, the most likely ones.“<sup>13,48</sup>

Eine andere Marketingstrategie setzt darauf, bestimmte Zielgruppen direkt anzusprechen. Als Beispiel kann hier „Dr. Moms - Treatment Guide“<sup>49</sup> herangezogen werden. Diese App richtet sich direkt an Mütter, die als „doctor of the family“<sup>49</sup> agieren, und soll ihnen bei Krankheitsfällen in der Familie eine Hilfe sein. Mütter werden dadurch beraten, welche Symptome sie noch zu Hause behandeln können und wann sie einen Arzt aufsuchen sollten.<sup>13,49</sup>

Aufgrund der eben erwähnten Punkte wird der Anschein erweckt, dass man durch die Nutzung solch eines Programms bedenkenlos auf einen Arztbesuch verzichten kann, wenn die App einem dazu geraten hat. Dieser Werbespruch fasst die durch diese App eventuell entstehende Suggestion gut zusammen: „It's like having a physician in your pocket!“<sup>13</sup>

### 3.2.2 Negative Kritikpunkte

#### Mangelnde Quellenangaben

Möchte der Benutzer jedoch herausfinden, wie medizinisch fundiert diese Programme sind, so bleiben ihm diese Informationen, bis auf einige Ausnahmefälle, meist verwehrt. Man findet bei den meisten Apps keinerlei direkten Bezug zur Wissensquelle, auf der die jeweilige App basiert. Interessiert man sich beispielsweise konkret für die Namen der Entwickler, sind diese trotz Recherche oftmals ungewiss.<sup>13</sup> Hamilton und Brady bemerkten, dass in diesem Fall meist nur von „Ärzten“ oder einem „medizinischen Team“ geredet wurde<sup>50</sup>. Dies lässt sich auch anhand des Beispiels einer App namens „Best Android Symptom Checker“<sup>51</sup> verdeutlichen: Die App wird damit umworben, dass deren Inhalt auf dem Wissen der Harvard Medical School beruhe. Bei dieser oberflächlichen Information soll es jedoch bleiben, es wird nirgendwo auf einzelne Namen der Entwickler oder Arbeitsgruppen verwiesen.<sup>13,51</sup>

Ein anderer Kritikpunkt an einigen dieser Apps ist das „Zusatzangebot“ an weiteren Apps, die von den gleichen Entwicklern angeboten werden aber keinen medizinischen Bezug aufweisen. Exemplarisch sollen hier Apps mit Buddha Zitaten oder Finanzrechnern aufgeführt werden. Dadurch wirken die Selbstdiagnoseprogramme bei genauerem Hinsehen schnell unglaubwürdig. Denn, wenn dieselben „Experten“ eine breite Palette an Apps anbieten, die sowohl laienhafte Gesundheitstipps als auch Smartphone-Games bis hin zu Medizin-Apps enthält, kommen rasch Zweifel an einer sorgfältig recherchierten Entwicklung auf.<sup>13</sup>

#### Versteckte Kommerzialisierung

Ein weiterer zu kritisierender Aspekt sind die auf den ersten Blick oftmals nicht direkt sichtbaren kommerziellen Zwecke der Apps. Noch während der Benutzung vieler Apps wird dem User nicht selten eine Weiterbehandlung durch gewisse namentlich genannte Ärzte vorgeschlagen oder zum Abschluss einer zusätzlichen Gesundheitsversicherung geraten. Andere Apps bieten einen kostenlosen Online-Gesundheitsreport an, falls die Benutzer zuvor deren Newsletter abonnieren. Anders als in dem genannten Werbespruch „It's like having a physician in your pocket“ versucht wird glaubhaft zu machen, tendieren die Apps

nicht dazu von einem Arztbesuch abzuraten, sondern werben im Gegenteil sogar für bestimmte Ärzte.<sup>13</sup>

### **Mangelhafter Datenschutz**

Datenschutz ist ebenfalls ein Faktor, der hier kurz behandelt werden soll. Lupton und Jutel beschrieben anhand verschiedener Selbstdiagnose-Apps, welche unterschiedliche Bedeutung dem Datenschutz zukommt<sup>13</sup>. Bei dem „Best Android Symptom Checker“<sup>51</sup> werden Benutzerdaten und Standortdaten zur Verbesserung der Benutzerfreundlichkeit und des Systems gesammelt. Allerdings werden die Daten laut Angabe nicht an Dritte weitergegeben. Bei dem populären Programm „WebMD“<sup>45</sup> hingegen, erhält der Nutzer trotz Datenschutzerklärung mit Hinweis auf Datenweitergabe keinerlei Informationen zum Zweck der Weitergabe. Es wird lediglich erklärt, dass die Daten an Dritte geleitet werden, die mit dem Unternehmen in Verbindung stünden.<sup>13</sup> Als Beispiel für äußerst mangelhaften Datenschutz sticht die App „iTriage Health“<sup>46</sup> hervor. Hierbei werden Personendaten zu Benutzergewohnheiten und Standortinformationen gesammelt und diese Daten zu Werbezwecken und zur Optimierung von Marketingzielen verwendet.<sup>13</sup>

Zusammenfassend lässt sich sagen, dass nutzerbezogene Daten durch Gebrauch der Apps oft gespeichert und weitergegeben werden, indem die Apps benutzt werden<sup>13,52</sup>. Zudem liegen leider teilweise nur wenig detaillierte Datenschutzerklärungen für die Benutzer vor<sup>13</sup>.

### **3.2.3 Positive Kritikpunkte**

#### **Auseinandersetzung mit dem Gesundheitszustand**

Neben den genannten negativen Kritikpunkten sollte man den eigentlichen Zweck solcher Apps nicht ungeachtet lassen. Der Werbeslogan der „Rapid Diagnosis“<sup>53</sup> App bringt deren Nutzen gut auf den Punkt: „[T]he primary objective of this software is to encourage an active interest in health related problems, their diagnosis and treatment, and to empower patients by providing them with accurate, up to date knowledge, so that they may understand and participate in their health care.“<sup>53</sup>

Es wird Folgendes klar: Selbstdiagnose-Apps zielen nicht darauf ab, dem Benutzer eine verlässliche Diagnose zu stellen. Es geht nicht darum, einen Arzt zu ersetzen.<sup>40,41</sup>

Um die Benutzer eben darauf hinzuweisen, beinhalten die Apps spezielle Hinweise, wie beispielsweise folgenden der „Symptify“<sup>54</sup> App:

„Symptify.com does not provide medical advice and it is NOT intended for medical diagnostic purposes. All material and content (i.e. text, images, videos, graphics or information) contained on Symptify.com are not intended to be a substitute for professional medical advice, diagnosis or treatment. If you believe you have a true medical emergency then you should call your doctor or dial 911 immediately. Symptify.com does NOT contain all causes of disease or possible injuries, and the percentages listed do NOT reflect the true probability of having a certain medical condition. Only a physician or a licensed healthcare provider may diagnose and treat injuries or disease.“<sup>54</sup>

Darüber hinaus werden die vorgeschlagenen Diagnosen auch meist bewusst nur „mögliche“ oder „wahrscheinliche“ Diagnosen genannt<sup>13</sup>. Dadurch werden die Benutzer stets daran erinnert, dass es sich bei den Apps nur um Programme handelt, die in ihrem Umfang beschränkt sind.

Entwickler dieser Selbstdiagnosesysteme distanzieren sich dadurch auch bewusst von der Aussage, solche Systeme könnten die Monopolstellung des Arztes angreifen. Denn obwohl diese Apps die eingegebenen Symptome völlig objektiv bewerten können, liegen deren diagnostische Möglichkeiten weit unter denen eines Arztes. Einen Arzt können die Programme nicht ersetzen.<sup>13,41</sup>

Aus diesem Grund geben einige Apps zusätzlich noch an, dass sie entweder nur zur Unterhaltung gedacht sind oder dass Nutzer dieser Programme zur Verifizierung der erhaltenen Informationen einen dazu qualifizierten Arzt aufsuchen sollten<sup>13</sup>.

Ein Nutzen dieser Gesundheits-Apps bleibt: Diese Produkte können dazu beitragen, dass sich die Bevölkerung mehr mit ihrem Gesundheitszustand auseinandersetzt.

### **Verdachtsgenerierung anhand einer medizinischen Wissensbasis**

Programme zur Selbstdiagnose beinhalten, anders als Medizinfores oder Suchmaschinen, medizinisches Wissen, das zu einem einzigen System zusammengefasst wird. Folglich werden die vom Patienten eingegebenen Informationen mit großen medizinischen Wissensbasen anhand von kalkulierten

Bewertungsschemata verarbeitet. Selbstverständlich können Fehler bei der Symptomabfrage entstehen, indem Benutzer Fragen missverstehen oder nicht zutreffende Symptome auswählen. Doch können solche Algorithmen in der Lage sein, Diagnosevorschläge unbeeinflusst abzugeben.<sup>13</sup>

### 3.2.4 Evaluation von Selbstdiagnoseprogrammen

Aktuell existiert ein breites Spektrum an Selbstdiagnosesystemen in Form von Apps oder Onlineprogrammen. In einer Evaluation der Harvard Medical School wurde erstmals die Genauigkeit der Diagnosestellung anhand von 19 Selbstdiagnoseprogrammen getestet. Zusätzlich dazu wurden Programme geprüft, welche dem Nutzer lediglich oder zusätzlich zur Verdachtsgenerierung einen Ratschlag hinsichtlich der Notwendigkeit eines Arztbesuches geben.<sup>55,56</sup> Auf diese Ergebnisse wird im Folgenden jedoch nicht eingegangen.

Die Genauigkeit der Verdachtsdiagnosengenerierung wurde anhand von 45 fiktiven Patientenfällen an 19 Programmen in Form von Apps getestet. Diese Programme sind, wie *SymptomCheck* auch, öffentlich frei zugänglich und nicht nur auf einen medizinischen Fachbereich, wie z. B. Kardiologie, begrenzt. Unter den getesteten Diagnosen gab es sowohl häufige, als auch weniger häufige Diagnosen, welche auf den Prävalenzen ambulanter Arztbesuche in den USA basierten. Anders als in dieser Studie zu *SymptomCheck*, wurden in dieser Evaluation keine Patienten rekrutiert.<sup>55</sup>

Durchschnittlich wurde die korrekte Diagnose von den Programmen in 34% auf dem ersten Platz gelistet. Hierbei gab es jedoch große Unterschiede hinsichtlich der Genauigkeit der Diagnoseerkennung. Während die korrekte Diagnose vom Programm „MEDoctor“<sup>57</sup> nur in 5% an erster Stelle erschien, erfolgte dies durch „DocResponse“<sup>58</sup> in 50%. Betrachtet man die ersten drei Diagnoseränge, so erschien dort die zutreffende Diagnose in 51% und innerhalb der wahrscheinlichsten 20 Diagnosen in 58% der Fälle. Doch auch in diesen Fällen fanden sich Schwankungen zwischen den einzelnen Programmen. Desweiteren sehen die Autoren dieses Papers solche Programme zur Selbstdiagnose einem Arzt zwar unter-, doch Suchmaschinen wie Google im alltäglichen Gebrauch weit überlegen.<sup>55</sup>

Diese Evaluation der Harvard Medical School war die erste, welche die Genauigkeit der Verdachtsgenerierung solcher Programme testete. Da diese Studie allerdings nur anhand von Fallbeispielen durchgeführt wurde, lässt sich darüber hinaus nichts über die tatsächliche Genauigkeit der Programme im alltäglichen Gebrauch sagen.<sup>55</sup>

Es wird angenommen, dass die im folgenden dargestellte Studie zu *SymptomCheck* die erste ist, in welcher die Verdachtsgenerierung an einem selbstentwickelten Programm direkt an Probanden und Patienten getestet wurde.

## Material und Methoden

### 4 Entwicklung der Wissensbasis

#### 4.1 Software

Für die Entwicklung der Wissensbasis und der Startseite von *SymptomCheck* verwendeten wir Microsoft Excel 2013. Diese Dateien wurden im Anschluss von Herrn Prof. Dr. Frank Puppe und den Mitarbeitern seines Lehrstuhls für künstliche Intelligenz und angewandte Informatik in ein Online-Programm umgewandelt und hochgeladen.

#### 4.2 Diagnosen

Ursprünglich beinhaltete *SymptomCheck* 241 Diagnosevorschläge [Tabelle 34]. Jedoch stellte sich durch die Pilotstudie I heraus, dass es sinnvoller ist, bestimmte Diagnosen zusammenzufassen und seltene Diagnosen komplett aus dem Programm herauszunehmen. So wurden beispielsweise diverse Tachykardieformen zur Diagnose „tachykarde Herzrhythmusstörung“ zusammengefasst. Dadurch erreicht die zutreffende Verdachtsdiagnose eine höhere Platzierung in der Rangfolge. Für den Laienbenutzer soll somit die Liste der Verdachtsdiagnosen geordneter und wesentlich übersichtlicher erscheinen. Aktuell beinhaltet *SymptomCheck* 178 Diagnosevorschläge, die grob in 18 Gruppen eingeteilt werden [siehe Tabelle 1]:

Tabelle 1: Anzahl der den Fachgebieten zugeordneten Diagnosen

Fachgebiet	Anzahl (n)	Fachgebiet	Anzahl (n)
Kardiologie	24	Psychiatrie	4
Pulmologie	16	Urologie	4
Gastroenterologie	28	Neurologie	5
Hämatonkologie	6	Gynäkologie	2
Endokrinologie	16	Dermatologie	1
Hepatology	2	HNO	8
Angiologie	9	Stoffwechsel	1
Infektiologie	11	Nicht pathologisch	22
Orthopädie	8	Sonstiges	11

Der Schwerpunkt liegt bei *SymptomCheck* auf internistischen Erkrankungen. Erkrankungen anderer Bereiche werden entweder nicht erfasst oder nur als entweder „urologische, neurologische, orthopädische oder dermatologische Erkrankung“ zusammengefasst [siehe **Tabelle 35**]. Dermatologische Diagnosen werden nicht im Einzelnen, sondern unter dem Begriff „dermatologische Erkrankung“ aufgeführt, da deren Symptomatik häufig sehr ähnlich bis gleich ist und eine genaue Abfrage durch das Programm vom Laien schwer zu beantworten ist. Dem Benutzer wird lediglich der Hinweis gegeben, dass es sich um eine dermatologische Diagnose handeln könnte.

Bei der Auswahl der Diagnosen konnten nicht alle internistischen Diagnosen erfasst werden. Die reine Selektion nach Inzidenzen gestaltete sich als schwierig, da für viele in der alltäglichen Praxis relevante Diagnosen keine Inzidenzen verzeichnet sind. Hierbei diente das Werk „INNERE MEDIZIN 2013“<sup>59</sup> von Gerd Herold als Selektionswerkzeug.

In jedoch allen außerinternistischen Gruppen außer der Dermatologie werden zusätzlich ICD-10 zugehörige Diagnosen aufgeführt, welche eine hohe Prävalenz besitzen. Als Beispiele können hier die Zystitis oder eine Schwangerschaft genannt werden. Obwohl beide Geschlechter von einer Zystitis betroffen sein können, wurde diese aufgrund ihrer hauptsächlich hohen Prävalenz bei Frauen hier formal zu den gynäkologischen Erkrankungen gezählt<sup>60,61</sup>.

Von den 178 Diagnosen sind jedoch nicht alle der ICD-10 Klassifikation zuzuordnen. Neben Diagnosen wie beispielsweise einer Pneumonie, enthält *SymptomCheck* auch „Diagnosen“ wie „Bewegungsmangel“ oder „schlechte Kondition“ [siehe **Tabelle 35**]. Solche Vorschläge sind keine Diagnosen im klassischen Sinne, jedoch häufige, nicht pathologische Erklärungen für in der Bevölkerung auftretende Beschwerden. Da dieses Programm seinen Benutzern gegebenenfalls auch eine Entwarnung geben soll, sind solche „Diagnosevorschläge“ wichtiger Bestandteil von *SymptomCheck*.

### 4.3 Symptome

Ein für den Erfolg von *SymptomCheck* essentieller Arbeitsschritt bei der Entwicklung war die Auswahl von möglichen Symptomen. Ursprünglich war geplant, ein

breites Spektrum an Symptomen zu sammeln, ehe mit der Integration der einzelnen Diagnosen in die Wissensbasis begonnen werden sollte. Für einen ersten Überblick über mögliche unentbehrliche Diagnosen dienten zahlreiche medizinische Fachbücher zum Thema Differentialdiagnose, welche mögliche, auf ein Leitsymptom zutreffende Differentialdiagnosen bieten<sup>62-67</sup>. Diese waren oftmals allerdings nur auf einige wichtige Symptome wie beispielsweise „Brustschmerz“ oder „Dyspnoe“ beschränkt. Da ein Ziel jedoch darin bestand, eine möglichst große Bandbreite an Symptomen zu bieten, musste ein anderes Konzept zur Symptomauswahl erarbeitet werden. Folglich wurden vor allem anhand der Fachbücher „INNERE MEDIZIN 2013“<sup>59</sup> und „INNERE MEDIZIN 2014“<sup>68</sup> von Gerd Herold die einzelnen internistischen Erkrankungen bezüglich ihrer Symptome analysiert und die Symptomliste schrittweise ergänzt. Im Anschluss erfolgte die zusätzliche Ergänzung durch Folgefragen auf Symptome, wenn deren Informationen zur Differenzierung der Diagnosen essentiell waren. Nach Abschluss des Entwicklungsprozesses enthält *SymptomCheck* nun 98 unterschiedliche Leitsymptome. Deren Gruppierung wird im nachfolgenden Kapitel noch genauer erläutert. Von diesen 98 Leitsymptomen können 50 noch genauer abgefragt werden. So soll eine fundierte Symptomabfrage ermöglicht werden. Am folgenden Beispiel „Bauchschmerz“ wird dies genauer erläutert [siehe **Abbildung 1**]:

Leidet eine Person unter starken Bauchschmerzen, dann stellt *SymptomCheck* automatisch, unter anderem, genauere Fragen zur Lokalisation und Schmerzbeginn. Hierbei darf vor allem auch die Auswahlmöglichkeit „nicht genau lokalisierbar“ nicht unterschätzt werden. So können beispielsweise Patienten mit einer Peritonitis auftretende Bauchschmerzen meist nicht exakt lokalisieren und würden ansonsten durch eine eingrenzende Lokalisationsangabe möglicherweise zu einer falschen Aussage verleitet werden.

Desweiteren wurde auch auf unterschiedliche Fragetechniken großen Wert gelegt. Je nach Anforderung werden die Symptome mittels Entscheidungsfragen, One-Choice-/ oder Multiple-Choice Fragen abgeprüft.

Abbildung 1: Detaillierte Symptomabfrage zum Beispiel "akuter Bauchschmerz"

<p><b>Wie schwer sind die Schmerzen?</b></p> <p><input type="checkbox"/> schwach</p> <p><input type="checkbox"/> stark</p>
<p><b>Wie lange haben Sie bereits starke Bauchschmerzen?</b></p> <p><input type="checkbox"/> schon seit längerer Zeit</p> <p><input type="checkbox"/> erst seit kurzem, plötzlich aufgetreten</p>
<p><b>Empfinden Sie die starken Bauchschmerzen verstärkt bei körperlicher Belastung oder Druck auf die schmerzende Bauchregion?</b></p> <p><input type="checkbox"/> Ja</p> <p><input type="checkbox"/> Nein</p>
<p><b>Ändert sich die Schmerzintensität der Bauchschmerzen durch die Nahrungsaufnahme oder indem Sie sich anders hinlegen/setzen/stellen?</b></p> <p><input type="checkbox"/> Ja</p> <p><input type="checkbox"/> Nein</p>
<p><b>Wo genau haben Sie die starken Bauchschmerzen?</b></p> <p><input type="checkbox"/> im linken Oberbauch</p> <p><input type="checkbox"/> im linken Unterbauch</p> <p><input type="checkbox"/> im rechten Oberbauch</p> <p><input type="checkbox"/> im rechten Unterbauch</p> <p><input type="checkbox"/> nicht genau lokalisierbar</p>

Zusätzlich wurde bei der Auswahl differenzierterer Fragen streng darauf geachtet, dass alle Fragen von einem Laien zu beantworten sind. Insgesamt beinhaltet *SymptomCheck* 145 solcher zusätzlichen Fragen. Von diesen 145 Fragen sind 116 nachfolgende Fragen auf die allgemeinen Symptome der Startseite und 29 vertiefendere Fragen, die auf einzelne der 116 Fragen folgen. Dadurch erstreckt sich die maximale Symptomabfrage auf 243 mögliche zu beantwortende Fragen. Wichtig ist jedoch, dass der Benutzer frei entscheiden kann, ob er alle ihm gestellten Fragen beantworten möchte oder nicht. *SymptomCheck* kann selbstverständlich durch detailliertere Informationen eine relevantere Verdachtsdiagnoseliste zusammenstellen als durch eine nur oberflächlich abgefragte Beschwerdesymptomatik. Es ist jedoch nicht auf eine vollständige Beantwortung der Fragen angewiesen. Ein Problem, das bei der Ausarbeitung der Fragen deutlich wurde, war die Vereinbarkeit von fundierter Symptomabfrage und alltagstauglicher Beantwortungsdauer. Dieser Aspekt

wurde mittels dieser Studie ebenfalls getestet und wird im Kapitel **Ergebnisse** vorgestellt.

Zudem enthält *SymptomCheck* bei eventuell unklaren Fragen jeweils einen Erklärungsbutton, welcher zusätzliche Erläuterungen liefert. Weiß der Benutzer beispielsweise nicht, was Atemnebengeräusche sind, so erhält er durch das Anklicken eines sich direkt daneben befindlichen Erklärungsbuttons folgende Erläuterung: „Atemnebengeräusche sind ungewöhnliche Atemgeräusche (Brummen, Pfeifen, Rasseln), die neben den normalen Atemgeräuschen (d.h. der Luftströmung) auftreten.“ Dadurch werden dem Benutzer die Fragen noch anschaulicher erklärt und eventuelle Missverständnisse verhindert.

Ursprünglich sollten auf der Startseite, welche im nachfolgenden Kapitel behandelt wird, ebenfalls Erklärungsbuttons erscheinen. Dies ließ sich aufgrund technischer Schwierigkeiten nicht umsetzen. Dennoch enthält *SymptomCheck* insgesamt 43 Erklärungsmöglichkeiten für den Benutzer.

#### **4.4 Entwicklung eines Bewertungssystems**

##### **Bewertungsschema von *SymptomCheck***

Mindestens ebenso relevant wie die Konzeption einer stimmigen Symptomabfrage ist die Bewertung der einzelnen Informationen, welche *SymptomCheck* durch den Benutzer erhält.

Neben einem Spaltenschema wurde in der Wissensbasis ein für *SymptomCheck* kompatibles Bewertungsschema entwickelt [siehe **Tabelle 2**]. Grundsätzlich lässt sich sagen, dass die Symptome mit den Diagnosen durch Zahlenwerte in Beziehung gesetzt werden. Diese Werte werden in Excel zur Übersichtlichkeit anhand von Bewertungseinheiten dargestellt. Hierbei werden die positiven Bewertungseinheiten mit P, die negativen mit N gekennzeichnet. Falls ein auf eine Diagnose zutreffendes Symptom bei einer Person vorliegt, so wird es bezüglich dieser Diagnose mit einer Bewertungseinheit der Kennzeichnung P positiv bewertet. Falls ein Symptom eher gegen eine Diagnose spricht, so erhält es, bezogen auf diese Diagnose, eine negative Bewertungseinheit mit der Kennzeichnung N.

Tabelle 2: Endgültiges Bewertungsschema

Bewertungs- einheit	Bedeutung	Bewertung
P	Diagnose wird wahrscheinlicher.	positive Bewertung
N	Diagnose wird unwahrscheinlicher.	negative Bewertung
<b>Diagnose...</b>		
P1	trifft selten auf das Symptom zu.	2
P2	trifft häufig auf das Symptom zu.	5
P3	trifft sehr häufig auf das Symptom zu.	10
P4	trifft am häufigsten auf das Symptom zu.	20
P7	folgt obligat auf das Symptom.	999
<b>Symptom...</b>		
N7	schließt, wenn verneint, Diagnose aus. (HN-Spalte)	-999
	schließt Diagnose aus. (H-Spalte)	-999
N3	spricht stark gegen die Diagnose.	-10
N2	spricht weniger stark gegen die Diagnose.	-5
N1	spricht wenig gegen die Diagnose.	-2

Die Bewertungseinheit P3 bedeutet, dass eine Diagnose X sehr häufig auf ein Symptom Z zutrifft. Symptom Z kann jedoch bei einer anderen Diagnose Y lediglich häufig (entspricht der Bewertungseinheit P2) oder selten (entspricht der Bewertungseinheit P1) zutreffen. Somit ist dieses Symptom für Diagnose X sensitiver als für Diagnose Y. Desweiteren wurde grundsätzlich festgelegt, dass es für jede Diagnose in der Wissensbasis in Excel zwei mögliche Spalten geben kann:

Hierbei enthält die erste Spalte „H“ sowohl positive als auch negative Bewertungseinheiten. Dadurch kann eine Diagnose als wahrscheinlich (P1-P4), unwahrscheinlich (N1-N3), obligat zutreffend (P7) oder ausschließbar (N7) bewertet werden. Im Gegensatz dazu enthält die zweite Spalte „HN“ einer Diagnose nur spezifische Bewertungseinheit „N7“. Dadurch wird eine Diagnose ausgeschlossen, wenn typisches Symptom verneint wird. Da nicht jede Diagnose durch bestimmte Informationen direkt ausgeschlossen werden kann, weist nicht jede Diagnose die Spalte „HN“ auf.

Dieses Spaltenschema wird im Anschluss an das im Folgenden erläuterte Bewertungsschema erklärt. Da diese Erläuterungen schwer nachvollziehbar

erscheinen können, soll das Spalten- und Bewertungsschema nun anhand eines einfachen Beispiels erklärt werden [siehe **Tabelle 3**].

**Tabelle 3: Bewertungsschema "Bauchschmerzen"**

Typ	Fragen-Antworten	H	HN
Lzeile		akute Magen-Darm-Entzündung	akute Magen-Darm-Entzündung
A	Bauchschmerzen	P2	N7
A	Verdauungsprobleme & Stuhlgangsbesonderheiten	P1	N7
JN	Verstopfung	N7	
OC	Seit wann leiden Sie unter Verstopfung?		
A	erst seit kurzem, plötzlich aufgetreten	N7	
A	schon seit längerer Zeit		
JN	ungewöhnliche Farbe des Stuhlgangs	P1	
OC	Wie lange haben Sie bereits starke Bauchschmerzen?		
A	schon seit längerer Zeit	N7	
A	erst seit kurzem, plötzlich aufgetreten	P4	

A=Antwort JN=Ja-Nein-Frage OC=Open-Choice-Frage

Bauchschmerzen allgemein sind ein häufiges Symptom einer akuten Gastroenteritis<sup>59</sup>. Ausgangspunkt ist hierbei diese Annahme: „Leidet ein Mensch unter Bauchschmerzen, dann hat er häufig eine akute Gastroenteritis.“ Aus diesem Grund erhalten die Bauchschmerzen bezüglich der akuten Gastroenteritis die positive Bewertungseinheit P3. Hat ein Mensch jedoch keine Bauchschmerzen, hat er laut *SymptomCheck* auch höchstwahrscheinlich keine akute Gastroenteritis. Deswegen wird eine akute Magen-Darm-Entzündung beim Verneinen von Bauchschmerzen durch die negative Bewertungseinheit N7 ausgeschlossen. Es ist logisch, dass *SymptomCheck* dadurch Patienten mit atypischer Symptomatik höchstwahrscheinlich nur unzureichend beraten kann. Allerdings übersteigt eine Integration von atypischen Beschwerden in die Wissensbasis derzeit noch die Möglichkeiten der Entwickler von *SymptomCheck*.

Außerdem lässt sich anhand des Beispiels zum Beschwerdebeginn starker Bauchschmerzen erklären, weshalb die „H-Spalte“ positive und negative Bewertungseinheiten enthalten kann. Nachdem eine Person „Bauchschmerzen“ als Symptom ausgewählt hat, fragt *SymptomCheck* automatisch nach der

Intensität. Gibt eine Person „starke Bauchschmerzen“ als Symptom an, so stellt *SymptomCheck* im Anschluss die Frage „Wie lange haben Sie bereits starke Bauchschmerzen?“. Wird die Antwortmöglichkeit „erst seit kurzem, plötzlich aufgetreten“ ausgewählt, so deutet diese Information mit der Bewertungseinheit „P4“ stark auf eine akute Gastroenteritis hin. Wählt man allerdings die Antwortmöglichkeit „schon seit längerer Zeit“ aus, ist dieses Symptom für die Diagnose ein Ausschlusskriterium und wird mit der Bewertungseinheit „N7“ bewertet. Gleichzeitig wird dadurch die Antwort „erst seit kurzem, plötzlich aufgetreten“ als Verneinung gedeutet. Somit ist die Verneinung dieses Symptoms mit der Bewertungseinheit „N7“ spezifisch für eine akute Gastroenteritis. Es wurde versucht, dieses Bewertungsschema weitestgehend umzusetzen. Allerdings war dies nicht immer möglich, da die strikte Einteilung in „P3“ (Diagnose trifft sehr häufig auf das Symptom zu.) „N1“ (Symptom spricht wenig gegen die Diagnose) und die anderen oben genannten Kategorien aus der Fachliteratur meist nicht zu entnehmen waren. Somit musste die Einteilung häufig nach Fachwissen der Entwickler vorgenommen werden, wenn keine genaueren Informationen vorhanden waren. Mittels Fachliteratur wurde versucht einen Überblick über die Leit- und Begleitsymptome jeder Diagnose zu gewinnen<sup>59,62-66,68</sup>. Anschließend wurde versucht dieses Wissen in spezifische Bewertungseinheiten zu transformieren. Als Erläuterung dient das oben aufgezeigte Beispiel des Beschwerdebegins der akuten Gastroenteritis. Da der akute Beginn laut Fachliteratur ein Leitsymptom ist, konnte eine schon seit längerer Zeit andauernde Beschwerdesymptomatik als Ausschlusskriterium hinzugezogen werden. Doch da auch Fachwerke oftmals nicht strikt in Leit-, Begleit- und seltene, aber mögliche Symptome unterscheidet, war bezüglich der Wahl der negativen Bewertungseinheiten das Einbringen eigenem Fachwissens zur Orientierung notwendig. Dies erklärt auch die auf den ersten Blick nicht direkt ersichtliche, jedoch enorme Schwierigkeit dieser Wissensbasisentwicklung. Aus diesem Grund ist es sehr gut möglich, dass andere Personen, die eine solche Wissensbasis entwickeln, in vielen Fällen über andere Bewertungseinheiten entschieden hätten. Wir, die Entwickler von *SymptomCheck*, sind der Meinung, dass man die Wissensbasis durch fundierteres Fachwissen, beispielsweise durch Fachärzte

und nicht, wie in diesem Fall, durch einen Studenten, optimieren kann. Ein zusätzlich wichtiger Aspekt von *SymptomCheck* ist die Möglichkeit, dass „Diagnosen“ vorgeschlagen werden können, welche keine ICD-10 Diagnosen, sondern nicht-pathologische Erklärungen für die Beschwerden sind [siehe **Tabelle 36**]. So weist das Programm die Benutzer gegebenenfalls darauf hin, dass deren Symptome auch beispielsweise eine psychische Ursache haben könnten oder auf eine zu geringe tägliche Trinkmenge, Bewegungsmangel oder eine ungesunde Ernährung zurückzuführen sind. Auch die Bewertungseinheiten wurden extra an diese „Diagnosen“ angepasst. So erhalten diese Verdachtsdiagnosen häufig die Bewertungseinheit „P4“, da sie im Alltag für ausgewählte Symptome die häufigsten Gründe sind. So ist eine Verstopfung beispielsweise meist das Resultat mangelnder Bewegung oder auf eine zu geringe tägliche Trinkmenge zurückzuführen<sup>59</sup>. Am Beispiel „Verstopfung“ wird auch deutlich, dass es dafür mehrere im Alltag oftmals zutreffende Erklärungen gibt. Symptome, wie eine zu geringe tägliche Trinkmenge oder eine unfallbedingte Verletzung, können jedoch durch zusätzliche Fragen, wie „Trinken Sie mindestens 2 Liter am Tag?“ oder – hier im Falle von Brustschmerzen - „Haben Sie sich an der Brust verletzt/gestoßen oder die Brustmuskeln stark beansprucht?“ ausgeschlossen werden.

Anhand dieser zwei Schemata – dem Bewertungs- und dem Spaltenschema - werden die Symptome mit den Diagnosen in die Wissensbasis von *SymptomCheck* umgesetzt. Genauer zu deren Alltagstauglichkeit wird im Kapitel **Ergebnisse** noch erläutert. Die Schemata sind das Resultat aus zahlreichen von uns zuvor durchgeführten Bewertungsversuchen, die auf unterschiedlichen Bewertungseinheiten und /-regeln basierten. Da die Bewertungsversuche jedoch für die endgültige Version von *SymptomCheck* nicht von Bedeutung sind, wird nicht genauer darauf eingegangen. Für jede Diagnose werden die positiven beziehungsweise negativen Zahlenwerte aller eingegebenen Symptominformationen zu einem Gesamtwert aufsummiert. Dieser Gesamtwert wird mit den Gesamtwerten aller anderer in *SymptomCheck* vorhandenen Diagnosen in Relation gesetzt. Je höher der Gesamtwert ist, desto höher erscheinen die Diagnosen im Ranking der Verdachtsdiagnosenliste.

### Ursprüngliches Bewertungsschema der Pilotstudie I

Erwähnenswert ist zudem das ursprüngliche Bewertungsschema von *SymptomCheck*. Dieses wurde für die Wissensbasis der Pilotstudie I verwendet, danach jedoch immens verbessert, umstrukturiert und resultiert nun in dem oben erläuterten endgültigen Schema. Daher ist es sehr interessant und wichtig zu erklären, wodurch die Verbesserungen zustande kamen. Vergleicht man das neue [siehe [Tabelle 2](#)], mit dem alten Bewertungsschema [siehe [Tabelle 4](#)] so fallen die geänderten negativen – d.h. N1-N7 – Bewertungen auf.

**Tabelle 4: Ursprüngliches Bewertungsschema**

Bewertungs- einheit	Bedeutung	Spalte	Wert
P	Diagnose wird wahrscheinlicher		positive Bewertung
N	Diagnose wird unwahrscheinlicher		negative Bewertung
<b>bezüglich eines Symptoms...</b>			
P1	selten zutreffende Diagnose	H	2
P2	möglich zutreffende Diagnose	H	5
P3	sehr häufig zutreffende Diagnose	H	10
P4	häufigste im Alltag zutreffende Diagnose	H	20
<b>Symptom...</b>			
N5	schließt, wenn verneint, Diagnose fast aus.	HN	-40
	schließt Diagnose fast aus.	H	-40
N4	spricht, wenn verneint, sehr stark gegen die Diagnose.	HN	-20
	spricht sehr stark gegen die Diagnose.	H	-20
N3	ist für diese Diagnose ein <b>Leitsymptom</b> .	HN	-10
	spricht stark gegen die Diagnose.	H	-10
N2	ist für diese Diagnose ein <b>Begleitsymptom</b> .	HN	-5
	spricht mäßig gegen die Diagnose.	H	-5
N1	ist für diese Diagnose ein <b>seltenes, aber mögliches Symptom</b> .	HN	-2
	spricht schwach gegen die Diagnose.	H	-2
<i>Anmerkung: Die Erklärungen sind immer auf ein Symptom bezogen.</i>			

Die ursprüngliche Orientierung erfolgte hauptsächlich anhand der Fachliteratur<sup>59,62-66,68</sup>. Hierbei wurden die Symptome eines Krankheitsbildes in Leit-, Begleit- und seltene, aber mögliche Symptome unterteilt. Diese Symptome sind unterschiedlich spezifisch für eine bestimmte Erkrankung. Daher enthielt die „HN-Spalte“ ursprünglich diese spezifischen Informationen in doppelter Negation. Dies wird anhand der Konstellation „Übelkeit, Erbrechen bei Appendizitis“ [siehe **Tabelle 5**] verdeutlicht:

**Tabelle 5: Erklärung der doppelten Negation anhand des Beispiels "Übelkeit, Erbrechen bei Appendizitis"**

Fragen-Antworten	H	HN
	Appendizitis	Appendizitis
Übelkeit, Erbrechen	P1	N2

Übelkeit und Erbrechen sind häufige Symptome einer Appendizitis<sup>59</sup>. Zur Integration dieser Information in die Wissensbasis war die Erstellung eines einheitlichen Schemas erforderlich. Um zu vermeiden, dass diese spezifischen Informationen von dem Programm als sensitive Informationen gewertet werden, durften die Informationen nicht der „H-Spalte“ zugeteilt werden. Denn Übelkeit und Erbrechen alleine sprechen im Alltag selten für eine Appendizitis. Daher wurden diese spezifischen Informationen der „HN-Spalte“ zugeordnet. Folglich ist diese so zu deuten: „Gibt eine Person nicht an, dass sie unter Übelkeit und/oder Erbrechen leidet, dann ist es weniger wahrscheinlich, dass bei ihr eine Appendizitis vorliegt.“ Auf diese Art wurde die Information in doppelter Verneinung in die Wissensbasis integriert.

Nach Auswertung der Pilotstudie I stellte sich jedoch heraus, dass die „HN-Spalten“ [**Tabelle 6**] im alltäglichen Gebrauch von *SymptomCheck* problematisch sind.

Zur Veranschaulichung dieser Problematik dient das Beispiel des Symptoms „Brustschmerzen“, einem Begleitsymptom einer COPD. Allerdings berichtet nicht jede Person, die an einer COPD leidet, über Brustschmerzen. Somit summieren sich in diesem Schema nicht angegebene Symptome schnell auf, wodurch für eine Erkrankung stark sensitive Symptome stark an Bedeutung verlieren. Dadurch erreicht eine Diagnose trotz des Zutreffens mehrerer Leitsymptome eine

nur mittelmäßige bis schlechte Platzierung im Ranking der Verdachtsdiagnosen, da andere Leit-, Begleit- oder mögliche, seltene Symptome zufälligerweise bei einer Person nicht zutreffen.

**Tabelle 6: Erklärung der Problematik des ursprünglichen Bewertungsschemas**

Fragen-Antworten	H	HN	H	HN
	COPD	COPD	akute Bronchitis	akute Bronchitis
Brustschmerzen	P1	N2	P1	N2
Fieber			P2	N2

Es gilt stets zu beachten, dass Krankheiten sich bei jedem Menschen individuell äußern. Daher war auch der Versuch der Umsetzung einer individuellen Berücksichtigung eine wesentliche Schwierigkeit bei der Entwicklung der Wissensbasis. Dazu wurde das Bewertungsschema deutlich verändert, indem alle „HN-Spalten“ zunächst entfernt wurden. In einem zweiten Schritt wurden nur bei solchen Diagnosen „HN-Spalten“ eingefügt, welche durch notwendig zutreffende Symptome mittels der Bewertungseinheit „N7“ bei Verneinung ausgeschlossen werden können. Desweiteren wurde die hochsensitive Bewertungseinheit „P7“ eingeführt, wodurch eine Diagnose bei Bejahung eines Symptoms zutreffen musste. Außerdem wurden die Bewertungseinheiten N4 und N5 völlig aus dem System entfernt, da solch eine genaue Bewertungsabstufung nicht mit unserem medizinischen Fachwissen durchführbar war und zu starken Verzerrungen führen würde.

Die Möglichkeit von *SymptomCheck*, neben Diagnosen nach ICD-10 auch plausible, im Alltag häufig zutreffende Erklärungen auf die Beschwerden zu liefern, ist ein positiv hervorzuhebender Aspekt. In der endgültigen Version von *SymptomCheck* wurden für diese Erklärungen zusätzlich Ausschlussfragen eingeführt. Solche Fragen, wie beispielsweise „Haben Sie sich an der Brust verletzt/gestoßen oder die Brustmuskeln stark beansprucht?“ oder „Nehmen Sie aktuell Medikamente ein?“, schließen bei Verneinung die Erklärungen „Verletzung durch einen Unfall/ sportliche Überanstrengung“ und „Medikamenten-nebenwirkungen“ aus. Wie bereits im Kapitel **4.2 Diagnosen** anhand der tachykarden Herzrhythmusstörungen erläutert, wurden viele Diagnosen zu einer

übergeordneten Diagnose zusammengefasst. Dies ist insofern besonders sinnvoll, da die Verdachtsdiagnoseliste für den Benutzer wesentlich übersichtlicher erscheint. So werden die verschiedenen Schweregrade einer Herzinsuffizienz, welche zuvor noch als einzelne Diagnosen in *SymptomCheck* erschienen, nun nur noch als Links-, Rechts- oder globale Herzinsuffizienz aufgeführt. Diese Zusammenfassung zu einem Oberbegriff betrifft neben den tachykarden Herzrhythmusstörungen vor allem die endokrinologischen Erkrankungen, Lymphome und Leukämien. Einen genauen Überblick über die zusammengefassten Diagnosen findet sich in der Appendix [siehe **Tabelle 34**].

Eine wichtige Änderung ist zudem die Intensitätsgewichtung der Schmerzen, Müdigkeit und Leistungsminderung. Zuvor wurden bei einigen Diagnosen starke Schmerzen mit z. B. „P3“ und schwache Schmerzen mit dem negativen Gegenwert „N3“ bewertet. Dadurch sollten bei der Angabe starker Schmerzen ungravierende Alltagsbeschwerden von schwerwiegenden Erkrankungen unterschieden werden. In der Pilotstudie I zeigte sich diese Methode jedoch nur wenig alltagstauglich. In der neuen Version wird die genauere Angabe von schwachen Schmerzen nicht extra bewertet. Lediglich die Gewichtung starker Schmerzen bleibt unverändert.

#### **4.5 Entwicklung der Startseite**

Ein weiterer wichtiger Faktor, der für die Benutzerfreundlichkeit ausschlaggebend ist, stellt die Entwicklung der Startseite dar.

Ursprünglich sollte die Startseite interaktiv gestaltet werden, wodurch der Benutzer die Möglichkeit haben sollte, anhand der Abbildung eines menschlichen Körpers durch das Anklicken einer bestimmten Körperregion die einzelnen Symptome auszuwählen. Allerdings wurde im Endeffekt vollkommen von dieser Idee abgewichen. Dies hatte verschiedene Gründe: Zum einen ist es nicht möglich, alle Symptome auf diese Art logisch zuzuordnen zu können. Für viele Symptome wie z. B. Appetitlosigkeit, Angst, Schüttelfrost oder geringe Leistungsfähigkeit hätten zusätzliche interaktive Gruppen entworfen werden müssen. Dadurch wäre die Startseite jedoch inhomogen und nicht schlicht geworden. Ein anderer wichtiger Grund ist die mangelnde Übersichtlichkeit einer solchen

Umsetzung. Da ein wesentlicher Vorteil von *SymptomCheck* darin besteht, viele verschiedene Symptome gleichzeitig zu erfassen und miteinander zu verrechnen, ist es für die korrekte Benutzung relevant, dass der Benutzer einen tatsächlichen Überblick über möglichst alle Symptome erhält. Die Befürchtung war jedoch, dass der Benutzer bei diesem Startseitenkonzept nicht alle gebotenen Körperregionen einzeln auswählen und somit wichtige Symptome entweder nur schwer oder gar nicht finden würde. Zudem besteht hierbei die Gefahr, dass der Benutzer wichtige Beschwerden auszuwählen vergisst.

Die Schwierigkeit der Startseitenentwicklung von *SymptomCheck* liegt auf der Hand: Damit die Startseite benutzerfreundlich erscheint, muss sie übersichtlich, schnell, verständlich, leicht bedienbar und optisch ansprechend sein. Die Umsetzung der endgültigen Startseite gestaltete sich äußerst schwierig, wie im Folgenden genauer erläutert wird: Grundsätzlich wurde versucht, alle für die Benutzerfreundlichkeit relevanten Aspekte miteinander zu kombinieren. Um das Kriterium der Übersichtlichkeit zu erfüllen, sollten alle Symptome potentiell über eine einzige Startseite aufrufbar sein.

Desweiteren mussten alle Symptome thematisch so angeordnet werden, dass sie von einem Laien und nicht nur von einem Mediziner leicht auffindbar sind. Auch mussten alle Symptome mit möglichst einem Wort beschrieben werden, damit die Übersichtlichkeit durch eine relativ homogene Optik erhalten bleibt. Nur in Ausnahmefällen sollten für die Beschreibung eines Symptoms mehrere Worte verwendet werden. Nichtsdestoweniger sollten alle auszuwählenden Beschwerden für den Benutzer verständlich sein. Die endgültige Startseite von *SymptomCheck* ist in **Abbildung 2** dargestellt.

Das Konzept wurde bewusst schlicht gewählt: Die Symptome sind in 18 Symptomkomplexe unterteilt. Hierbei findet sich bei manchen eine Zusammenfassung mehrerer Teilbereiche. Ein Beispiel hierfür ist die Gruppe „Brust, Lunge & Herz“. Grund für den Zusammenschluss war die gemeinsame thorakale Lokalität. Es schien logisch, beispielsweise „Atembeschwerden“ zusammen mit einem „Druck-/Engegefühl in der Brust“ in einer Kategorie aufzuführen.

Abbildung 2: Startseite von SymptomCheck

<b>Schmerzen</b> Kopf/Gesicht Mund/Zahn/Zunge/Lippe Arm Schulter Bauch Schulter Hand /Fuß Hals Ohr Brust Bein Auge Knie Rücken		<b>Basisdaten</b> zu hoher/niedriger Blutdruck zu hoher Blutzucker zu hoher / zu niedriger / unregelmäßiger Puls Fieber		<b>Kopf/Gesicht</b> Probleme mit der Gesichtshaut vergrößerter Kopfumfang Sonstige Symptome		<b>Auge</b> schlechteres Sehvermögen gerötete Augen geschwollene Augenlider Hervortreten der Augen aus der Augenhöhle Sonstige Augenbeschwerden		<b>Haare</b> Haarausfall Sonstige Haarsymptome	
<b>Hand &amp; Fuß &amp; Bein &amp; Arm &amp; Schulter &amp; Rücken</b> <b>Hände &amp; Füße:</b> kalte/helle Hände/Füße Kribbeln Blässe an Händen/Füßen Zitternde Hände/Füße oder Muskelzucken Größen-/Formveränderung starke Verhornung brüchlige Fingernägel <b>Arme &amp; Beine:</b> geschwollene Arme, Beine, Fußgelenke Gehbeschwerden, Beinverbiegungen Knochenbrüche Krämpfe		<b>Bauch &amp; Verdauung &amp; Stuhlgang &amp; Urin</b> Verdauungsprobleme Stuhlgangsbesonderheiten Urinbeschwerden Übelkeit, Erbrechen Druckgefühl im Bauch Blähungen breiartiger Bauch starke Fetttalagerung am Bauch		<b>Kreislaufprobleme:</b> Ohnmachtsanfälle, Bewusstlosigkeit, Sturzanfalle Schock Schüttelfrost, Frösteln Schwindel <b>Frauenbeschwerden</b> <b>Männerbeschwerden</b>		<b>Schlafprobleme</b> Atmung setzt im Schlaf aus nächtliches Luftpappen Atembeschwerden im Schlaf Schnarchen Schlafstörungen <b>Ohr</b> Hörbeschwerden Ohrgeräusche		<b>Nase</b> Schnupfen plötzlichliches Nasenbluten, ungewöhnlich langes, starkes Bluten <b>Brust, Lunge &amp; Herz</b> Atembeschwerden Herzklopfen, Herzstolpern Sodbrennen häufiges Aufstoßen asymmetrische Brustkorbewegung Druck-/Engegefühl in der Brust	
<b>Mund</b> Nahrungsrückfluss aus der Speiseröhre in den Mund Zahnfleischbeschwerden Mundgeruch blaue Lippen Sonstige Mundbeschwerden		<b>Hals</b> Husten Halsschwellung / Schluckbeschwerden Heiserkeit dicke Halsgefäße		<b>Haut</b> Schwitzen Hautausschlag, Juckreiz, schuppige Haut Sonstige Hautauffälligkeiten		<b>Psyche &amp; Gehirn:</b> Benommenheit, Schläfrigkeit, komatöser Zustand Konzentrations-/Gedächtnisstörungen Persönlichkeitsveränderung Angst/Beklemmungsgefühl störende Gedanken / angstbeeinflussende Gefühle Halluzinationen depressive Stimmung		<b>Sonstiges</b> Schwäche, Müdigkeit, geringe Leistungsfähigkeit Appetitlosigkeit, Heißhungerattacken, Salzhunger Libidoverlust, Unfruchtbarkeitsverdacht plötzliche Lähmung eines Körperteils übermäßiges Kälte-/Wärmempfinden starke Gewichtszu-/abnahme starken / keinen Durst plötzliche Größenzunahme/-abnahme Sprechstörungen	

Desweiteren ist unter anderem auch „Sodbrennen“ und „Herzklopfen/Herzstolpern“ aufgeführt, da die Patienten diese Symptome auch thorakal wahrnehmen. Stark auf die wesentlichen Aspekte reduziert ist vor allem die Gruppe „Schmerzen“. Hier wird der Benutzer nur durch die Überschrift auf die Art der Symptomatik hingewiesen. Innerhalb der Gruppen findet man nur noch mögliche Lokalisationen. Somit konnten all diese Symptome jeweils um den Wortteil „-schmerzen“ reduziert werden. Aus „Brustschmerzen“ und „Bauchschmerzen“ wurde in der endgültigen Startseite somit nur noch „Brust“ und „Bauch“. Dadurch sparten wir 50% der für diese Begriffe benötigten Fläche. Eine andere wichtige Gruppe heißt „Basisdaten“. Hierunter findet der Benutzer häufig abgefragte Items zu Blutdruck, Puls, Blutzucker und möglicher erhöhter Körpertemperatur.

In vielen Gruppen existieren zusätzliche Symptommöglichkeiten, wie „Sonstige Augenbeschwerden“ oder „Sonstige Hautauffälligkeiten“ [siehe [Abbildung 3](#)]. Hierbei wurde versucht, schwer zusammenfassbare oder seltene Symptome dennoch leicht zugänglich zu machen. Wählt man beispielsweise „Sonstige Augenbeschwerden“ und „Sonstige Hautauffälligkeiten“ aus, so gelangt man zu folgenden Symptomen:

Abbildung 3: Beispiele für zusätzliche Symptommöglichkeiten

**Welche Augenbeschwerden treffen zu?**

- Doppelbilder, Lichtblitze
- Tageslichtempfindlichkeit ?
- unnormal große / kleine Pupillen
- seltenes Blinzeln ?
- sonstige Veränderungen am Auge ?

**Welche Hautprobleme treffen zu?**

- dicke / dünne Haut (z.B. pergamentartig)
- stark pigmentierte Schleim-/ Haut
- bläulich verfärbte Haut und/oder Schleimhaut
- gelblich verfärbte Haut
- knotenartige Verdickungen / geschwollene Lymphknoten
- blaue Flecken ohne Grund
- feuchte, kalte Haut

Auch geschlechtsspezifische Symptome sind in *SymptomCheck* vorhanden und auf der Startseite unter den Oberbegriffen „Frauenbeschwerden“ und

„Männerbeschwerden“ zu finden. Dadurch gelangt der Benutzer zu einer breiten Palette an – für das jeweilige Geschlecht – körperlichen Auffälligkeiten. In **Abbildung 4** sieht man dies am Beispiel von „Frauenbeschwerden“:

Abbildung 4: Frauenbeschwerden

**Welche Frauenbeschwerden haben Sie?**

- Regelblutungsbeschwerden ?
- allgemein starke Unterleibsschmerzen ?
- ausbleibende Milchsekretion ?
- grundlose Milchsekretion ?
- Jucken/Brennen der Vagina oder starker vaginaler Ausfluss, veränderter Geruch der Vagina
- außergewöhnlich große Klitoris
- starke Körperbehaarung ?
- kleine Brust ?
- stark ausgeprägte Muskulatur ?
- männliche Stimme ?

Körperliche Auffälligkeiten, welche sich nicht in eine der Kategorien einordnen ließen, wurden unter der Gruppe „Sonstiges“ zugeordnet. Somit sind letztlich alle 243 Auswahlmöglichkeiten über die 98 Symptome der Startseite zu erreichen.

### Erläuterung der endgültigen Startseite

Die Startseite ist thematisch in Gruppen geordnet, in welchen sich jeweils Symptomlisten befinden. Die verschiedenen Symptomgruppen sind durch diverse Überschriften voneinander zu unterscheiden. Zutreffende Symptome sind dabei vom Patienten auszuwählen. Nach Auswahl der zutreffenden Symptome fragt *SymptomCheck* anschließend die einzelnen Symptome beispielsweise bezüglich Intensität, Dauer oder Lokalisation differenzierter ab. Sollten bestimmte Symptome unklar formuliert sein, bietet das Programm eine Erklärungshilfe, wenn auf den Erklärungsbutton neben dem jeweiligen Symptom gedrückt wird. Dem Patienten bleibt es jedoch selbst überlassen, wie genau er die Fragen beantworten möchte.

Parallel zur Beantwortung wird der Testperson ein Ranking mit möglichen Diagnosen aufgezeigt. Hierbei werden Diagnosen, die mit hoher Wahrscheinlichkeit zutreffen, visuell hervorgehoben. Zudem kann der Benutzer

bei Bedarf durch einen Klick auf den Infobutton neben einer Verdachtsdiagnose auf eine öffentlich zugängliche Webseite weitergeleitet werden, auf der er sich zur jeweiligen Verdachtsdiagnose informieren kann.

## 5 Klinische Studie zu *SymptomCheck*

### 5.1 Studienziele und Fragestellungen/Hypothesen

Im Folgenden werden die Rahmenbedingungen der Studie zu *SymptomCheck* dargestellt. Zur Beschreibung der klinischen Studie wurden zusätzlich einige Absätze aus dem Studienprotokoll entnommen, das bei Bedarf im Anhang eingesehen werden kann. Zusätzlich zur klinischen Studie wurde eine Online-Evaluation durchgeführt. Diese Erhebung wurde über das Rechenzentrum der Universität Würzburg validiert.

#### 5.1.1 Primäres Studienziel

Es sollte getestet werden, ob *SymptomCheck* Benutzerfreundlichkeit (und somit Alltagstauglichkeit) mit ausführlicher Symptomabfrage und fundierter Verdachtsgenerierung und Verdachtsüberprüfung vereinen kann. Die vorliegende Arbeit befasst sich ausschließlich mit der Verdachtsgenerierung.

Folgende Fragestellungen wurden adressiert:

##### Benutzerfreundlichkeit

Sind die Benutzer grundsätzlich in der Lage, *SymptomCheck* zu bedienen? Finden sich die Nutzer gut mit der Suche von für sie zutreffenden Symptomen zurecht und kommen sie mit dem Frage-Stil zurecht? Führt *SymptomCheck* die Benutzer rasch zu den gewünschten Informationen (mögliche Verdachtsdiagnosen)? Erachten die Probanden *SymptomCheck* zur „Selbstdiagnose“ als nützlich?

##### Überprüfung der Sensitivität der Verdachtsgenerierung

Hilft *SymptomCheck* Patienten, mögliche Erkrankungen zu erkennen? Schlägt *SymptomCheck* richtige Diagnosen vor ohne auf der anderen Seite ungefährliche Symptome unnötig zu dramatisieren?

### 5.1.2 Sekundäre Studienziele

Es sollte untersucht werden, anhand welcher Medien sich die Probanden normalerweise über Ihre Symptome oder Erkrankungen informieren (z. B. Internet, Bücher, Apothekenumschau etc.).

Es sollte untersucht werden, ob die Probanden befürchteten krank zu sein und ob *SymptomCheck* diese Befürchtungen lindern konnte.

Es sollte untersucht werden, ob demographische Parameter die obigen Fragestellungen möglicherweise in relevantem Ausmaß beeinflussen (z. B. Alter, Geschlecht, Bildungsgrad).

### 5.1.3 Hypothesen

1. *SymptomCheck* wird als benutzerfreundlich empfunden.
2. Die zutreffende Diagnose wird von *SymptomCheck* „erkannt“, d.h. die korrekte Diagnose wird unter den 10 wahrscheinlichsten Diagnosen aufgelistet.
3. Jüngere Personen und Personen mit einem höheren Bildungsgrad kommen mit *SymptomCheck* besser zurecht.

## 5.2 Studiendesign

Querschnittstudie ohne Kontrollgruppe. Keine Intervention.

## 5.3 Studienpopulation

*SymptomCheck* wurde an folgenden Probanden getestet:

- a) Für eine Pilotstudie I wurden Probanden (N=16) in der Ambulanz des Deutschen Zentrums für Herzinsuffizienz in Würzburg rekrutiert. Diese Gruppe setzte sich aus Patienten zusammen, die die ambulante Sprechstunde besuchten und somit entweder an konkreten Symptomen litten oder die Sprechstunde aus anderen Gründen (z. B. Verlaufskontrollen) besuchten. Dabei handelte es sich primär um kardiologische Diagnosen. Viele litten

jedoch auch unter Begleiterkrankungen (z. B. Diabetes mellitus, Hypertonie etc.). *SymptomCheck* wurde mittels Computer angewendet.

- b) Im Anschluss daran erfolgte eine Online-Befragung zur Evaluation der Benutzerfreundlichkeit von *SymptomCheck*, indem das Programm online frei zugänglich hochgeladen wurde. Hierbei beteiligten sich 76 Teilnehmer, wobei nicht auszuschließen ist, dass sich Teilnehmer mehrmals an der Umfrage beteiligt haben könnten. Da allerdings keine identisch ausgefüllten Evaluationsbögen eingingen, entspricht jeder Evaluationsbogen in der Auswertung einem Teilnehmer.
- c) Aus der Pilotversion von *SymptomCheck* wurde nach einigen Überarbeitungen [siehe **4.4 Entwicklung eines Bewertungssystems**] die endgültige Version entwickelt, die in der Medizinischen Klinik und Poliklinik in Würzburg an 33 Patienten getestet wurde. Diese Patienten befanden sich in stationärer Behandlung und wurden von *SymptomCheck* zu ihren Beschwerden befragt, wegen derer sie in die Klinik aufgenommen wurden.

## 5.4 Ein- und Ausschlusskriterien der Studienpopulation

- |  |   |
|--|---|
| <p>a) <u>Einschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Volljährigkeit (18 Jahre)</li> <li>• Schriftliche Einverständniserklärung</li> </ul> | <p><u>Ausschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Keine</li> </ul> |
| <p>b) <u>Einschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Jeder mit Zugang zum Internet</li> </ul>   | <p><u>Ausschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Keine</li> </ul> |
| <p>c) <u>Einschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Volljährigkeit (18 Jahre)</li> <li>• Schriftliche Einverständniserklärung</li> </ul> | <p><u>Ausschlusskriterien</u></p> <ul style="list-style-type: none"> <li>• Keine</li> </ul> |

## 5.5 Fragebogen zur Datenerhebung

Mit einem Fragebogen werden unter anderem Informationen zu Alter, Geschlecht, Bildungsgrad, Nutzungsverhalten diverser Quellen für medizinische Informationen und Selbsteinschätzung des Gesundheitszustands erhoben. Die Datenerhebung zur Benutzungsqualität erfolgt mittels eines standardisierten Fragebogens [siehe [5.6 UEQ-Fragebogen zur Benutzerfreundlichkeit](#)] und speziell auf *SymptomCheck* bezogene, einzustufende Aussagen.

Die gesamten Fragebögen sind im Anhang einsehbar.

## 5.6 UEQ-Fragebogen zur Benutzerfreundlichkeit

Zur Evaluation der Benutzerfreundlichkeit verwendeten wir als Teil unseres Fragebogens den standardisierten Fragebogen „User Experience Questionnaire“<sup>69</sup>. Dieser wurde speziell zur Messung der Benutzerfreundlichkeit von Softwareprodukten entwickelt und enthält insgesamt 26 bipolare Items, die über eine Faktorenanalyse selektiert wurden. Diese Items wiederum werden sechs Faktoren zugeordnet [siehe [Abbildung 5](#)]. Sie umfassen Vorhersagbarkeit, Stimulation, Attraktivität, Originalität, Durchschaubarkeit und Effizienz und dienen als Maßstab für die Benutzerfreundlichkeit.

Hierbei liefert die Skala „Attraktivität“ einen allgemeinen Gesamteindruck, wohingegen 3 Skalen die pragmatische und 2 Skalen die hedonistische Qualität berücksichtigen<sup>70</sup>. Somit liefert der UEQ keinen Gesamtwert, sondern anhand der Mittelwerte der sechs Skalen einen differenzierteren Gesamteindruck.

Die einzelnen Items und deren Polung erscheinen dabei im Fragebogen völlig randomisiert und werden „in Form eines siebenstufigen semantischen Differentials“<sup>69</sup> abgefragt. Dabei besteht die Skala „Attraktivität“ aus 6 Items, die restlichen aus nur 4 Items. Für die Auswertung werden alle Werte eines Items von einer ursprünglichen siebenstufigen Skala von +1 bis +7 in eine neue weiterhin siebenstufige Skala von -3 bis +3 transformiert. (Dabei entspricht der Wert +1 dem Wert -3, der Wert +2 dem Wert -2 etc.) Die Deutung der Werte ist [Tabelle 7](#) zu entnehmen.

Abbildung 5: 6 Skalen des UEQ-Fragebogens; übersetzt und verändert aus dem Englischen<sup>70</sup>

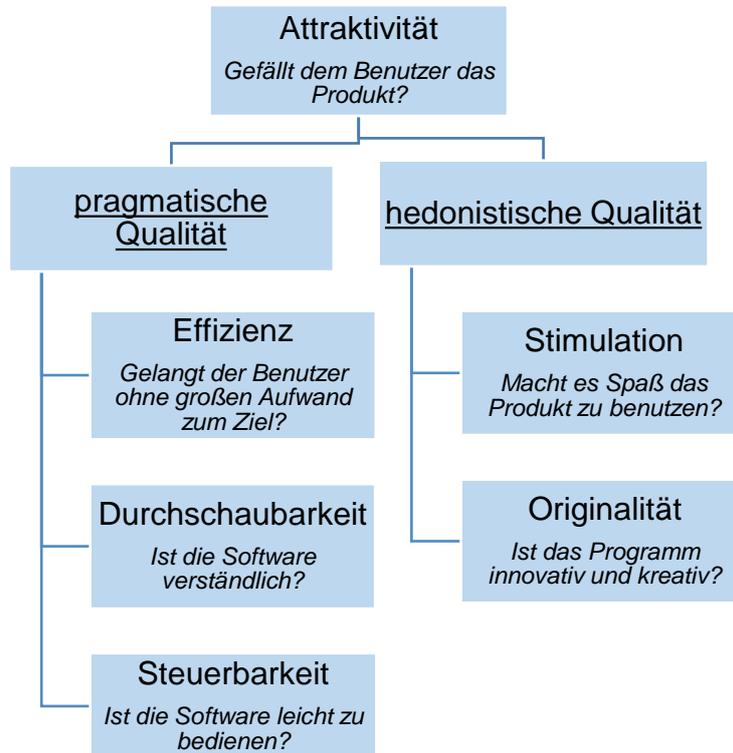


Tabelle 7: Wertebereiche des UEQ-Fragebogens

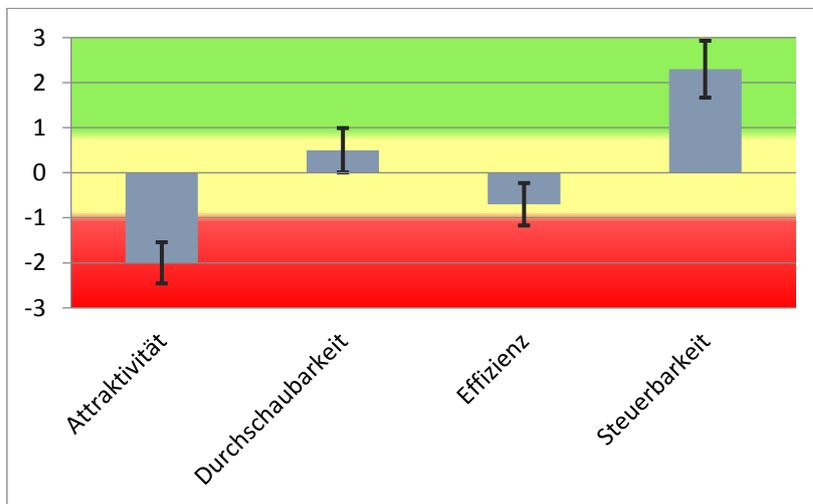
Wertebereich	Bedeutung
> 0,8 bis 3	positive Beurteilung
≥ -0,8 bis ≤ 0,8	neutrale Beurteilung
< -0,8 bis -3	negative Beurteilung

Hierbei wurde die Einteilung der Werte als positiv, neutral oder negativ von den Entwicklern des Fragebogens festgelegt<sup>70</sup>. Demnach sind Bewertungen von  $\geq -0,8$  bis  $\leq 0,8$  als neutral zu werten, wohingegen Bewertungen von  $> 0,8$  bis 3 als positiv und von  $< -0,8$  bis -3 als negativ zu deuten sind.

Im Anschluss wird aus allen gesammelten Werten eines Items der Mittelwert gebildet. Ebenso wird für die anderen Items verfahren. Die Mittelwerte der Items, welche derselben Skala zugeordnet werden, bilden nun wiederum den gemeinsamen Mittelwert dieser Skala. Somit resultieren daraus sechs Mittelwerte, anhand derer die Benutzerfreundlichkeit gemessen wird. Die auf **Tabelle 7** basierende Deutung der Abbildungen und Tabellen wird kurz anhand

der fiktiven Beispiele in **Abbildung 6** und **Tabelle 8** erläutert. Hierbei beträgt der Mittelwert der Skala Attraktivität -2,0 und wird somit in **Abbildung 6** im roten, negativen Bereich abgebildet. In **Tabelle 8** ist das Symbol „↓“ negativ zu werten. Analog dazu werden die Skalen Durchschaubarkeit und Effizient neutral mit 0,5 und -0,7 bewertet und erscheinen folglich in **Abbildung 6** im neutralen gelben Bereich. In **Tabelle 8** wird die neutrale Bewertung mit dem Symbol „↔“ dargestellt. Die Steuerbarkeit erreicht den Mittelwert 2,3 und erscheint in **Abbildung 6** im positiven, grünen Bereich und erhält in **Tabelle 8** analog dazu das Symbol „↑“.

**Abbildung 6: Fiktives Beispiel zur Bewertung der 6 Skalen (UEQ-Fragebogen)**



**Tabelle 8: Fiktives Beispiel zur Bewertung der 6 Skalen des UEQ-Fragebogens**

Skalenbewertung			
Attraktivität	↓ -2,0	Effizienz	↔ -0,7
Durchschaubarkeit	↔ 0,5	Steuerbarkeit	↑ 2,3
<b>Bewertungsschema</b>	> 0,8 bis 3 positive Bewertung ≥ -0,8 bis ≤ 0,8 neutrale Bewertung < -0,8 bis -3 negative Bewertung		

## 5.7 Studienablauf

*SymptomCheck* wurde an insgesamt drei Kollektiven getestet. Diese sind in Kapitel **5.3 Studienpopulation** und **5.4 Ein- und Ausschlusskriterien der Studienpopulation** genauer erläutert. Zuerst fand die Pilotstudie I statt, in der *SymptomCheck* initial getestet wurde. Nach einigen Modifikationen kam die verbesserte Version von *SymptomCheck* in der Online-Evaluation und in der Pilotstudie II zum Einsatz. In den Pilotstudien I und II wurden die Benutzerfreundlichkeit und besonders die Überprüfung der Verdachtsgenerierung anhand klinischer Patienten eruiert. Im Rahmen einer Online-Evaluation wurde die Benutzerfreundlichkeit schwerpunktmäßig durch ein Online-Kollektiv evaluiert. Die einzelnen Kollektive werden im Folgenden genau erläutert.

### Pilotstudie I am DZHI

Im Rahmen einer Pilotstudie I wurden in der Ambulanz des Herzinsuffizienzentrums in Würzburg (DZHI) 16 Patienten während ihrer Wartezeit befragt. Die Befragung wurde innerhalb vom 26.05.2015 bis zum 05.06.2015 durchgeführt, wobei wir pro Patient zwischen 15 und 30 Minuten benötigten. Die Patienten wurden dabei willkürlich ausgewählt. Vor Beginn der Testung wurden die Patienten anhand einer Patientenaufklärung über die Studie informiert und unterzeichneten eine Einwilligungserklärung zur Teilnahme an der Datenerhebung. Im Anschluss füllten sie handschriftlich einen Evaluationsbogen aus. Daraufhin wurde den Probanden auf einem Computer die Startseite von *SymptomCheck* aufgezeigt und die Probanden aufgefordert, deren aktuelle Beschwerden auszuwählen. Die Testpersonen sollten dabei das Programm eigenständig bedienen. Nach Beenden von *SymptomCheck* wurden die Probanden abschließend gebeten einen zweiten Fragebogen auszufüllen. Nach Abschluss der Patientenbefragung wurden die Ergebnisse des Diagnoserankings mit den entnommenen Diagnosen der Patientenakten abgeglichen.

### Online-Evaluation

Nach Auswertung der Datenerhebung im DZHI wurde die Online-Evaluation vom 07.09.2015 bis zum 06.10.2015 durchgeführt. Der Link, über welchen man zu

einer initialen Informationsseite der Online-Evaluation gelangte, wurde über das soziale Netzwerk Facebook, der Webseite des DZHI, sowie über E-Mails verbreitet. Die Informationsseite von *SymptomCheck* erläutert unter anderem den Sinn der Online-Evaluation von *SymptomCheck*, die Anonymität der Datenerhebung und weist den Probanden darauf hin, nach Testung von *SymptomCheck* die Evaluation sorgfältig durchzuführen. Desweiteren enthält die Probandeninformationsseite auch eine Verlinkung zu einem Impressum, welches Informationen zum Haftungsausschluss, Datenschutz, Urheberrecht und Kontaktadressen liefert [siehe **Abbildung 30**].

Über einen Startbutton gelangte der Proband auf die Startseite von *SymptomCheck* [siehe **Abbildung 2**], welche dem Benutzer über einen kurzen Informationstext nochmals die Bedienung des Programms erläutert. Nach Abschluss der Testung konnten die Probanden über einen Feedback-Button zur Evaluation gelangen. Der Fragebogen dazu ist die leicht gekürzte Version der klinischen Fragebögen [siehe **Anhang**], die im DZHI und der Medizinischen Klinik und Poliklinik I des Universitätsklinikums Würzburg zum Einsatz kamen. Nach Abgabe der Evaluation [siehe **Abbildung 31**] wurden die Daten auf dem Server der Fakultät für Informatik in Würzburg unter der Leitung von Herrn Prof. Frank Puppe (Lehrstuhl Informatik VI) gespeichert.

### **Pilotstudie II in der Medizinischen Klinik und Poliklinik I**

Die Pilotstudie II erfolgte vom 09.09.2015 bis zum 22.09.2015 in der Medizinischen Klinik und Poliklinik I in Würzburg auf den Stationen M31, M33 und M34. Hierbei wurden insgesamt 33 Patienten befragt. Diese befanden sich in stationärer Behandlung und wurden einzeln in ihren Zimmern befragt. Der Ablauf erfolgte analog zur Befragung im DZHI. Hierbei wurden die Patienten gebeten, die Symptome in *SymptomCheck* auszuwählen, welche sie bei Aufnahme in die Klinik hatten. Im Anschluss an die Befragung verglichen wir deren Hauptdiagnose, die dem SAP-System entnommen wurden, mit der Liste an Verdachtsdiagnosen von *SymptomCheck*. Es wurde geprüft, ob sich die tatsächliche Diagnose unter den 10 der am wahrscheinlichsten zutreffenden Verdachtsdiagnosen befindet.

## 5.8 Datenschutz und Ethikvotum

Vor Beginn der Testung mussten die Probanden nach angemessener Vorabinformation eine Einwilligungserklärung zur Teilnahme an der Studie unterschreiben, die sie ebenfalls als Kopie erhielten. Dadurch bestätigten die Probanden, dass ihre Daten zu Studienzwecken benutzt werden durften. Im Anschluss erhielt die Testperson den Evaluationsbogen, auf dem eine Kennnummer (Stud-PID) vermerkt ist.

Auf einer Liste in Papierform wurde jede Stud-PID einem Patienten zugeordnet und so pseudonymisiert. Diese Liste verblieb im DZHI und wurde nach Ablauf von 3 Monaten vernichtet, so dass die Informationen dann ausschließlich anonymisiert prozessiert wurden. Die zugeteilte Stud-PID wurde bei Bearbeitung des Programms in den Computer eingegeben. Die auf dem Computer erhobenen Informationen wurden anonym auf dem Server des Rechenzentrums der Universität Würzburg gespeichert. Diese Daten wurden nur für die Auswertung dieser Studie gespeichert und benutzt. Das Rechenzentrum hatte keinen Zugang zur Patientenidentifikationsliste, die im DZHI verblieb.

Im Anschluss daran wurde das Diagnoseergebnis des Programms mit den Diagnosen der Patientenakte verglichen. Dabei wurden außer Angaben zu Geschlecht, Alter, Bildungsstand und Diagnosen keine weiteren Informationen verwendet. Weitere Informationen, insbesondere die Namen der Probanden blieben somit pseudonymisiert. Nach Ablauf von drei Monaten wurden die Daten anonymisiert.

Es wurde ein Studienprotokoll (5/15) der Ethikkommission vorgelegt, das nach kleinen Modifikationen (Amendment 01) als ethisch unbedenklich eingestuft werden konnte.

## 5.9 Fallzahlen

Für die Pilotstudie I im DZHI wurden 16 Patienten befragt. In der Medizinischen Klinik und Poliklinik in Würzburg wurden Daten von insgesamt 33 stationär aufgenommenen Patienten erhoben. Eine formale Fallzahlschätzung erfolgte nicht, da zunächst die Benutzerfreundlichkeit von *SymptomCheck* qualitativ erfasst werden sollte. Die in einem späteren Abschnitt vorgestellten Ergebnisse

der Pilotstudie I sollten als Basis für die Folgestudie in der Medizinischen Klinik und Poliklinik in Würzburg dienen, die die Benutzerfreundlichkeit, Diagnosesicherheit und die klinische Wertigkeit des Einsatzes eines solchen Systems quantifiziert. Die Daten wurden rein deskriptiv ausgewertet. Statistische Kenngrößen zur Diagnosesicherheit wurden explorativ erhoben. Desweiteren wurde auf mögliche Unterschiede verschiedener demographischer Variablen geachtet. Dies erfolgte allerdings nicht mittels Regressionsmodellen, sondern anhand direkter Vergleiche.

### **5.10 Nutzen-Risiko-Abwägung**

Die an der Studie teilnehmenden Probanden hatten selbst keinen Nutzen durch die Studie. Allerdings trugen sie durch ihre Teilnahme zur Evaluierung und Verbesserung des Programms *SymptomCheck* bei. Da die Patienten lediglich einen Evaluationsbogen ausfüllten und das Programm durcharbeiten sollten, gab es für sie kein gesundheitliches Risiko. Die Patienteninformationen wurden pseudonymisiert, so dass auch die Gefahr, dass diese an unberechtigte Personen gelangen, minimiert wurde.

### **5.11 Datenanalyse**

Die statistische Auswertung erfolgte mit der Software SPSS 20.

Hierbei lag der Schwerpunkt auf deskriptiven Analysen der gesammelten Daten. Die Auswertung wurden zum einen mittels Kreuztabellen durchgeführt, die die relativen und absoluten Häufigkeiten aufzeigten. Außerdem wurden auch metrische Daten bezüglich Standardabweichung, Mittel-, Minimum-, Maximum- und Medianwert analysiert.

Gruppenunterschiede wurden mit Hilfe des exakten Tests nach Fisher getestet. Alle Testungen wurden zweiseitig durchgeführt. Für den exakten Test nach Fisher wurde das Signifikanzniveau  $\alpha$  auf 0,05 festgelegt und p-Werte berechnet. Bei einem p-Wert  $\leq 0,05$  galt der Zusammenhang als signifikant. Alle Analysen sind als explorativ anzusehen. Es wurde deshalb keine Anpassung auf multiples Testen vorgenommen.

## 6 Vergleich von *SymptomCheck* mit einem Arzt

Zusätzlich zur klinischen Studie wurde die Verdachtsdiagnosefindung von *SymptomCheck* mit der eines Arztes verglichen. Hierfür diente ein medizinisches Fallbuch<sup>71</sup> aus dem Bereich der Inneren Medizin. Darin werden in jeweils einem Beispiel die Symptome eines fiktiven Patienten geschildert, woraufhin sich die Leser mögliche Verdachtsdiagnosen, notwendige Diagnostik und die passende Therapie überlegen sollen. Diesem Buch wurden die ersten 26 Fallbeispiele entnommen. Die Informationen zu den Symptomen des darin dargestellten Patienten eines jeden Fallbeispiels wurden zum einen in *SymptomCheck* eingegeben zum anderen in identischer Form einem Arzt vorgelegt. Während *SymptomCheck* bei der Auswahl von zutreffenden Symptomen automatisch mögliche Verdachtsdiagnosen auflistet, sollte der Arzt in absteigender Reihenfolge die drei wahrscheinlichsten Verdachtsdiagnosen auswählen. Im Anschluss wurde untersucht, wie häufig die korrekten Diagnosen unter den ersten drei Verdachtsdiagnosen gelistet wurden. Desweiteren wurde zwischen *SymptomCheck* und dem Arzt verglichen, wie häufig die korrekten Diagnosen der Fallbeispiele für die wahrscheinlichste Diagnose („Platz 1“) gehalten wurden.

## Ergebnisse

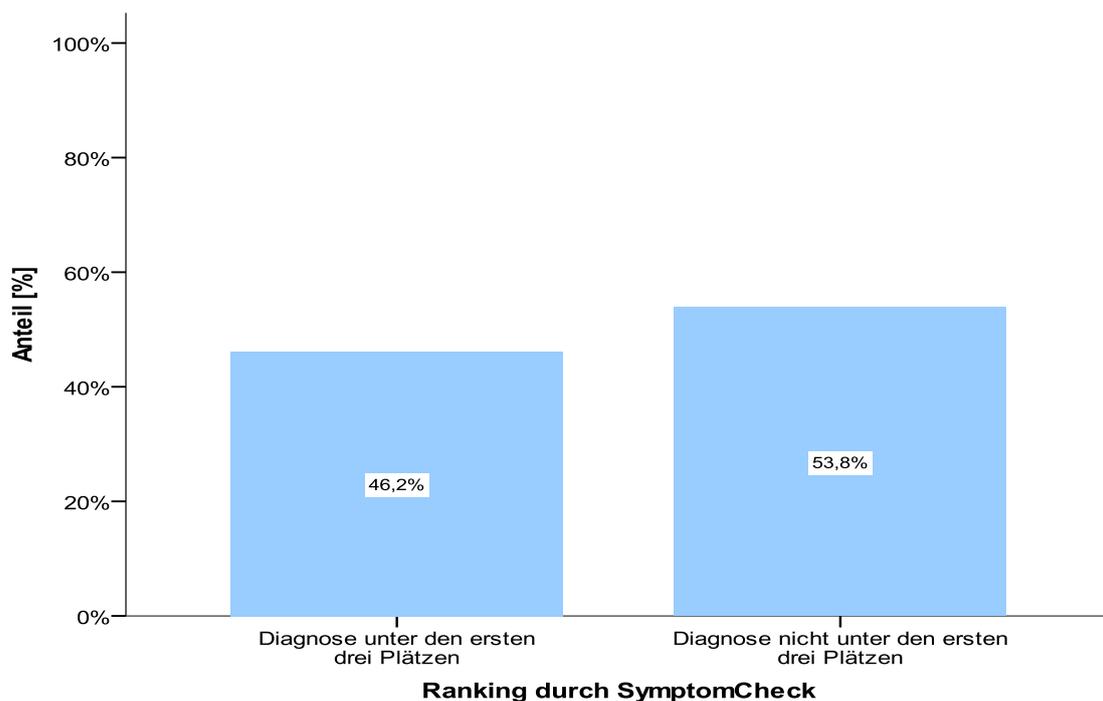
### 7 *SymptomCheck* vs. Arzt

Das Programm *SymptomCheck* ist frei zugänglich unter dem Link: <https://go.uniwue.de/symptomcheck>.

Wie bereits erläutert, wurde *SymptomCheck* anhand von 26 Fallbeispielen mit der Genauigkeit der Diagnosefindung eines erfahrenen Arztes verglichen.

Hierbei wurde zum einen untersucht, wie häufig die korrekte Diagnose unter den ersten drei Verdachtsdiagnosen gelistet wurde. Bei *SymptomCheck* erschien die korrekte Diagnose in 12 Fällen (46,2%) unter den wahrscheinlichsten drei Diagnosen (siehe **Abbildung 7**).

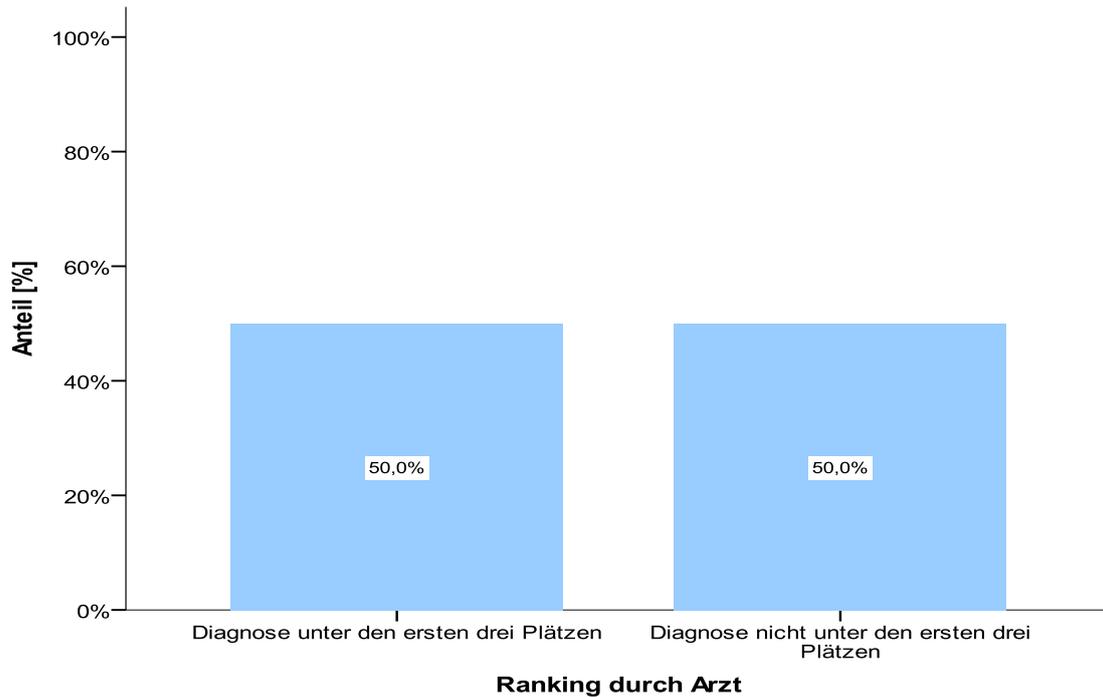
**Abbildung 7: Korrekte Diagnose unter den ersten 3 Plätzen (*SymptomCheck*; n=26)**



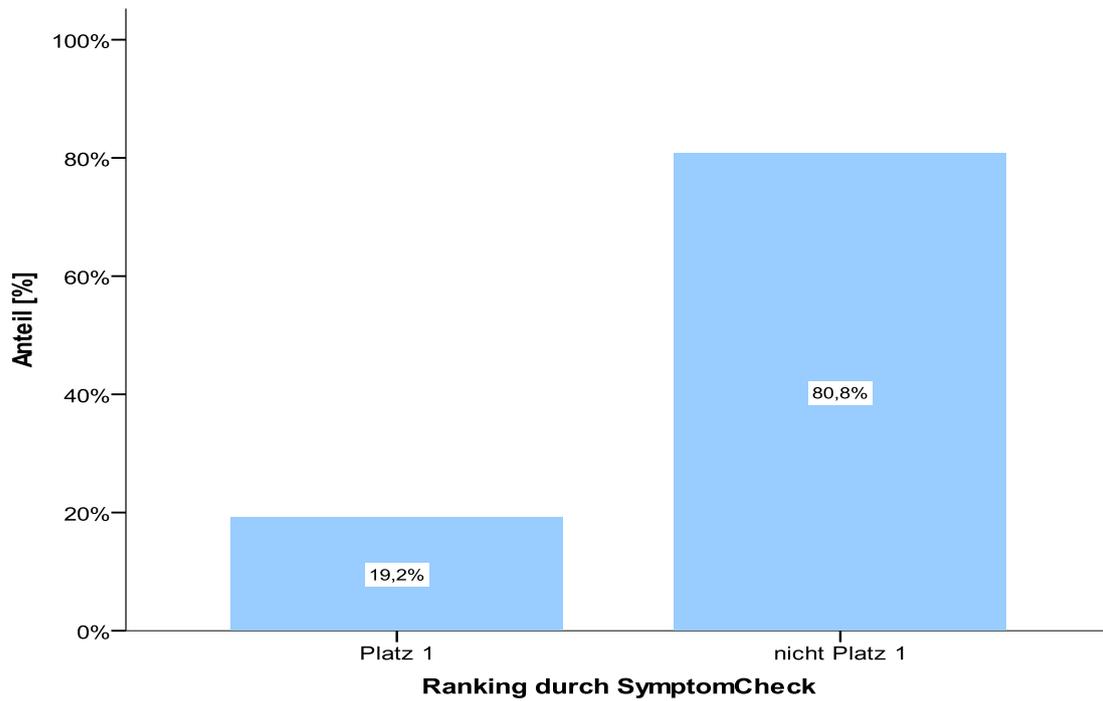
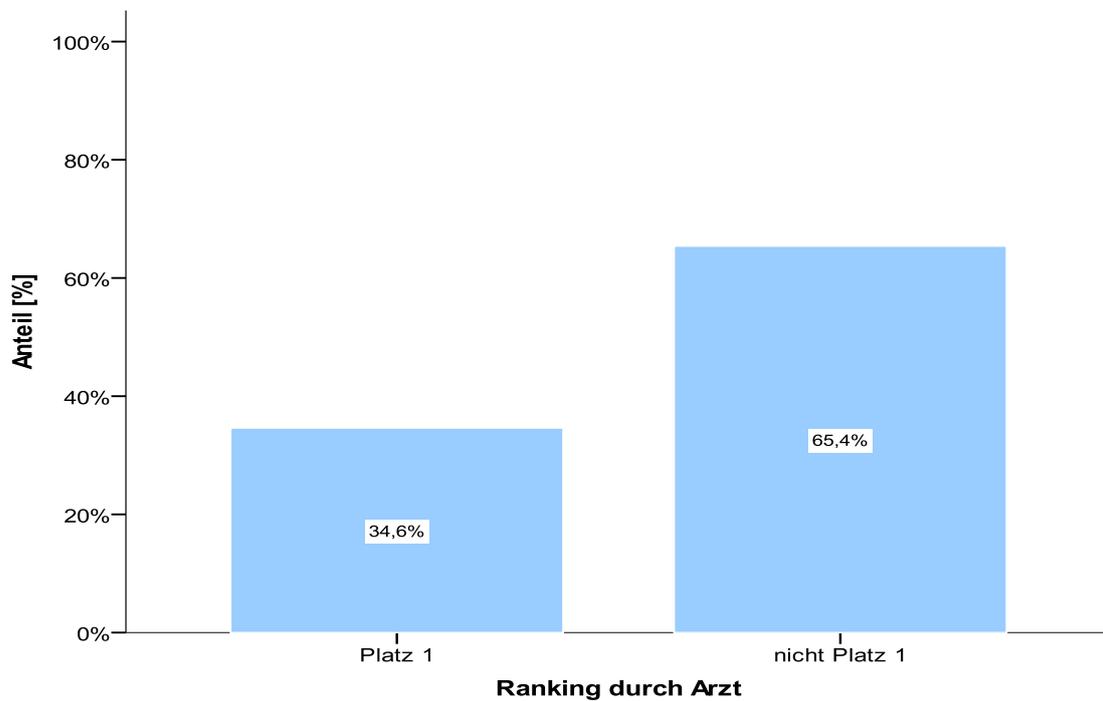
In 14 Fällen (53,8%) erschien die Diagnose in *SymptomCheck* zwar unter den Verdachtsdiagnosen, hierbei konnte sie jedoch auch an beispielsweise 50. Stelle gelistet werden, da in *SymptomCheck* alle enthaltenen Diagnosen prinzipiell aufgelistet werden können. Im Gegensatz dazu wurde in 50,0% aller 26 Fällen die zutreffende Diagnose durch den Arzt auf einem der ersten drei Plätze aufgeführt (siehe **Abbildung 8**). Da der Mediziner jedoch für jedes Fallbeispiel

nur drei Verdachtsdiagnosen angeben konnte, entsprechen die 13 Fälle der Gesamtmenge an durch den Arzt erkannten Diagnosen. Somit lässt sich sagen, dass der Mediziner die korrekte Diagnose mit einer gering höheren Wahrscheinlichkeit unter den besten drei Rängen platzieren konnte, als *SymptomCheck*.

**Abbildung 8: Korrekte Diagnose unter den ersten 3 Plätzen (Arzt; n=26)**



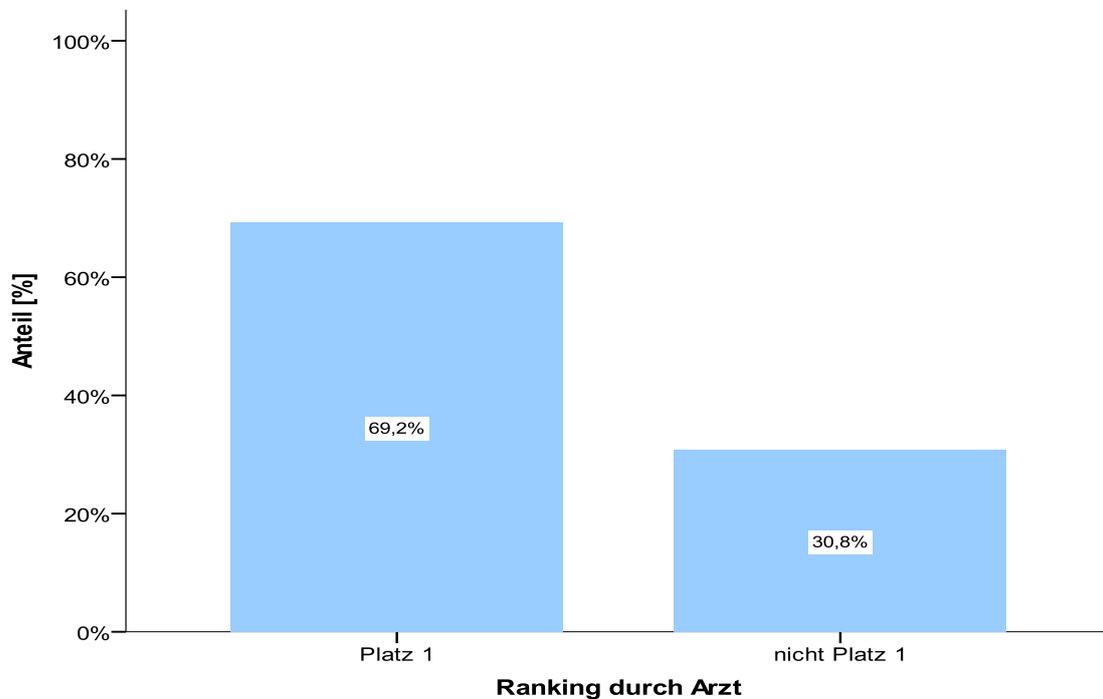
Zum anderen wurde getestet, wie häufig die zutreffende Diagnose sowohl von *SymptomCheck* als auch vom Arzt für die wahrscheinlichste Verdachtsdiagnose gehalten wurde. Von allen 26 Fallbeispielen listete *SymptomCheck* fünf Diagnosen (19,2%) auf Platz 1 (siehe **Abbildung 9**). Dem Arzt hingegen gelang es, neun (34,6%) von 26 Diagnosen der Fallbeispiele korrekt als wahrscheinlichste Diagnose auf Platz 1 zu setzen (siehe **Abbildung 10**).

**Abbildung 9: Korrekte Diagnose auf erstem Platz (SymptomCheck; n=26)****Abbildung 10: Korrekte Diagnose auf erstem Platz (Arzt; n=26)**

Betrachtet man jedoch nur die Diagnosen, welche überhaupt vom Arzt erkannt wurden (n=13), dann konnte der Mediziner sogar in neun (69,2%) von 13 Fällen

die richtige Diagnose auf den ersten Rang platzieren (siehe **Abbildung 11**). Somit schnitt auch hier der Mediziner besser ab, als das Programm.

**Abbildung 11: Korrekte Diagnose auf erstem Platz (Arzt; n=13)**



## 8 Pilotstudie I

### 8.1 Rahmendaten

An der Pilotstudie I des DZHI nahmen 16 Patienten teil. Allerdings brach ein Patient die Befragung nach Beantwortung des ersten Fragebogens und während der Testung von *SymptomCheck* ab. Somit konnten dessen Daten nicht in die Auswertung der Benutzerfreundlichkeit und der Verdachtsgenerierung einbezogen werden.

Insgesamt gab es 10 männliche (62,5%) und 6 weibliche (37,5%) Testpersonen [siehe **Abbildung 12; Tabelle 37**]. Davon werden 62,6% der Altersgruppe der 51- bis 70-Jährigen zugeordnet [siehe **Tabelle 38; Abbildung 13**].

Der Mittelwert innerhalb der Pilotierung liegt bei 58 Jahren, der Median bei 59 Jahren. Die Verteilung auf die verschiedenen Schulabschlüsse ist relativ ausgewogen [siehe **Tabelle 39; Abbildung 14**].

Abbildung 12: Geschlechterverteilung (Pilotstudie I; n=16)

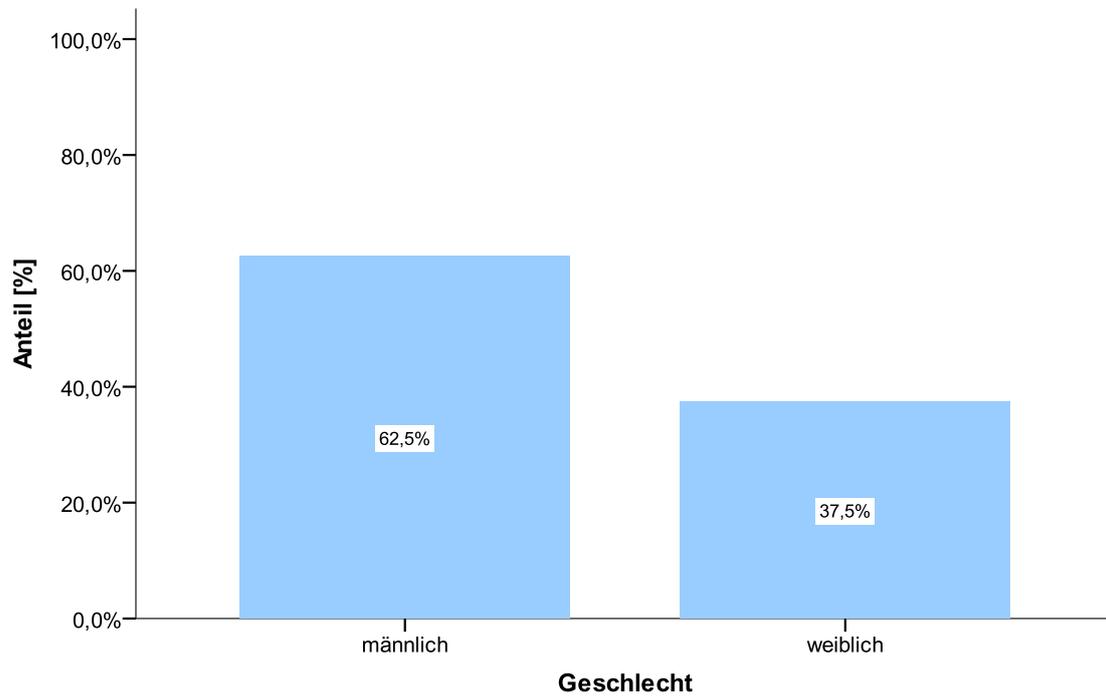


Abbildung 13: Altersverteilung (Pilotstudie I; n=16)

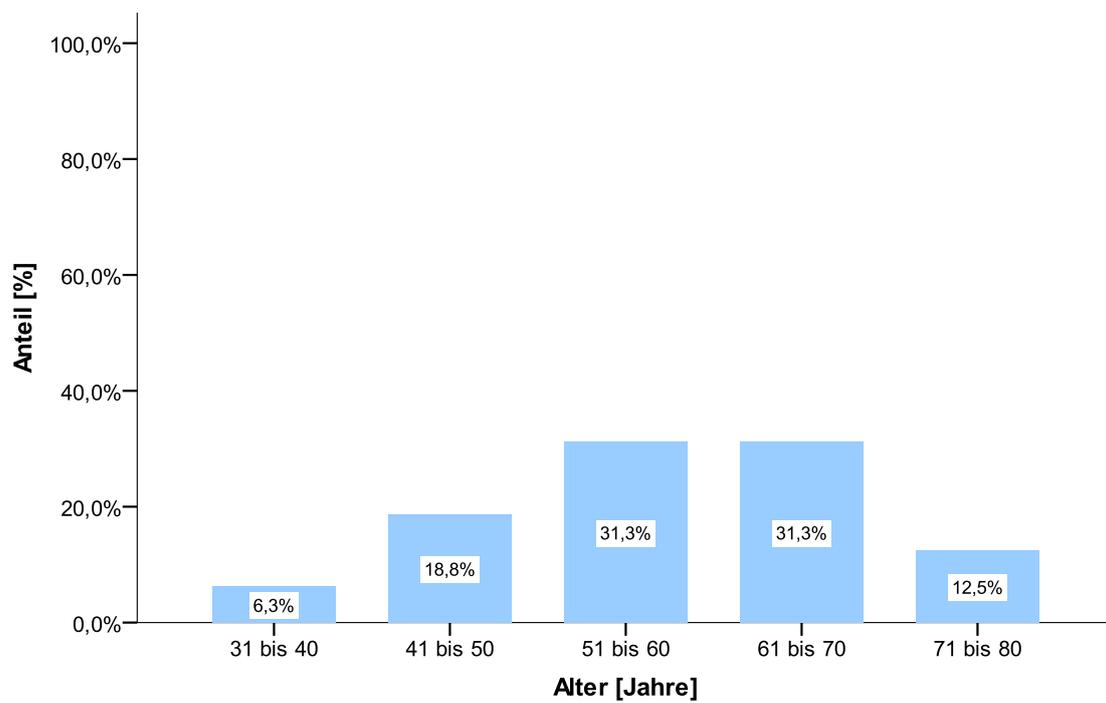
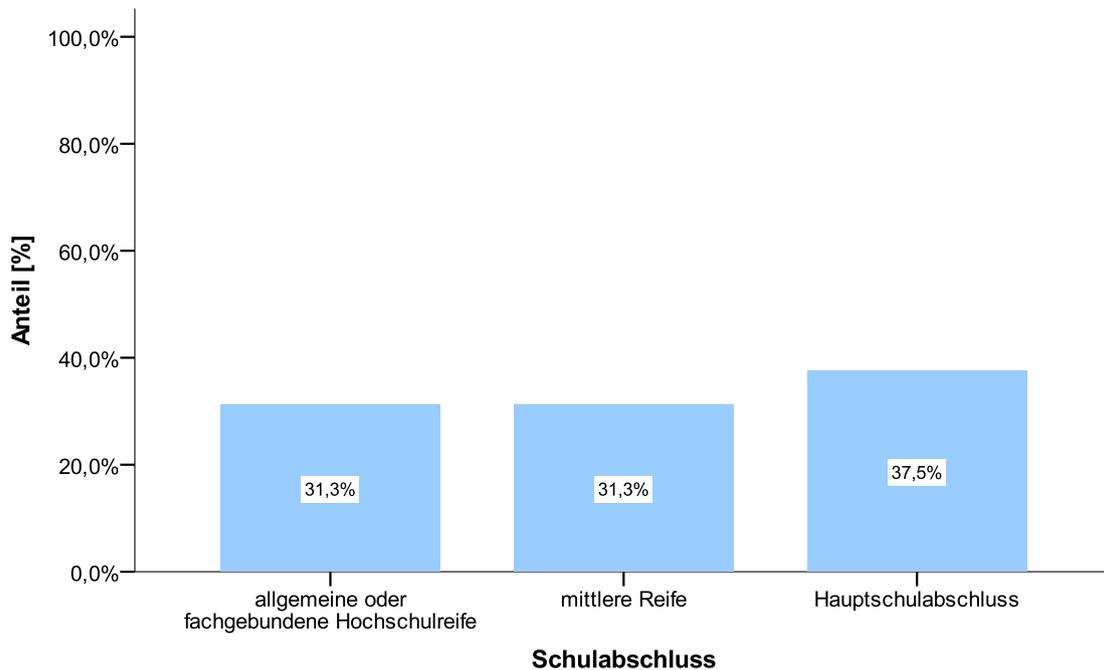


Abbildung 14: Aufteilung der Patienten nach Schulabschlüssen (Pilotstudie I; n=16)



Eine Darstellung der geschlechtsspezifischen Verteilung bezüglich der Altersklassen oder des Schlussabschlusses ist hierbei aufgrund der geringen Fallzahl nicht sinnvoll.

## 8.2 Benutzungsqualität

Im Folgenden werden die Ergebnisse der Benutzungsqualität der Pilotierung vorgestellt. Hierbei wird auf eine detaillierte Auswertung der Benutzerfreundlichkeit verzichtet und nur auf die Ergebnisse des UEQ-Fragebogens eingegangen.

Anhand des „User Experience Questionnaire“<sup>69</sup> wurden bipolare Items bewertet, welche sich wiederum in 6 Skalen einteilen lassen. Die genaue Bewertung der einzelnen Items ist der Appendix [siehe **Tabelle 40**] zu entnehmen.

Die Interpretation der Tabellen und Abbildungen ist in **5.6 UEQ-Fragebogen zur Benutzerfreundlichkeit** erläutert. *SymptomCheck* erzielt in der Pilotstudie I bezüglich aller 6 Skalen positive Ergebnisse [siehe **Tabelle 9**; **Abbildung 15**].

Abbildung 15: Bewertung der 6 Skalen (Pilotstudie I)

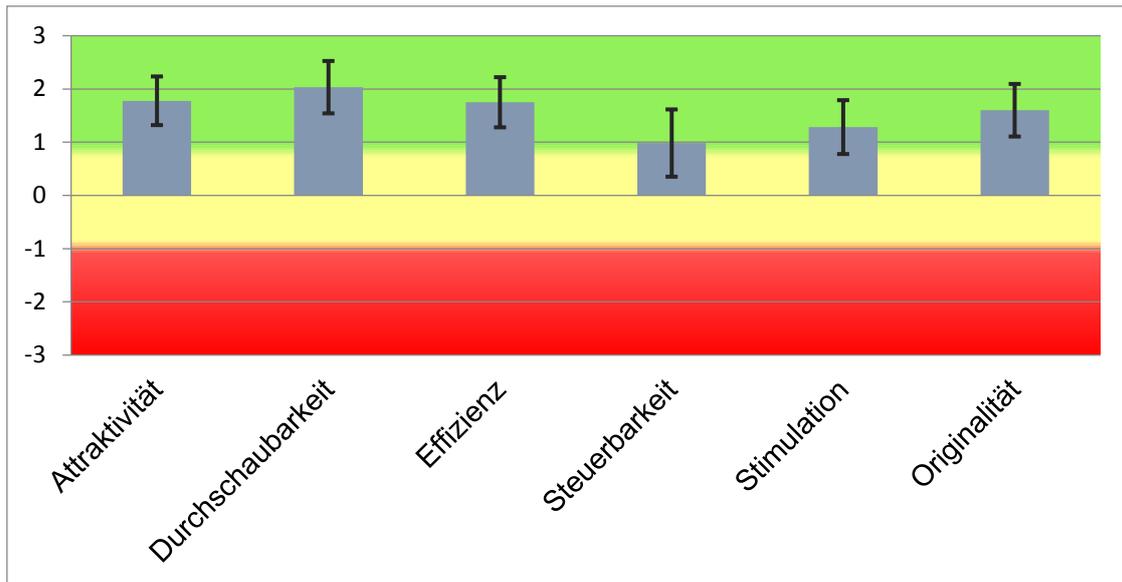


Tabelle 9: Bewertung der 6 Skalen des UEQ-Fragebogens (Pilotstudie I)

Skalenbewertung					
Attraktivität	↑	1,8	Steuerbarkeit	↑	1,0
Durchschaubarkeit	↑	2,0	Stimulation	↑	1,3
Effizienz	↑	1,8	Originalität	↑	1,6
<i>Bewertungsschema</i>		<i>&gt; 0,8 bis 3 positive Bewertung</i> <i>≥ -0,8 bis ≤ 0,8 neutrale Bewertung</i> <i>&lt; -0,8 bis -3 negative Bewertung</i>			

Besonders positiv bewerteten die Patienten die Durchschaubarkeit mit einer Wertung von 2,0 bei einem maximal erreichbaren Wert von +3. Am schlechtesten schnitt die Steuerbarkeit mit einer Wertung von 1,0 ab. Doch liegt dieser Wert damit trotzdem noch in einem positiven Ergebnisbereich.

Werden die einzelnen Skalen geschlechtsabhängig betrachtet, so wurde *SymptomCheck* in allen sechs Faktoren von den weiblichen Patienten der Pilotstudie I besser bewertet [siehe [Abbildung 16](#); [Tabelle 10](#)].

Besonders deutliche Unterschiede zeigen sich hierbei bezüglich der Steuerbarkeit und der Attraktivität, kaum Unterschiede ergeben sich in Hinblick auf die Stimulation. Die Steuerbarkeit wird von den männlichen Patienten bei einem Mittelwert von 0,7 neutral bewertet.

Abbildung 16: Geschlechtsabhängige Bewertung der 6 Skalen (Pilotstudie I)

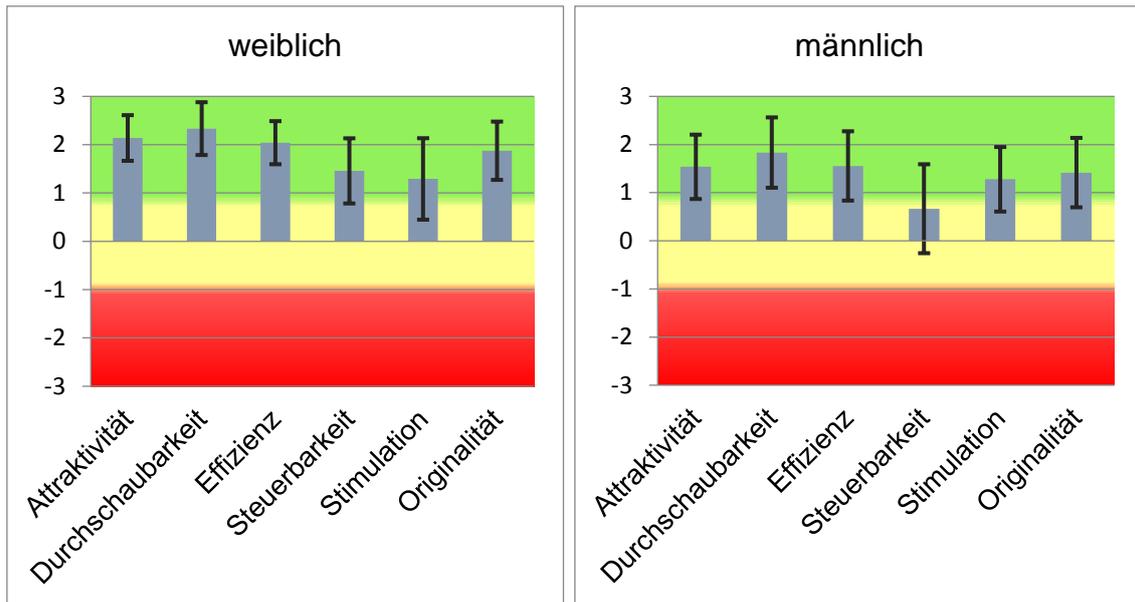


Tabelle 10: Geschlechtsabhängige Bewertung der 6 Skalen des UEQ-Fragebogens (Pilotstudie I)

Skalenbewertung		
	weiblich	männlich
Attraktivität	↑ 2,1	↑ 1,5
Durchschaubarkeit	↑ 2,3	↑ 1,8
Effizienz	↑ 2,0	↑ 1,6
Steuerbarkeit	↑ 1,5	↔ 0,7
Stimulation	↑ 1,3	↑ 1,3
Originalität	↑ 1,9	↑ 1,4

### 8.3 Überprüfung der Verdachtsgenerierung

Neben der Benutzerfreundlichkeit untersuchten wir, ob die Version der Pilotstudie I von *SymptomCheck* die zutreffenden Diagnosen unter den Top 10 Verdachtsdiagnosen aufführen kann. Dafür verglichen wir die Verdachtsdiagnosen von *SymptomCheck* mit der aktuellen, aus dem Arztbrief ersichtlichen, Diagnose.

Hierbei gestaltete sich die Auswertung schwierig, wie im Folgenden erläutert wird:

Neun der insgesamt 16 Patienten kamen lediglich zur Verlaufskontrolle. Von diesen neun Patienten waren drei bezüglich der für die Verlaufskontrolle relevanten Diagnose beschwerdefrei und die Diagnosen eines Patienten mit

schwerwiegender Vorerkrankung (Z.n. Herz- und Nierentransplantation) nicht zur Auswertung geeignet. Denn für Personen mit einer derartigen Krankengeschichte wurde *SymptomCheck* nicht konzipiert. Zudem lassen sich viele chronische Erkrankungen der Patienten, wie z. B. eine Hypothyreose, nicht mit *SymptomCheck* auswerten, da deren Symptome bereits medikamentös therapiert werden. Somit sind viele aus dem Arztbrief ersichtlichen Erkrankungen kein Vorstellungsgrund der Patienten. Ein Patient kam im Rahmen einer Erstvorstellung für eine Adipositas-Studie, ein anderer zur Komplettierung seiner Staging-Untersuchungen und ein dritter zur kardialen Abklärung vor Beginn einer Chemotherapie. Eine andere schwangere Patientin kam nach stationärem Aufenthalt aus der Frauenklinik, wo sie sich zum Ausschluss eines Apoplex befand. Ein Patient brach die Patientenbefragung vorzeitig ab, sodass dessen Diagnosen nicht mit *SymptomCheck* abgeglichen werden konnten. Nur 3 Patienten kamen erstmalig zur kardialen Abklärung. Bei einem der drei Patienten zeigte sich in der Untersuchung ein unauffälliger Befund. Details sind **Tabelle 47** zu entnehmen. Um die Verdachtsgenerierung der Pilotversion überprüfen zu können, beschränkten wir uns in der Auswertung lediglich auf die Diagnosen der 10 Patienten, die anhand der eingegebenen Symptome reliabel registriert werden konnten [siehe **Tabelle 11**].

**Tabelle 11: Diagnoseliste (Pilotstudie I)**

<b>Diagnose</b>	<b>Anzahl (n)</b>
stabile Angina pectoris	2
tachykarde Herzrhythmusstörung	1
Vorhofflimmern	1
chronische Herzinsuffizienz	4
akute Aortenklappeninsuffizienz	1
chronische Mitralklappeninsuffizienz	1
Depression	2
COPD	1
Asthma bronchiale	1
Bronchialkarzinom	1
Schlafapnoesyndrom	2
arterielle Hypertonie	3
Schwangerschaft	1
Rauchen, Nikotinabusus	1
psychische Ursache (Stress, Aufregung, Trauer, etc.)	1
Diagnose nicht in <i>SymptomCheck</i> enthalten	1

Für die Auswertung einer aus dem Arztbrief zu entnehmenden Diagnose können mehrere Verdachtsdiagnosen in *SymptomCheck* zutreffen. In unseren Fällen waren dies die Herzinsuffizienz, die Depression, das Vorhofflimmern und die Aortenklappeninsuffizienz. Ist bei einem Patienten eine Herzinsuffizienz diagnostiziert, so kann diese unter der Pilotversion von *SymptomCheck* als diversen Unterformen aufgeführt werden:

- Akute Linksherzinsuffizienz
- Chronische Herzinsuffizienz
- Leichte Linksherzinsuffizienz
- Mäßige Linksherzinsuffizienz
- Starke Linksherzinsuffizienz
- Chronische Rechtsherzinsuffizienz

Für unsere Auswertung berücksichtigten wir lediglich die am höchsten im Ranking platzierte Diagnose. Diese war stets die chronische Herzinsuffizienz, weshalb nur diese in der Auswertung mitberücksichtigt ist. In gleicher Weise wurde die Diagnose „Aortenklappeninsuffizienz“ ausgewertet.

Bezogen auf die Depression berücksichtigten wir die Verdachtsdiagnose „Depression“, aber auch die allgemeine Verdachtsdiagnose „psychische Ursache“. Dasselbe gilt für das „Vorhofflimmern“, welches sowohl als „Vorhofflimmern“ als auch als allgemein „tachykarde Herzrhythmusstörung“ gewertet wird. Die Diagnose eines Patienten lautete Non-Hodgkin-Lymphom. Diese Erkrankung ist in der Pilotversion von *SymptomCheck* nur in ihren diversen Unterformen (z. B. als „Diffuses großzelliges B-Zell-Lymphom“ oder „Burkitt-Lymphom“) aufgeführt. Da im Arztbrief keine Angabe zur genaueren Bezeichnung gemacht wurde, ist diese Diagnose in der Auswertung nur unter „nicht erkannt“ aufgeführt. Von den 24 Diagnosen befanden sich 4 unter den Top 10 der Verdachtsdiagnosen [siehe [Tabelle 12](#)]. Dies entspricht einem Anteil von 16,7% [siehe [Abbildung 17](#)]. Betrachtet man die Top 20, so wurden 37,5% der Diagnosen innerhalb dieses Intervalls angezeigt. In der Pilotversion von *SymptomCheck* wurde der größte Anteil mit 33,3% zwischen Platz 21 und 40 in der Verdachtsdiagnoseliste angezeigt.

Abbildung 17: Überprüfung der Verdachtsgenerierung: Überblick über die Platzierung der Diagnosen (Pilotstudie I)

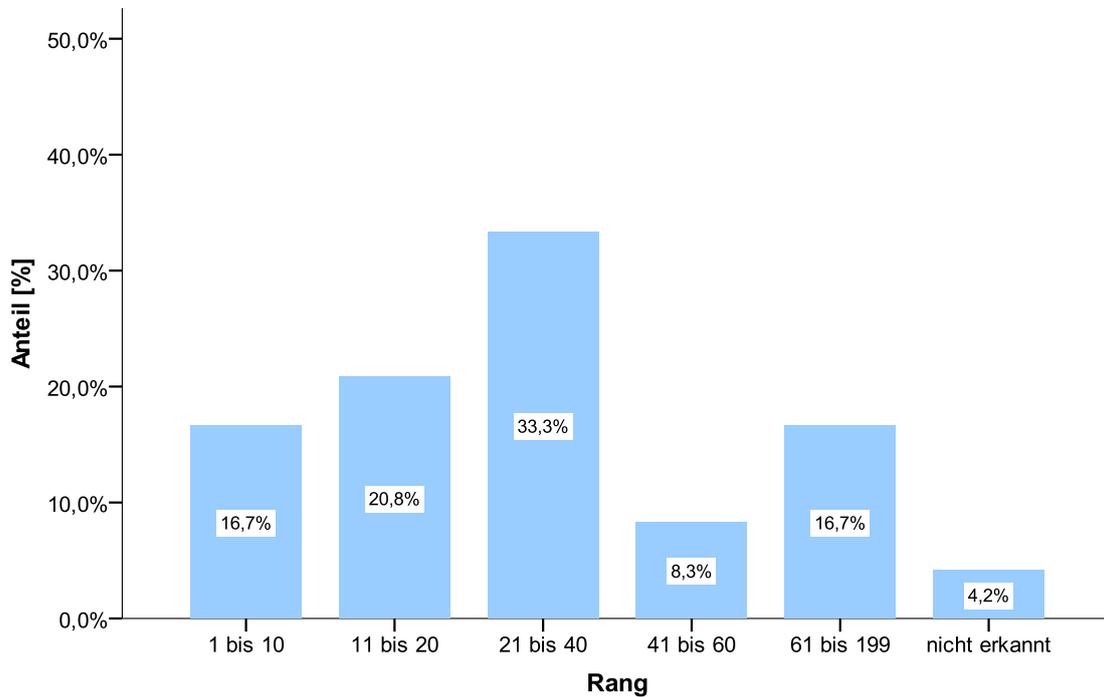


Tabelle 12: Platzierung der Diagnosen (Pilotstudie I)

Rang	Häufigkeit [n]	Anteil [%]	Kumulierte Prozente
1 bis 10	4	16,7	16,7
11 bis 20	5	20,8	37,5
21 bis 40	8	33,3	70,8
41 bis 60	2	8,3	79,2
61 bis 199	4	16,7	95,8
nicht erkannt	1	4,2	100,0
Gesamt	24	100,0	

Zusammenfassend lässt sich sagen, dass die Evaluation der Benutzerfreundlichkeit sehr positiv ausfiel. Bei der Auswertung zur Genauigkeit der Verdachtsgenerierung zeigte sich *SymptomCheck* für seine endgültige Version noch verbesserungsfähig.

## 9 Online-Evaluation

### 9.1 Rahmendaten

Die Online-Evaluation von *SymptomCheck* erfolgte vom 07.09.2015 bis zum 06.10.2015. Insgesamt wurden 75 Evaluationsbögen ausgefüllt und ausgewertet. Jedoch lässt sich dadurch nicht direkt auf 75 Teilnehmer schließen, da ein Teilnehmer eventuell auch mehrfach daran teilgenommen hat. Für die Auswertung wird aber von dem Fall ausgegangen, dass jeder Evaluationsbogen einem Teilnehmer entspricht.

Ein zusätzlicher 76. Fragebogen wurde gänzlich unausgefüllt abgeschickt, so dass dieser nicht in die Bewertung mit einbezogen wurde.

Von diesen 75 Teilnehmern waren 26 männlich (34,7%) und 49 weiblich (65,3%) [siehe **Abbildung 18**; **Tabelle 37**]. Der Mittelwert liegt bei 30, der Median hingegen bei 26 Jahren. Betrachtet man die Altersgruppen, so befindet sich mit 77,3 % der Hauptanteil der Befragten in der Altersgruppe der 21- bis 30-Jährigen [siehe **Abbildung 19**; **Tabelle 38**]. Dies ist höchstwahrscheinlich auf die Veröffentlichung der Evaluation mittels Facebook zurückzuführen, da man davon ausgehen kann, dass besonders diese Altersgruppe sehr präsent in diesem sozialen Netzwerk ist. Die anderen Altersgruppen sind nur in geringen Anteilen vertreten.

Auch die Verteilung der Bildungsabschlüsse ist nicht gleichmäßig. 89,3% der Befragten gaben an, eine allgemeine oder fachgebundene Hochschulreife zu besitzen, nur 9,3% haben die mittlere Reife erlangt und lediglich 1,3% einen Hauptschulabschluss [siehe **Abbildung 20**; **Tabelle 39**].

Dies könnte zum einen daran liegen, dass weniger Personen mit mittlerer Reife oder einem Hauptschulabschluss über die Befragung informiert wurden oder dass die Testpersonen mit Abitur oder Fachabitur motivierter waren an dieser Evaluation teilzunehmen.

Abbildung 18: Geschlechterverteilung (Online-Evaluation; n=75)

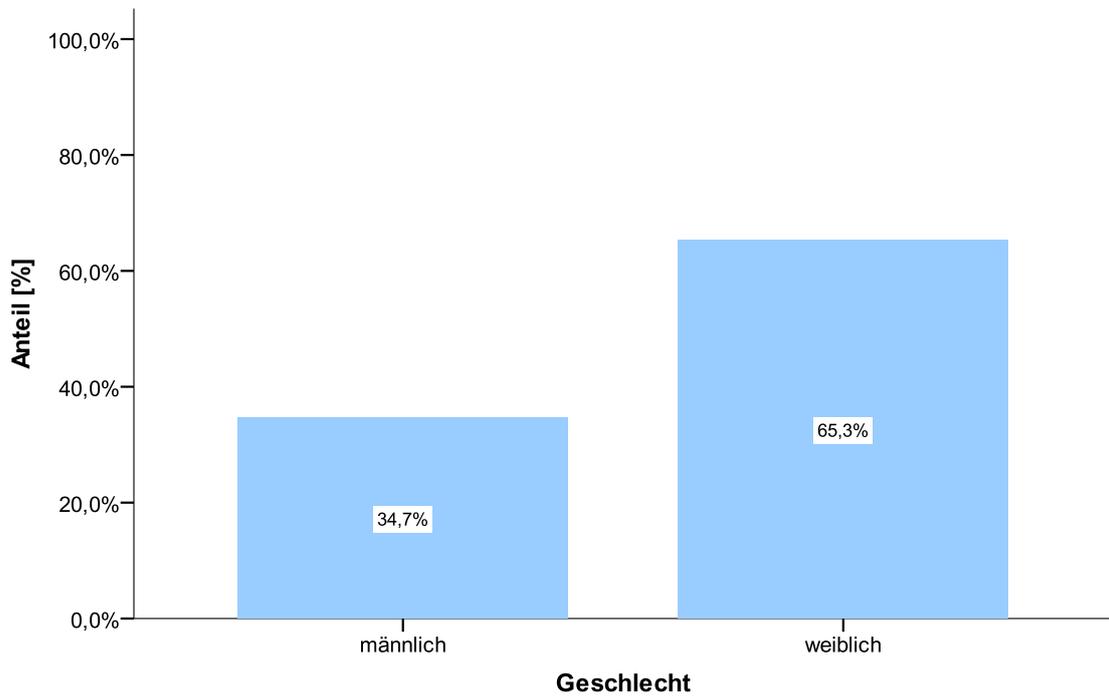


Abbildung 19: Altersverteilung (Online-Evaluation; n=75)

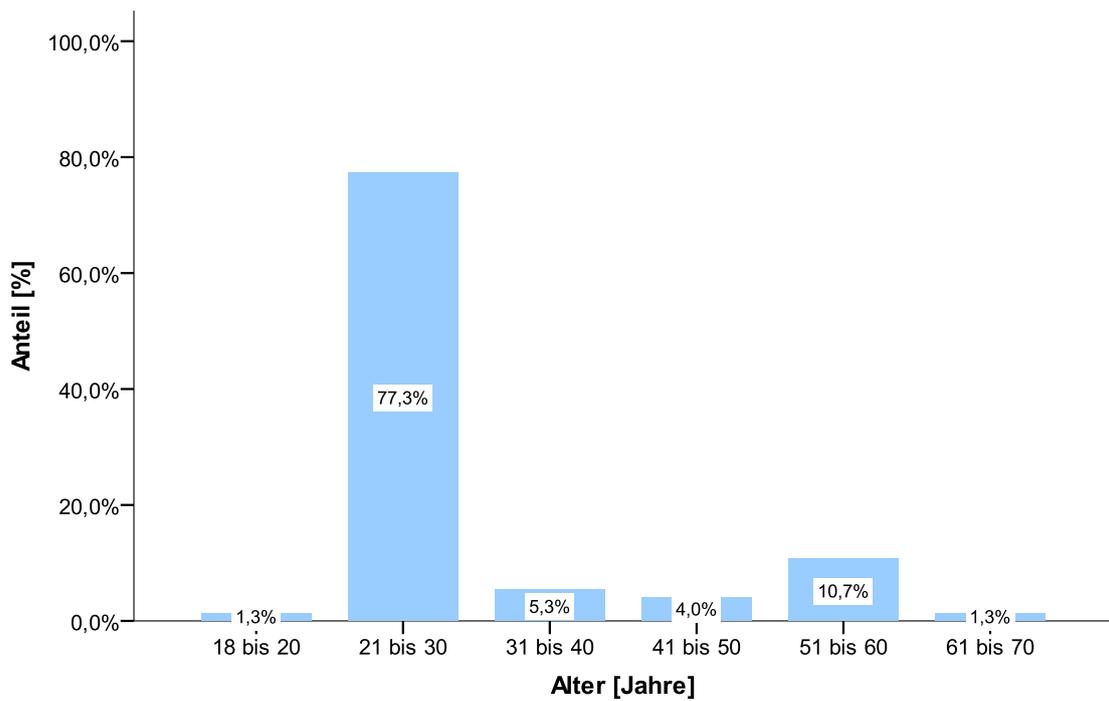
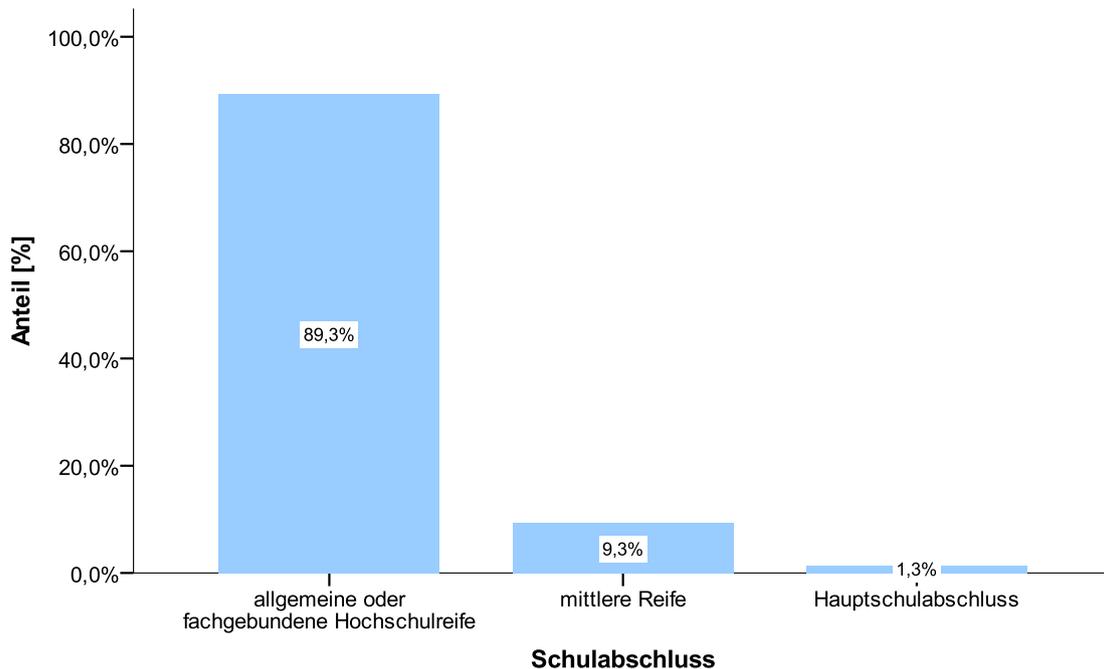


Abbildung 20: Aufteilung der Probanden nach Schulabschlüssen (Online-Evaluation; n=75)

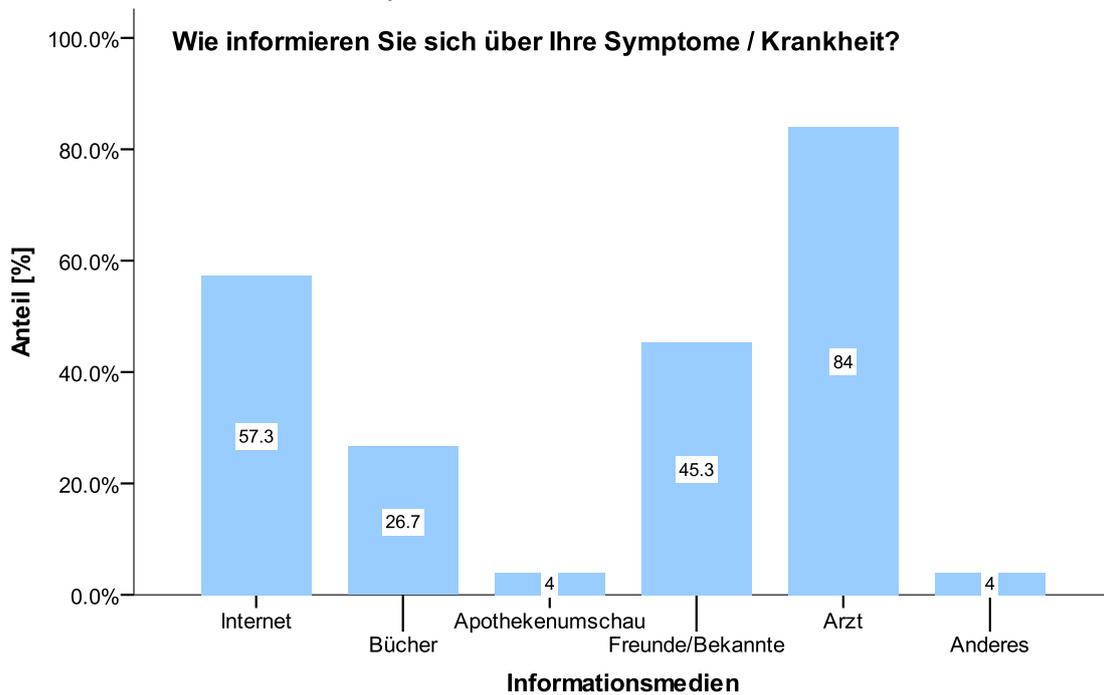


Wie in **Tabelle 13** ersichtlich, stellt der Arzt für 84,0% der Befragten die häufigste Informationsquelle dar. Dessen Konsultation steht heutzutage laut der Online-Evaluation noch immer an erster Stelle. Das Internet wird von 57,3% für symptom- oder krankheitsspezifische Informationen hinzugezogen und stellt das zweithäufigste Informationsmedium dar.

Tabelle 13: Medien zur Information über Symptome und Krankheiten (Online-Evaluation)

Wie informieren Sie sich über Ihre Symptome / Krankheit?	Antworten		
	Anzahl (n)	Anteil (%)	Prozent der Fälle
Internet	43	25,9%	57,3%
Bücher	20	12,0%	26,7%
Apothekenumschau	3	1,8%	4,0%
Freunde/Bekannte	34	20,5%	45,3%
Arzt	63	38,0%	84,0%
Anderes	3	1,8%	4,0%
Gesamt	165	100,0%	221,3%

Abbildung 21: Medien zur Information über Symptome und Krankheiten (Online-Evaluation; n=75)



Zwei Personen, die sich über das Internet über Ihre Symptome und Krankheiten informieren, enthielten sich der Angabe einer konkreten Internetseite. Von den übrigen Personen, die das Internet als medizinische Informationsquelle zu Rate ziehen, benutzen 75,6% „Wikipedia“, 36,6% „Netdoktor“, 9,8% „Onmeda“, 14,6% „Gutefrage.net“, 12,2% „Apotheken-umschau.de“ und 12,2% eine andere Internetseite [siehe [Tabelle 14](#)].

Tabelle 14: Internetseiten als medizinische Informationsquelle (Online-Evaluation)

Welche Seite benutzen Sie dafür am häufigsten?	Antworten		
	Anzahl [n]	Anteil [%]	Prozent der Fälle
Wikipedia	31	47,0%	75,6%
Netdoktor	15	22,7%	36,6%
Onmeda	4	6,1%	9,8%
Gutefrage.net	6	9,1%	14,6%
Apotheken-umschau.de	5	7,6%	12,2%
Andere	5	7,6%	12,2%
Gesamt	66	100,0%	161,0%

Da diese Frage von vielen Personen mit mehr als einer Antwort beantwortet wurde, lässt sich, wie auch in der klinischen Studie, keine Aussage darüber treffen, welche Seite am häufigsten benutzt wird.

## 9.2 Benutzungsqualität

Für eine Aussage zur Benutzerfreundlichkeit ist die Auswertung der Online-Evaluation essentiell, da *SymptomCheck* für Personen konzipiert wurde, die sich bei Sorgen über ihre Beschwerden selbstständig zu Hause am Computer durch *SymptomCheck* eine Übersicht über mögliche Diagnosen erstellen lassen können.

### 9.2.1 UEQ-Fragebogen

Von den insgesamt 75 Fragebögen wurden 53 vollständig, 14 teilweise und 8 gar nicht bezüglich der 26 Items zur Benutzungsqualität ausgefüllt. Die 8 gänzlich unausgefüllten UEQ-Fragebögen wurden nicht in die Auswertung einbezogen. Unvollständige Fragebögen wurden hingegen berücksichtigt.

Die Interpretation der Tabellen und Abbildungen ist in [5.6 UEQ-Fragebogen zur Benutzerfreundlichkeit](#) erläutert.

*SymptomCheck* erzielte in der Online-Studie bezüglich der Skalen „Durchschaubarkeit“ und „Steuerbarkeit“ positive Ergebnisse. Betrachtet man die Skalen „Attraktivität“, „Effizienz“, „Stimulation“ und „Originalität“, so wurden diese neutral-positiv bewertet [siehe [Tabelle 15](#); [Abbildung 22](#)].

Genauere Details zur Bewertung der einzelnen Items sind in [Tabelle 41](#) ausführlich dargestellt.

Abbildung 22: Bewertung der 6 Skalen (Online-Evaluation)

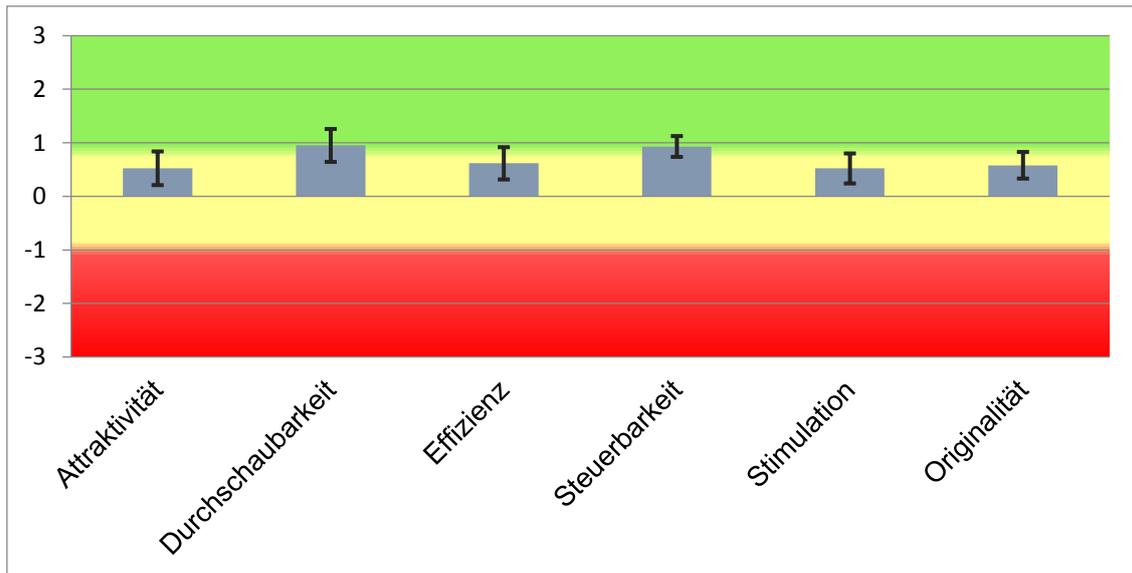


Tabelle 15: Bewertung der 6 Skalen (Online-Evaluation)

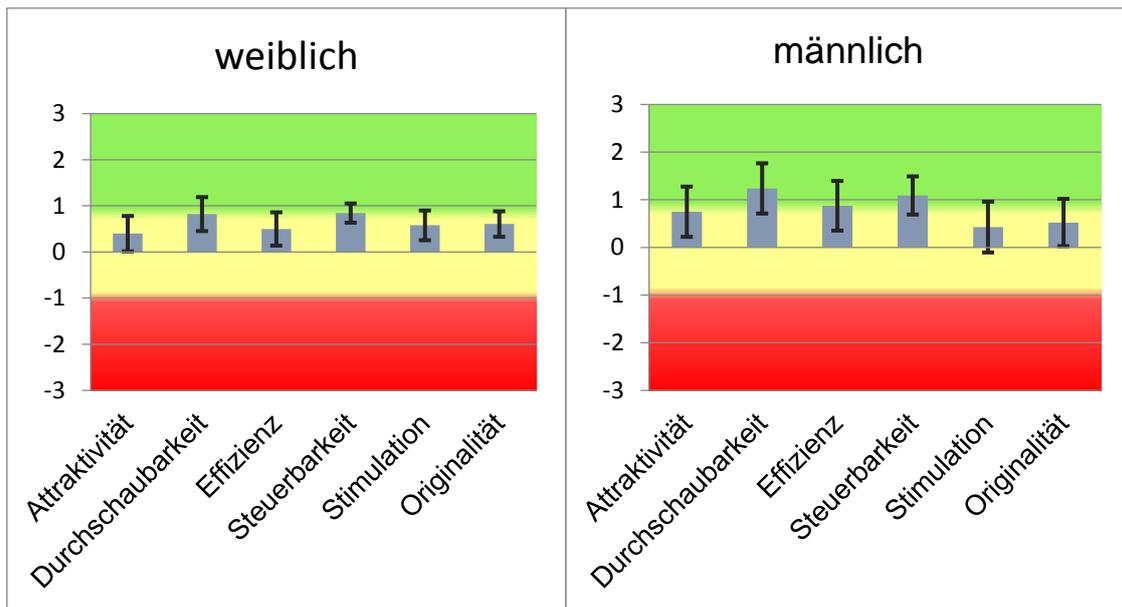
Skalenbewertung					
Attraktivität	↔	0,5	Steuerbarkeit	↑	0,9
Durchschaubarkeit	↑	1,0	Stimulation	↔	0,5
Effizienz	↔	0,6	Originalität	↔	0,6
<i>Bewertungsschema</i>	<i>&gt; 0,8 bis 3 positive Bewertung</i> <i>≥ -0,8 bis ≤ 0,8 neutrale Bewertung</i> <i>&lt; -0,8 bis -3 negative Bewertung</i>				

Betrachtet man die einzelnen Skalen geschlechtsabhängig, so wurde *SymptomCheck* größtenteils von den männlichen Personen besser bewertet [siehe [Tabelle 16](#); [Abbildung 23](#)].

Tabelle 16: Geschlechtsabhängige Bewertung der 6 Skalen (Online-Evaluation)

Skalenbewertung		
	weiblich	männlich
Attraktivität	↔ 0,4	↔ 0,8
Durchschaubarkeit	↑ 0,8	↑ 1,2
Effizienz	↔ 0,5	↑ 0,9
Steuerbarkeit	↑ 0,8	↑ 1,1
Stimulation	↔ 0,6	↔ 0,4
Originalität	↔ 0,6	↔ 0,5

Abbildung 23: Geschlechtsabhängige Bewertung der 6 Skalen (Online-Evaluation)



Während *SymptomCheck* von den weiblichen Personen nur bezüglich der Skalen „Durchschaubarkeit“ und „Steuerbarkeit“ positiv und bezüglich der restlichen Skalen neutral bewertet wurde, schnitt das Programm bei den männlichen Probanden zusätzlich im Bereich „Effizienz“ positiv ab.

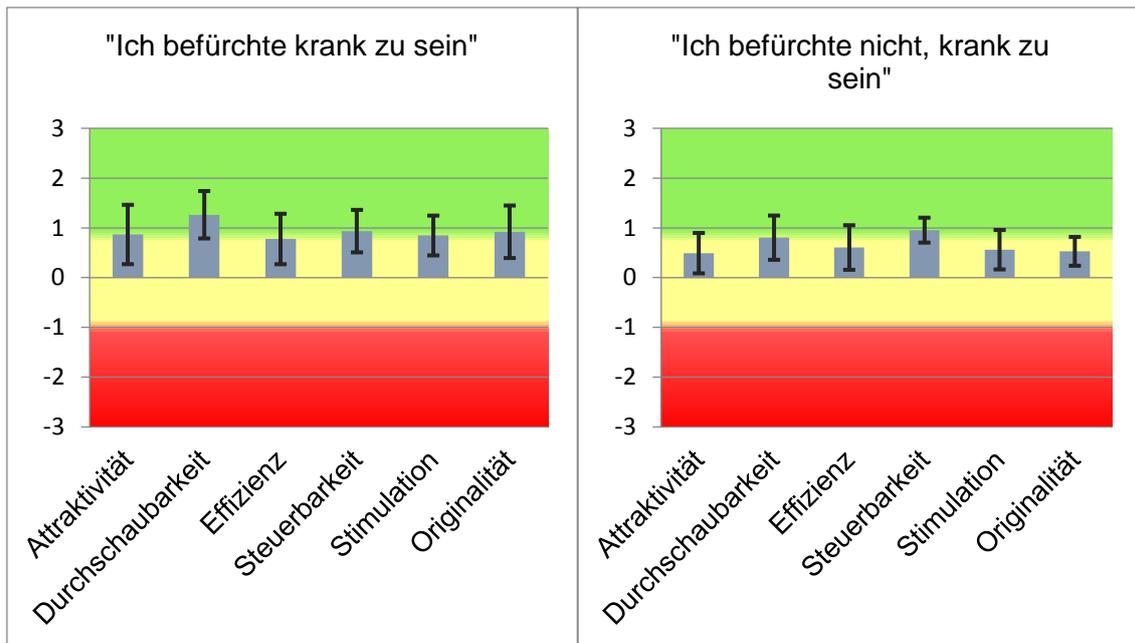
Desweiteren wurde untersucht, ob Unterschiede zwischen Personen, welche aktuell befürchten krank zu sein (Gruppe 1) und denen, die nicht befürchten krank zu sein (Gruppe 2), bestehen. Zwei der 75 Personen enthielten sich der Angabe und wurden nicht in die Auswertung einbezogen [siehe [Tabelle 17](#)]. Insgesamt gaben 21 Personen (28,8%) an, dass sie befürchten krank zu sein. 43 Personen (58,9%) befürchten dies nicht, 9 verhielten sich neutral (12,3%).

Tabelle 17: Bewertung der Aussage "Ich befürchte krank zu sein." (Online-Evaluation)

	Aussage „Ich befürchte krank zu sein.“	Anzahl (n)	Anteil [%]	Gültige Prozente	Kumulierte Prozente
Gruppe 1	trifft völlig zu	8	10,7	11,0	11,0
	trifft eher zu	13	17,3	17,8	28,8
	neutral	9	12,0	12,3	41,1
Gruppe 2	trifft eher nicht zu	21	28,0	28,8	69,9
	trifft gar nicht zu	22	29,3	30,1	100,0
	Gesamt (angegeben)	73	97,3	100,0	
	nicht angegeben	2	2,7		
	Gesamt	75	100,0		

Interessant ist, dass *SymptomCheck* von den Personen, die befürchten krank zu sein, in 5 von 6 Skalen besser bewertet wurde als von denen, die nicht befürchten krank zu sein [siehe **Abbildung 24**; **Tabelle 18**].

**Abbildung 24:** Unterschiede bezüglich der Bewertung der 6 Skalen hinsichtlich der unterschiedlichen Bewertung der Aussage "Ich befürchte krank zu sein" (Online-Evaluation)



**Tabelle 18:** Unterschiede bezüglich der Bewertung der 6 Skalen hinsichtlich der unterschiedlichen Bewertung der Aussage "Ich befürchte krank zu sein" (Online-Evaluation)

	„Ich befürchte krank zu sein.“	
	Ja (Gruppe 1) <sup>1</sup>	Nein (Gruppe 2) <sup>2</sup>
Attraktivität	↑ 0,9	↔ 0,5
Durchschaubarkeit	↑ 1,3	↑ 0,8 <sup>3</sup>
Effizienz	↔ 0,8	↔ 0,6
Steuerbarkeit	↑ 0,9	↑ 1,0
Stimulation	↑ 0,8 <sup>3</sup>	↔ 0,6
Originalität	↑ 0,9	↔ 0,5

<sup>1</sup>entspricht den Antworten „trifft völlig zu“ und „trifft eher zu“  
<sup>2</sup>entspricht den Antworten „trifft eher nicht zu“ und „trifft gar nicht zu“  
<sup>3</sup>Mittelwert >0,8, durch die Darstellung mit nur einer Dezimalen jedoch in der Tabelle nicht darstellbar

Während Gruppe 2 nur die Skalen „Durchschaubarkeit“ und „Steuerbarkeit“ positiv bewertete, schnitten bei Gruppe 1 auch die Skalen „Attraktivität“, „Stimulation“ und „Originalität“ positiv ab.

Bei Bedarf können die Ergebnisse der einzelnen Items in der Appendix eingesehen werden [siehe **Tabelle 42**; **Tabelle 43**]. Alle restlichen Skalen wurden von beiden Gruppen neutral bewertet. Dieses Ergebnis ist äußerst erfreulich, da im Alltag vor allem diejenigen Personen *SymptomCheck* benutzen sollen, wenn sie Angst haben krank zu sein. Folglich kann in **Abbildung 24** dargestellte Teil „Ich befürchte krank zu sein“ stellvertretend für die wohl am besten zutreffende Benutzerfreundlichkeit gesehen werden.

Desweiteren wurde überprüft, ob es signifikante Unterschiede bezüglich der Bewertung von *SymptomCheck* zwischen Gruppe 1 und Gruppe 2 gibt. Im exakten Test nach Fisher<sup>72</sup> zeigten sich in keiner der untersuchten Domäne signifikante Unterschiede. Die Details sind der Appendix zu entnehmen [siehe **Tabelle 29**]. Aufgrund der sehr ungleichen Alters- und Bildungsgradverteilung lassen sich diesbezüglich keine Aussagen zu möglichen Unterschieden in der Benutzerfreundlichkeit treffen.

### **9.2.2 Bewertung spezifischer Aussagen zu *SymptomCheck***

Von den insgesamt 75 Fragebögen wurden 58 vollständig, 15 teilweise und 2 gar nicht bezüglich der spezifischen Aussagen zu *SymptomCheck* ausgefüllt.

Neben dem UEQ-Fragebogen wurden in **Tabelle 19** die Nutzerbewertungen folgender wichtiger Aussagen erhoben, die spezifischere Informationen zu *SymptomCheck* liefern.

Es werden nun lediglich die relevanten Punkte hervorgehoben:

Im Allgemeinen schnitt *SymptomCheck* bezüglich dieser Aussagen durchweg positiv ab. Relativ ausgewogen ist jedoch nur die Bewertung von Aussage 5. Demnach half *SymptomCheck* zwar 38,5% der befragten besser als andere Informationsplattformen im Internet, jedoch standen auch 30,8% der Aussage neutral gegenüber und weitere 30,8% verneinten diese Aussage.

Tabelle 19: Spezifische Bewertungen zu *SymptomCheck* (Online-Evaluation)

	<b>Aussage</b>	<b>Anteil (%)</b>		
		<b>positiv</b>	<b>neutral</b>	<b>negativ</b>
1.	Ich konnte meine Symptome gut finden.	73,8	1,5	24,6
2.	Unklare Begriffe wurden mir gut erklärt.	65,6	14,8	19,7
3.	Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.	96,9	1,5	1,5
4.	Dieses Programm ist im Alltag sehr nützlich.	64,6	15,4	20
5.	Das Programm hat mir besser geholfen, als andere Informationsplattformen im Internet.	38,5	30,8	30,8
6.	Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	77,8	6,3	15,9
7.	Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel).	90,8	3,1	6,2
		<b>zu lang</b>	<b>genau richtig</b>	<b>zu kurz</b>
8.	Die Dauer für die Beantwortung der Fragen war...	4,3	81,4	14,3

Genauere Ergebnisse zu den einzelnen Aussagen sind der Appendix [siehe **Tabelle 30**] zu entnehmen.

Auch hier wurde, um ein in der Realität relevantes Patientenkollektiv zu untersuchen, nach Unterschieden [siehe **Tabelle 20**] zwischen Personen gesucht, welche aktuell befürchten krank zu sein (Gruppe 1) und denen, die nicht befürchten krank zu sein (Gruppe 2).



Gruppe 1 bewertete 6 von 7 Aussagen positiver als Gruppe 2. Lediglich Aussage 7 („Es ist nützlich, dass mir auch andere Erklärungen [...] vorgeschlagen werden.“) wird von Gruppe 2 um 1,4% geringfügig positiver bewertet. Desweiteren überwiegen in allen Aussagen außer Aussage 5 von Gruppe 2 anteilmäßig die positiven gegenüber den neutralen und negativen Bewertungen. Die Personen von Gruppe 1 konnten mit 84,2% ihre Symptome gut finden. Gerade Personen, die befürchten krank zu sein, leiden eher an Beschwerden als Personen aus Gruppe 2. Somit ist es besonders wichtig, dass genau diese Gruppe ihre Symptome einfach und schnell finden kann. Auch die alltägliche Nützlichkeit wurde mit 78,9% von Gruppe 1 im Vergleich zu den 54,1% von Gruppe 2 wesentlich positiver bewertet. Desweiteren hielten 52,6% (Gruppe 1) das Programm für eine bessere Hilfe als andere Informationsplattformen im Internet. In Gruppe 2 waren dies nur 35,1%. Auch hier sind, wie oben erklärt, die Ergebnisse von Gruppe 1 besonders aussagekräftig, da genau diese Personen-Gruppe im Alltag am ehesten *SymptomCheck* benutzen würde. Äußerst relevant ist zudem die Evaluation der Benutzungsdauer von *SymptomCheck*. Erfreulicherweise wurde diese von Gruppe 1 mit 84,2% und Gruppe 2 mit 78,0% als „genau richtig“ lange dauernd eingeschätzt. Dieser Aspekt war bei der Entwicklung des Programms ein wichtiger Faktor, da wir der Ansicht sind, dass der Alltagserfolg eines solchen Programms nur durch eine fundierte Symptomabfrage innerhalb eines für den Benutzer angenehmen Zeitrahmens gewährleistet werden kann. Zur Überprüfung eines möglichen signifikanten Unterschieds zwischen den beiden Gruppen bezogen auf diese sieben Aussagen wurde der Chi-Quadrat-Test<sup>72</sup> durchgeführt. Hierbei war jedoch kein Unterschied signifikant. Beim Chi-Quadrat-Test wurde die erwartete Häufigkeit aufgrund einer für diesen Test zu geringen Fallzahl meist nicht erreicht. Daher wurde der exakte Test nach Fisher angewendet, welcher auch auf einen kleinen Stichprobenumfang angewendet werden kann und zuverlässige Resultate liefert<sup>72</sup>. Doch auch dessen Ergebnisse waren nicht signifikant [siehe [Tabelle 31](#)].

Ebenso wurden diese Aussagen auf geschlechtsspezifische Unterschiede anhand dieser beiden eben erläuterten Tests untersucht. Doch auch in diesem Aspekt zeigten sich keine signifikanten Ergebnisse [siehe [Tabelle 32](#)].

Altersspezifische Unterschiede waren schwer auswertbar, da der Schwerpunkt der Altersverteilung auf den 20- bis 30-Jährigen liegt und die restlichen Testpersonen sehr ungleichmäßig verteilt sind [siehe **Abbildung 19**]. Das Personenkollektiv wurde in die Gruppe der 18- bis 30-Jährigen und in die der 31- bis 63-Jährigen eingeteilt, da der älteste Patient 63 Jahre alt war. Dadurch sollte der Hauptanteil der jungen Testpersonen mit dem Rest verglichen werden. Auch hierbei war nur der exakte Test nach Fisher auswertbar, laut dem es keine signifikanten Unterschiede zwischen beiden Altersgruppen gibt [siehe **Tabelle 33**]. Ebenso können keine Aussagen über mögliche Unterschiede bezüglich der Benutzerfreundlichkeit getroffen werden, da auch hier die Verteilung sehr ungleichmäßig erfolgte.

## **10 Pilotstudie II**

### **10.1 Rahmendaten**

An der Studie in der Medizinischen Klinik I nahmen insgesamt 33 Patienten teil. Von diesen Patienten testeten allerdings drei das Programm nicht selbstständig und konnten sich somit bezüglich der Benutzerfreundlichkeit nicht äußern.

Insgesamt wurden 22 männliche (66,7%) und 11 weibliche (33,3%) Patienten befragt [siehe **Abbildung 25; Tabelle 37**].

Der Altersmittelwert beträgt hierbei 58 Jahre der Altersmedian liegt bei 60 Jahren. Die genaue Verteilung ist in **Tabelle 38** und in **Abbildung 26** ersichtlich.

Bezüglich der Schulabschlüsse lässt sich sagen, dass 12 Patienten (36,4%) einen allgemeinen oder fachgebundenen Hochschulabschluss, 7 Patienten (21,2%) die mittlere Reife, 11 Patienten (33,3%) einen Hauptschulabschluss und 2 Patienten (6,1%) keinen oder einen anderen Abschluss besitzen. Ein Patient (3%) enthielt sich der Altersangabe [siehe **Abbildung 27, Tabelle 39**].

Wie auch in der Pilotstudie I ist hier eine geschlechtsspezifische Verteilung bezüglich der Altersklassen oder des Schulabschlusses nicht sinnvoll, da die Fallzahl (N=33) auch hier zu gering ist.

Abbildung 25: Geschlechterverteilung (Pilotstudie II; n=33)

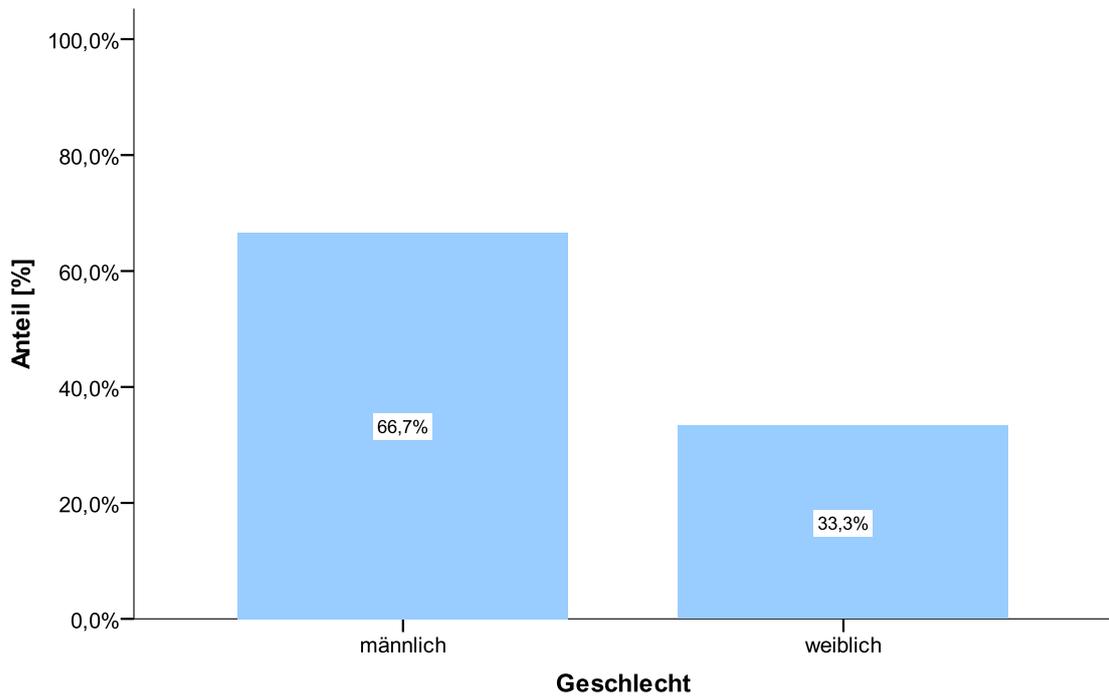


Abbildung 26: Altersverteilung (Pilotstudie II; n=33)

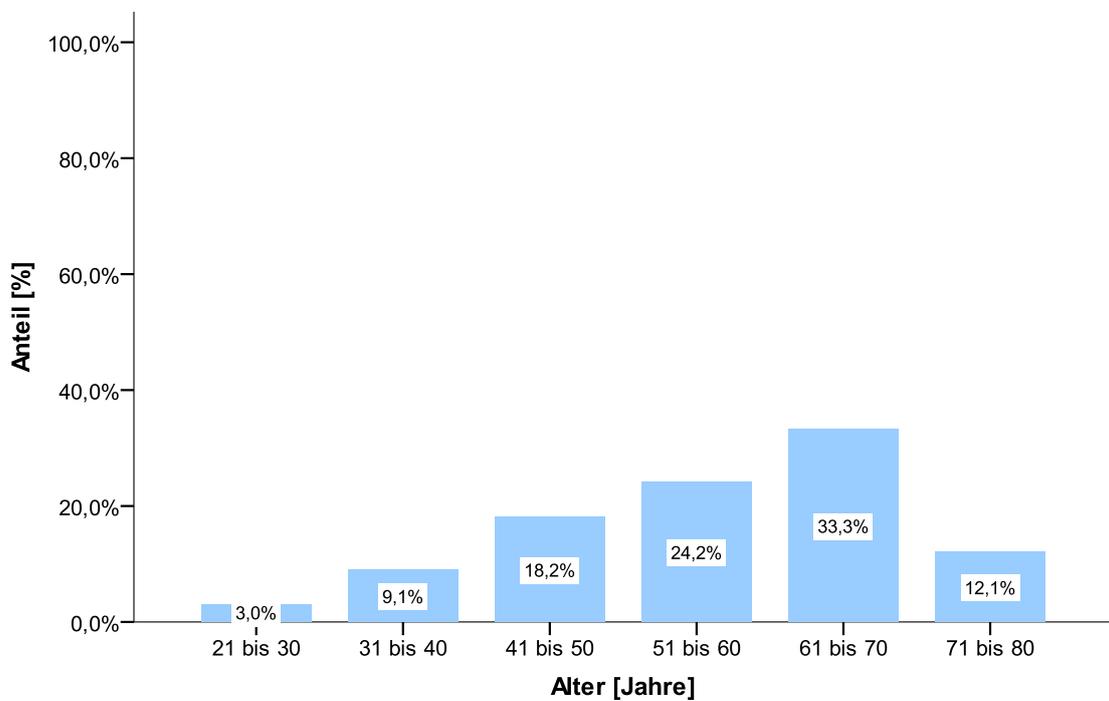
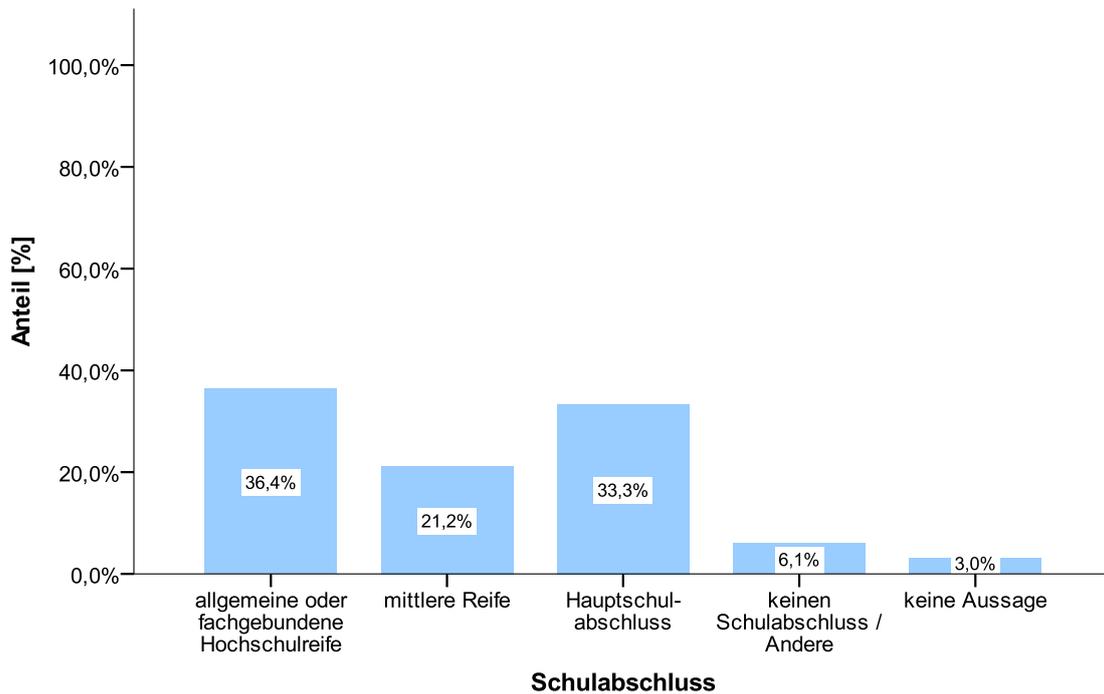


Abbildung 27: Aufteilung der Patienten nach Schulabschlüssen (Pilotstudie II; n=33)



Für dieses Patientenkollektiv ist der Arzt mit 84,8% die häufigste Informationsquelle bezüglich Symptome und Krankheiten. 45,5% der Befragten gaben an, das Internet für medizinische Fragen zu benutzen, 24,2% die Apothekenumschau. Desweiteren gaben 12,1% der Patienten an, Bücher zu Rate zu ziehen, weitere 12,1% wenden sich an Freunde und Bekannte und 15,2% informieren sich anderweitig [siehe **Tabelle 21**].

Tabelle 21: Medien zur Information über Symptome und Krankheiten (Pilotstudie II)

Wie informieren Sie sich über Ihre Symptome / Krankheit?	Antworten		
	Anzahl (n)	Anteil (%)	Prozent der Fälle
Internet	15	23,4	45,5
Bücher	4	6,3	12,1
Apothekenumschau	8	12,5	24,2
Freunde/Bekannte	4	6,3	12,1
Arzt	28	43,8	84,8
Anderes	5	7,8	15,2
<b>Gesamt</b>	<b>64</b>	<b>100,0</b>	<b>193,9</b>

Eine Person, die sich über das Internet über Ihre Symptome und Krankheiten informiert, enthielt sich der Angabe einer konkreten Internetseite. Von den übrigen Personen, die das Internet als medizinische Informationsquelle zu Rate ziehen, benutzen 100,0% „Wikipedia“, 28,6% „Netdoktor“, 0,0% „Onmeda“, und jeweils 7,1% „Gutefrage.net“, „Apotheken-umschau.de“ und eine andere Internetseite [siehe **Tabelle 22**]. Da diese Frage von vielen Personen mit mehr als einer Antwort beantwortet wurde, lässt sich keine Aussage darüber treffen, welche Seite am häufigsten benutzt wird.

**Tabelle 22: Internetseiten als medizinische Informationsquelle (Pilotstudie II)**

Welche Seite benutzen Sie dafür am häufigsten?	Antworten		
	Anzahl [n]	Anteil [%]	Prozent der Fälle
Wikipedia	14	66,7	100,0
Netdoktor	4	19,0	28,6
Onmeda	0	0,0	0,0
Gutefrage.net	1	4,8	7,1
Apotheken-umschau.de	1	4,8	7,1
Andere	1	4,8	7,1
Gesamt	21	100,0	150,0

## 10.2 Studiendauer für die einzelnen Teilnehmer

Die Studiendauer für jeweils einen Teilnehmer variierte stark von der Schnelligkeit und Gründlichkeit der Beantwortung der Fragebögen und von *SymptomCheck*. Die Personen benötigten hierbei zwischen 10 und 30 Minuten.

## 10.3 Benutzungsqualität

Auch die Patienten der Medizinischen Klinik wurden zur Benutzerfreundlichkeit von *SymptomCheck* befragt.

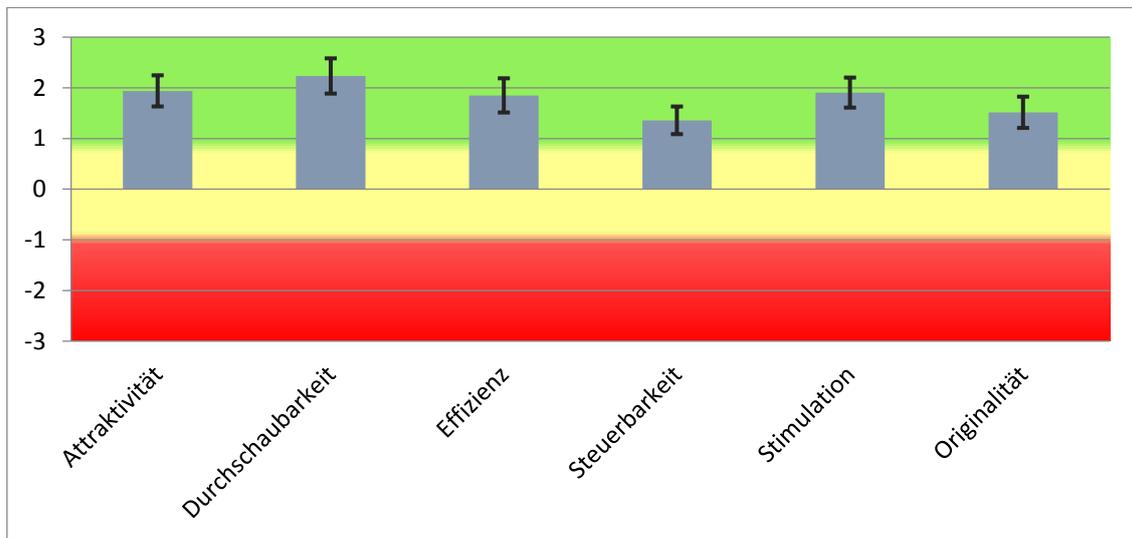
### 10.3.1 UEQ-Fragebogen

Hierbei wurden von den 33 Fragebögen 26 vollständig, 4 teilweise und 3 überhaupt nicht bezüglich der Benutzungsqualität ausgefüllt. Die Auswahl und

Auswertung der Fragebögen erfolgte analog zur Auswertung der Online-Evaluation.

Die Interpretation der Tabellen und Abbildungen ist in **5.6 UEQ-Fragebogen zur Benutzerfreundlichkeit** erläutert. *SymptomCheck* erzielt in der Pilotstudie II der Medizinischen Klinik I bezüglich aller sechs Skalen positive Ergebnisse [siehe **Abbildung 28; Tabelle 23**]. Genaue Details zur Bewertung der einzelnen Items sind in **Tabelle 44** dargestellt.

**Abbildung 28: Bewertung der 6 Skalen (Pilotstudie II)**



**Tabelle 23: Bewertung der 6 Skalen (Pilotstudie II)**

Skalenbewertung			
Attraktivität	↑ 1,9	Steuerbarkeit	↑ 1,4
Durchschaubarkeit	↑ 2,2	Stimulation	↑ 1,9
Effizienz	↑ 1,9	Originalität	↑ 1,5
<i>Bewertungsschema</i>	<i>&gt; 0,8 bis 3 positive Bewertung</i> <i>≥ -0,8 bis ≤ 0,8 neutrale Bewertung</i> <i>&lt; -0,8 bis -3 negative Bewertung</i>		

Vergleicht man die Ergebnisse dieser Studie mit denen der Online-Evaluation, so wurde die Benutzerfreundlichkeit von *SymptomCheck* von den Patienten durchweg besser bewertet als von den Online-Probanden [siehe **Abbildung 22; Tabelle 15**].

Ein Störfaktor für dieses Ergebnis könnte sein, dass die Patienten direkten Kontakt zum Doktoranden hatten, der diese über diese Studie informiert und den

Ablauf der Befragung geleitet hatte. Dadurch könnten die Patienten durch Sympathie zum Doktoranden das Programm besser bewertet haben. Eine andere auch denkbare Erklärung könnte sein, dass die Patienten in der Klinik ihre Diagnose teilweise schon wussten und diese in der Verdachtsdiagnoseliste von *SymptomCheck* als wahrscheinlich zutreffende Diagnose angezeigt bekamen. Somit konnten sie die Leistung von *SymptomCheck* anhand ihrer eigenen Diagnose validieren und bewerteten das Programm positiver.

Vergleicht man die Skalen geschlechtsabhängig [siehe **Tabelle 45**], schneidet *SymptomCheck* bei männlichen Patienten in den Skalen Effizienz, Verlässlichkeit und Originalität besser ab als bei weiblichen Patienten. Die Frauen bewerten die Skalen Attraktivität, Durchschaubarkeit und Stimulation besser. Somit zeichnet sich in der Gesamtheit kein geschlechtsspezifischer Unterschied aus. Dieses Ergebnis ist positiv, da *SymptomCheck* für alle Geschlechter konzipiert ist.

Es ist nicht möglich, valide Aussagen bezüglich altersspezifischer Unterschiede in der Benutzungsqualität abzugeben, da dafür die Fallzahl (N=30) zu gering ist [siehe **Tabelle 46**].

### **10.3.2 Bewertung spezifischer Aussagen zu *SymptomCheck***

Wie auch in der Online-Evaluation wurde *SymptomCheck* bezüglich dieser Aussagen durchweg positiv bewertet [siehe **Tabelle 24**]. Diese Ergebnisse wurden jedoch nicht so detailliert untersucht und dargestellt wie in der Online-Evaluation, da vor allem dort der Fokus auf der Untersuchung der Benutzungsqualität lag.

Verglichen mit der Online-Evaluation [siehe **Tabelle 19**] schnitt *SymptomCheck* in der Pilotstudie II bezüglich der Aussagen 1, 2, 4, 5, 6 und 8 besser ab, als in der Online-Evaluation. *SymptomCheck* wurde lediglich von den Befragten der Online-Evaluation bezüglich Aussage 3 mit 96,9% und Aussage 7 mit 90,8% positiver bewertet als von den Patienten der Pilotstudie II (89,7% und 86,7%).

Tabelle 24: Spezifische Bewertungen zu *SymptomCheck* (Pilotstudie II)

	Aussage	Anteil (%)		
		positiv	neutral	negativ
1.	Ich konnte meine Symptome gut finden.	100	0	0
2.	Unklare Begriffe wurden mir gut erklärt.	93,3	6,7	0
3.	Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.	89,7	10,3	0
4.	Dieses Programm ist im Alltag sehr nützlich.	86,7	10	3,3
5.	Das Programm hat mir besser geholfen, als andere Informationsplattformen im Internet.	72,4	27,6	0
6.	Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	83,3	16,7	0
7.	Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel).	86,7	10	3,3
		<b>zu lang</b>	<b>genau richtig</b>	<b>zu kurz</b>
8.	Die Dauer für die Beantwortung der Fragen war...	3,3	93,3	3,3

#### 10.4 Überprüfung der Verdachtsgenerierung

Ebenso wie in der Pilotstudie I zuvor wurde untersucht ob die überarbeitete, endgültige Version von *SymptomCheck* die zutreffenden Diagnosen unter den Top 10 Verdachtsdiagnosen aufführen kann. Daher wurden die aktuellen Diagnosen der Entlassungsbriefe mit den Verdachtsdiagnosen von *SymptomCheck* verglichen. Die Auswertung erfolgte analog zur Pilotierung.

Von den 33 Patienten bedienten nur 30 das Programm eigenständig, weshalb nur deren Diagnosen in die Gesamtauswertung einfließen [siehe [Tabelle 25](#)].

Diese 64 Diagnosen sind in ihren medizinischen Fachgebieten zugeordnet. Den Hauptanteil der getesteten Diagnosen bilden dabei die kardiologischen Diagnosen mit 56,25 % [siehe [Tabelle 26](#)].

Tabelle 25: Diagnoseliste (Pilotstudie II)

<b>Diagnose</b>	<b>Anzahl (n)</b>
allgemein akutes Koronarsyndrom	4
stabile Angina Pectoris	1
Pericarditis constrictiva	1
bradykarde Herzrhythmusstörung	2
tachykarde Herzrhythmusstörung	6
Vorhofflattern	2
Vorhofflimmern	3
globale Herzinsuffizienz	8
Rechtsherzinsuffizienz	1
Linksherzinsuffizienz	3
Endokarditis	1
Aortenklappenstenose	1
akute Mitralklappeninsuffizienz	1
chronische Mitralklappeninsuffizienz	1
Mitralklappenstenose	1
Lungenödem	2
Lungenkarzinom	3
Diabetes Mellitus	3
hormonelle Ursache	1
akute Gastritis	1
Gastroösophageale Refluxkrankheit	1
akute Pankreatitis	2
chronische Pankreatitis	2
Medikamentennebenwirkung	1
Infektion	2
COPD	2
Synkope	2
Hauterkrankung	1
akutes Nierenversagen	1
Pneumonie	4
<b>Gesamt</b>	<b>64</b>

Tabelle 26: Diagnoseeinteilung nach medizinischem Fachgebiet (Pilotstudie II)

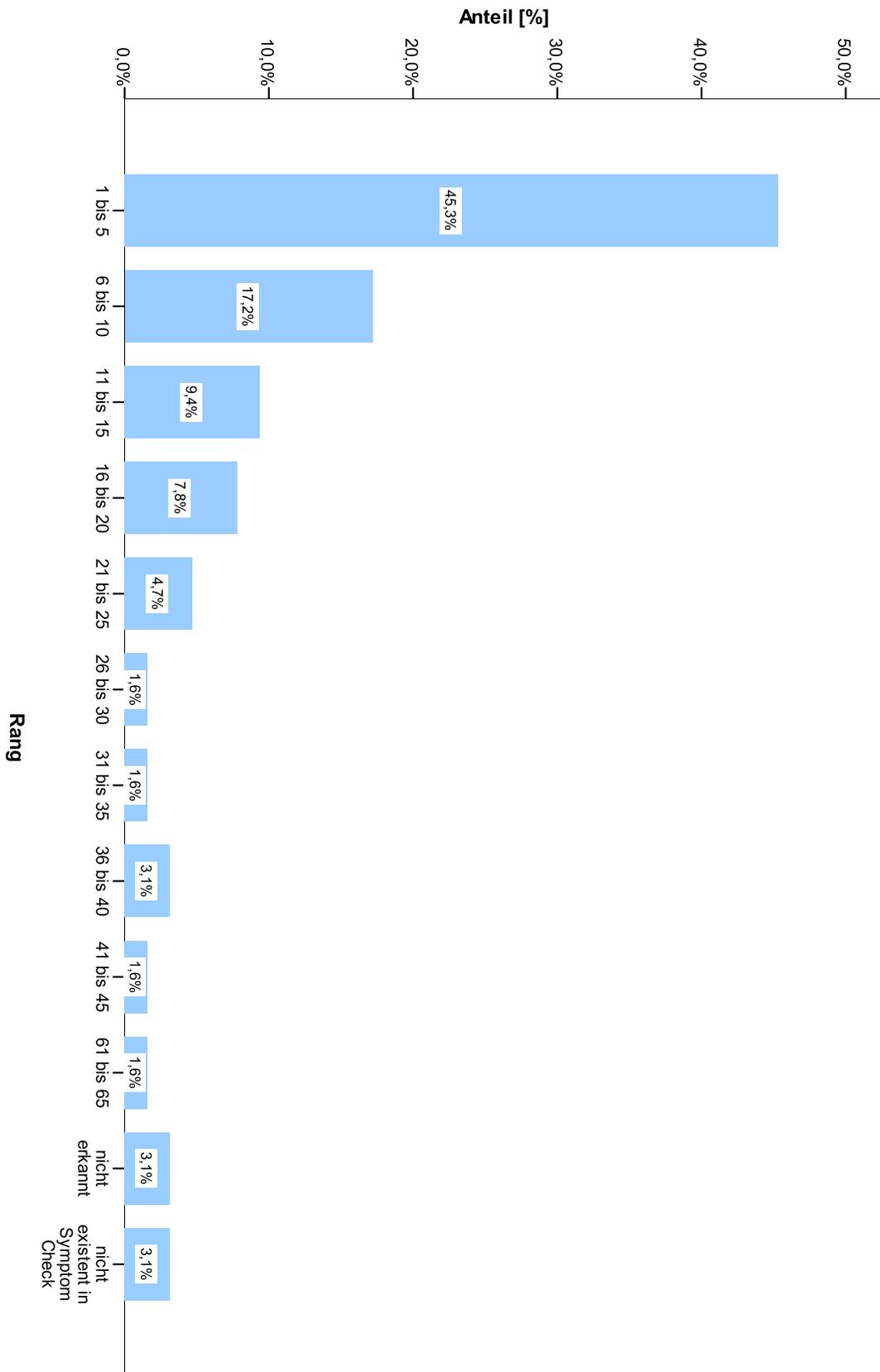
medizinisches Fachgebiet	Anzahl (n)
Kardiologie	36
Pulmologie	11
Endokrinologie	4
Gastroenterologie	6
Infektiologie	2
Dermatologie	1
Nephrologie	1
Sonstiges	3
Insgesamt	64

Im Rahmen der Studie war es nicht möglich, jede einzelne in *SymptomCheck* aufgeführte Diagnose einzeln anhand mehrerer Patienten mit der jeweiligen Diagnose zu testen. Dies hätte die Rahmenbedingungen der Testung weit überschritten. Doch lässt sich anhand des im Folgenden vorgestellten Rankings [siehe **Tabelle 27**; **Abbildung 29**], lässt sich ein guter Eindruck über die Verlässlichkeit der Generierung von Verdachtsdiagnosen erzeugen:

Tabelle 27: Platzierung der Diagnosen (Pilotstudie II)

Rang	Häufigkeit [n]	Anteil [%]	Kumulierte Prozenze
1 bis 5	29	45,3	45,3
6 bis 10	11	17,2	62,5
11 bis 15	6	9,4	71,9
16 bis 20	5	7,8	79,7
21 bis 25	3	4,7	84,4
26 bis 30	1	1,6	85,9
31 bis 35	1	1,6	87,5
36 bis 40	2	3,1	90,6
41 bis 45	1	1,6	92,2
61 bis 65	1	1,6	93,8
nicht erkannt	2	3,1	96,9
nicht existent in <i>SymptomCheck</i>	2	3,1	100,0
Gesamt	64	100,0	

Abbildung 29: Überprüfung der Verdachtsgenerierung: Überblick über die Platzierung der Diagnosen (Pilotstudie II)



Von den 64 Entlassdiagnosen erschienen 40 (62,5%) unter den 10 wahrscheinlichsten Diagnosen. Davon befanden sich 29 Diagnosen (45,3%) unter den Top 5 der Verdachtsdiagnoseliste. Betrachtet man die Top 20 der Verdachtsdiagnosen, so befinden sich insgesamt 51 Diagnosen (79,9%) darunter. Verglichen mit dem Ergebnis der Pilotstudie I, bei der nur 37,5% der Diagnosen dieses Intervalls angezeigt werden konnten, ist dies eine enorme Verbesserung. Bemerkenswert ist ebenfalls, dass der größte Anteil der Diagnosen unter den Top 5 in der Verdachtsdiagnoseliste erschienen. In der Pilotversion von *SymptomCheck* wurde der größte Anteil mit 33,3% zwischen Platz 21 und 40 in der Verdachtsdiagnoseliste angezeigt. Der Mittelwert bezüglich der Platzierung einer Diagnose liegt bei Rang 11, wohingegen der Median bei Rang 7 liegt [siehe **Tabelle 28**]. Eine Hypothese dieser klinischen Studie war, dass *SymptomCheck* die korrekte Diagnose unter seinen 10 für am wahrscheinlichsten zutreffend ausgewählten Diagnosen listen kann [siehe **5.1.3 Hypothesen**]. Betrachtet man den Median, so erfüllt *SymptomCheck* diese Hypothese. Orientiert man sich hingegen am Mittelwert, so gelingt es *SymptomCheck* ganz knapp nicht, korrekte Diagnose unter den Top 10 zu listen.

Hierbei sollte man sich jedoch am Median orientieren, da es – wie in **Abbildung 29** ersichtlich – nur vereinzelt Ausreißer nach unten, d.h. eine Platzierung schlechter als Platz 25, gibt.

**Tabelle 28: Mittelwert und Median der Diagnoseplatzierung (Pilotstudie II)**

	Anzahl (n)	Platzierung
Gültig	60	-
Fehlend <sup>1</sup>	4	-
Mittelwert	-	11,1
Median	-	6,5
Standardabweichung	-	12,1
Minimum	-	1,0
Maximum	-	63,0

*darunter fallen Diagnosen, welche von SymptomCheck nicht erkannt wurden oder nicht in SymptomCheck existent sind*

Folglich kann *SymptomCheck*, nach der Überarbeitung der Pilotversion, deren Ergebnisse in Kapitel **8.3 Überprüfung der Verdachtsgenerierung** dargestellt wurden, erfreuliche Ergebnisse erzielen.

## Diskussion

### 11 Besonderheiten des Patienten- und Probandenkollektivs

Für die Pilotstudien I und II zu *SymptomCheck* wurden Patienten rekrutiert, die zur Erfüllung der Einschlusskriterien volljährig sein und eine schriftliche Einverständniserklärung unterzeichnen mussten. Es wurden nur deutschsprachige Patienten befragt, da nur diese das Programm testen konnten. Abgesehen von der Altersgrenze gab es keine Ausschlusskriterien. Diese Kollektive entsprechen somit keinem typischen Studienkollektiv, allerdings war dies so beabsichtigt. Es sollte keine Vorselektion anhand von physiologischen Parametern erfolgen, da *SymptomCheck* sich in Zukunft an ein breites Benutzerspektrum richten soll. Für das Online-Kollektiv fiel die Vorselektion noch geringer aus, da jede Person mit einem Internetzugang an der Online-Evaluation teilnehmen konnte.

### 12 Kritikpunkte

#### **SymptomCheck vs. Arzt**

Beim Vergleich der Genauigkeit der Diagnosefindung eines Arztes mit der von *SymptomCheck* fiel auf, dass der Mediziner in allen Vergleichen dem Programm, wenn teilweise auch nur gering, überlegen war (siehe **Abbildung 6-10**). Dies ließe sich dadurch erklären, dass die klinische Erfahrung eines Arztes mit in die Beurteilung von Symptomkonstellationen einfließt. *SymptomCheck* ist nicht in der Lage, Erfahrung zu erlangen und kann sich nur auf reine medizinische Fakten berufen. *SymptomCheck* kann dagegen jedoch auch nicht von Vorurteilen getäuscht werden. Die Untersuchung dieser Aspekte müsste jedoch in künftigen Studien in einer noch größeren Anzahl an Fallbeispielen als auch an Ärzten erfolgen.

#### **Pilotstudie II**

Der Altersmittelwert des Kollektivs der Pilotstudie II liegt bei 58 Jahren. Dies ist damit zu begründen, dass sich als stationäres Patientenklientel vorwiegend Personen oberhalb des 40. Lebensjahres befanden. An einem größeren Patientenkollektiv mit einer breiteren Altersverteilung könnte eine Überprüfung

möglicher altersbezogener Unterschiede bezüglich der Genauigkeit der Verdachtsgenerierung erfolgen.

Antworten auf folgende Fragen wären zusätzlich interessant:

- *Geben jüngere Patienten andere Symptome bei der gleichen Diagnose an als ältere Patienten?*
- *Gibt es Diagnosen die sich bezogen auf ihr Ranking signifikant innerhalb der verschiedenen Altersgruppen unterscheiden?*
- *Wie groß ist der Einfluss von Komorbiditäten als Störfaktor?  
Wenn man hypothetisch davon ausgeht, dass jüngere Patienten weniger Komorbiditäten besitzen als ältere: Wird die gleiche Diagnose bei einem jungen Patienten als wahrscheinlicher eingestuft als bei einem älteren Patienten mit Komorbiditäten?*

Desweiteren konnten anhand der 33 Patienten auch nur ein sehr geringes Krankheitsspektrum abgefragt werden.

Diese klinische Querschnittstudie soll und kann nur einen ersten Eindruck von der Güte der entwickelten Bewertungsschemata [siehe [4.4 Entwicklung eines Bewertungssystems](#)] und deren Integration in eine stimmige Oberfläche [siehe [4.5 Entwicklung der Startseite](#)] vermitteln. Da *SymptomCheck* jedoch insgesamt eine Bandbreite von 178 Diagnosen aufweist, würde sich eine Anschlussstudie empfehlen, in der die einzelnen Diagnosen anhand eines großen Patientenkollektivs getestet würden. Hierbei können auch die einzelnen internistischen Fachbereiche auf signifikante Unterschiede hin getestet werden.

### **Online-Evaluation**

An der Online-Evaluation beteiligten sich 75 Personen. Hierbei kann eine Verzerrung der Ergebnisse jedoch nicht ausgeschlossen werden, da eine wiederholte Teilnahme an der Evaluation möglich war. Es wurde jedoch bei der Auswertung von keiner wiederholten Teilnahme ausgegangen. Auch gleichen die Ergebnisse der Fragebögen einander nicht.

Eine Antwort auf die Frage, ob *SymptomCheck* die Befürchtungen der Benutzer, von einer schlimmen Erkrankung betroffen zu sein, reduzieren konnte, wurde nicht gefunden. In der Online-Evaluation wurden die Probanden nur nach der Bearbeitung von *SymptomCheck* zu ihrer Befürchtung befragt [siehe [Tabelle 17](#)]. Auf eine zusätzliche Beantwortung vor der Testung wurde verzichtet, da ansonsten der zeitliche Gesamtrahmen der Befragung überschritten würde. Es

wurde befürchtet, dass dadurch eine geringere Anzahl an Testpersonen die Fragebögen vollständig beantworten würden. Es wäre nicht sinnvoll gewesen, zu untersuchen, ob bei den klinischen Patienten eine Verringerung dieser Befürchtung durch *SymptomCheck* resultiert, da die meisten Patienten wussten, dass sie krank waren.

Eine Stärke dieser Evaluation ist der hohe Beteiligungsanteil der 21- bis 30-Jährigen Personen. Da gerade diese Altersgruppe im Allgemeinen ein sehr aktiver Nutzer des Internets ist, geht man davon aus, dass auch ein großer Anteil davon dem künftigen Klientel von *SymptomCheck* entsprechen wird<sup>73</sup>. Somit sind die Ergebnisse der Online-Evaluation besonders für diese Altersgruppe repräsentativ. Auffallend ist auch, dass der Beteiligungsanteil mit 89,3% für Personen mit der allgemeinen oder fachgebundenen Hochschulreife sehr hoch ist. Zusammenfassend lässt sich sagen, dass die Ergebnisse der Online-Evaluation vor allem für die Probandengruppe der 21- bis 30-Jährigen Personen mit Fach-/Abitur aussagekräftig sind. Für eine Unterscheidung der Benutzerfreundlichkeit bezüglich der einzelnen Alters- und Bildungsgruppen ist eine langjährige, groß angelegte Online-Evaluation mit hoher Fallzahl denkbar.

## 13 Überprüfung der Hypothesen

Zu den in Kapitel **5.1.3 Hypothesen** genannten Hypothesen, lässt sich sagen, dass diese nur teilweise zutreffen. Hypothese 1, laut der *SymptomCheck* als benutzerfreundlich empfunden wird, trifft in allen durchgeführten Evaluationen zu. Jedoch trifft Hypothese 2, welche besagt, dass die korrekte Diagnose unter den 10 wahrscheinlichsten Diagnosen aufgelistet wird, nur teilweise zu. Ausgehend vom Median, erfüllt *SymptomCheck* diese Hypothese mit einer medianen Platzierung einer Diagnose auf Rang 7. Dient jedoch der Mittelwert zur Orientierung, so gelingt es *SymptomCheck* bei einer mittleren Platzierung einer Diagnose auf Rang 11 ganz knapp nicht, die korrekte Diagnose unter den Top 10 zu listen. Gemäß der dritten Annahme, kämen jüngere Personen und Personen mit einem höheren Bildungsgrad mit *SymptomCheck* besser zurecht als andere. Es lässt sich nach Auswertung dieser Studie lediglich sagen, dass *SymptomCheck* für dieses, im vorherigen Abschnitt dargestellte Klientel

benutzerfreundlich ist. Ob das Programm aber von dieser Personengruppe als bedienungsfreundlicher eingestuft wird als von anderen, bleibt offen.

## **14 Schwächen von *SymptomCheck***

### **Optisches Erscheinungsbild**

Die Startseite von *SymptomCheck* wurde für Bildschirmformate ab 15 Zoll entwickelt. Für kleinere Geräte und Smartphones ist diese Version bislang ungeeignet, da die Schrift und das Design sich nicht automatisch an das kleinere Format anpassen. Folglich kann es sein, dass sich die einzelnen Teile der Startseite – z. B. Lösungskasten und Symptomkasten – überschneiden. Dadurch sind nicht mehr alle Abschnitte für den Benutzer einsehbar. Um dies zu vermeiden, könnte *SymptomCheck* zukünftig als App zur Verfügung gestellt werden. Hierfür müsste die Startseite zweistufig gestaltet werden, sodass erst die übergeordneten Bereiche (z.B. Schmerzen oder Beschwerden im Bereich der Brust, der Lunge oder des Herzens) und im Anschluss die Symptome auszuwählen sind.

Desweiteren erhielten wir per E-Mail zusätzliches Feedback von Probanden, die sich größere auszuwählende Kreise für die Beantwortung der einzelnen Fragen wünschen. Auch wurden wir von ihnen darauf hingewiesen, dass es nicht immer auf den ersten Blick ersichtlich war, dass die Symptome direkt anklickbar sind. Hierfür wären eventuell markierbare Kästchen neben den einzelnen Symptomen in der Startseite zu integrieren. Um Verständnisprobleme zu vermeiden, könnte zusätzlich hinter jeden Symptomkasten ein „?“-Button eingefügt werden, mit dem erklärende Informationen zu dem jeweiligen Bereich und dessen Symptomen geliefert werden. Für den künftigen Gebrauch sollten diese Änderungen in einer überarbeiteten Version vorgenommen werden.

Nach Auswertung der klinischen Studie zeigte sich, dass 79,7% der Diagnosen unter den 20 wahrscheinlichsten Verdachtsdiagnosen erschienen. Insgesamt empfanden in der Online-Evaluation 77,8% der Befragten die Übersicht über mögliche Diagnosen als gut. Bei denjenigen, die befürchteten krank zu sein, waren dies sogar 83,3%. Diese Zahlen sind bereits sehr positiv, könnten jedoch noch optimiert werden.

Zur Verbesserung der Übersichtlichkeit könnte man die Liste an Vorschlägen lediglich auf das Aufzeigen der 20 wahrscheinlichsten Verdachtsdiagnosen reduzieren. Damit müsste man zwar das Risiko in Kauf nehmen, dass zu etwa 20% die aktuelle Diagnose nicht aufgeführt wird, jedoch würden sich die Benutzer dadurch in 80% der Fälle lediglich an den 20 dargestellten Diagnosen orientieren können. Dadurch würde sich das Risiko minimieren, dass die Personen durch sehr unwahrscheinliche, jedoch angezeigte Diagnosen verwirrt und verunsichert würden.

### **Fachgebietsgruppierung**

Eine hilfreiche Verbesserung wäre zudem die Gruppierung der Diagnosen nach medizinischen Fachgebieten mittels farblicher Text hervorhebung. Dadurch erhielten Patienten bei der Anhäufung von Diagnosen einer bestimmten Farbe den Hinweis, dass sie beispielsweise bei multiplen rot hinterlegten Diagnosen für den Fall einer ärztlichen Konsultation einen Kardiologen aufsuchen sollen. Die Patienten würden folglich richtungshinweisende Informationen erhalten.

### **Schweregradeinteilung**

Wie bereits in der Studie der Harvard Medical School erwähnt, bieten einige Selbstdiagnoseprogramme Ratschläge hinsichtlich der Dringlichkeit eines Arztbesuches<sup>55</sup>. Solch ein Zusatz ließe sich auch in *SymptomCheck* leicht umsetzen. Ein Vorschlag wäre die symbolische Kennzeichnung von folgenden Zuständen: „Notfall – sofort einen Arzt aufsuchen“ (rote Ampel), „Ärztlichen Rat innerhalb der nächsten Wochen einholen“ (gelbe Ampel) und „Selbstbehandlung; ärztlicher Rat nur bei anhaltender Symptomatik“ (grüne Ampel). Dadurch könnten die Personen visuell leicht zur Dringlichkeit einer ärztlichen Konsultation beraten werden. Dieser Aspekt könnte auch zur Verbesserung der Nützlichkeit und Übersichtlichkeit des Programms beitragen.

### **Fachärztliche Kooperation**

Die gesamte Wissensbasis wurde allein mit dem medizinischen Wissen von Fachbüchern und den medizinischen Kenntnissen und logischen Überlegungen

eines Doktoranden entwickelt<sup>59,62-68</sup>. Wir gehen davon aus, dass die Sensitivität von *SymptomCheck* durch die Zusammenarbeit mit Fachärzten der jeweiligen internistischen und außerinternistischen Wissensgebiete noch verbessert würde. Diese könnten neben den aus der Literatur extrahierbaren Informationen vor allem deren klinische Erfahrung in die Wissensbasis einfließen lassen.

Desweiteren wäre mit deren Hilfe das Krankheitsspektrum auch stark erweiterbar. Ein Vorteil der Bewertungsschemata ist, dass diese auf multiple Diagnosen anwendbar sind.

### **Bewertungsschema**

Beim Bewertungsschema von *SymptomCheck* wurde auf der Startseite bewusst nicht zwischen verneinten und nicht angegebenen Symptomen unterschieden, da der Aufwand für die explizite Angabe von Verneinungen zu groß ist und implizite Verneinungen nicht zuverlässig genug sind. Es zeigte sich jedoch durch die Evaluationen, dass eine Unterscheidung in einigen Fällen sinnvoll ist. Unsere Vermutung ist, dass das Programm spezifischer werde, wenn für jeden übergeordneten Bereich (z.B. „Schmerzen“) auf der Startseite eine weitere Antwortalternative (z.B. „Keine Probleme“) als explizite Verneinung aufgeführt werde. Zur Anpassung des Bewertungsschemas zur Auswertung von solchen Negationen wäre, wie oben erwähnt, eine fachärztliche Kooperation äußerst sinnvoll.

### **Technische Fehler**

Da die vollständige Wissensbasis von *SymptomCheck*, manuell von einem Doktoranden entwickelt wurde, können Unstimmigkeiten oder Fehler innerhalb der Bewertung von Diagnosen nicht ausgeschlossen werden. Zwar wurde bei allen Änderungen der Bewertungsschemata und Verknüpfung der Diagnosen mit spezifischeren Symptomabfragen stets auf eine korrekte Datenverarbeitung geachtet, jedoch kann es trotzdem zu Bearbeitungsfehlern gekommen sein.

### **Störfaktoren**

Wichtig für die Beurteilung von Selbstdiagnoseprogrammen wie *SymptomCheck* ist der Einfluss von Störfaktoren auf das Gesamtergebnis. Optische Störfaktoren

wurden durch schlichtes Design versucht zu minimieren. Der Fokus sollte nur auf relevante Aspekte des Programms gelenkt werden. Die Erklärungen an den Benutzer sind auf das nötigste reduziert, ebenso wurden auf Animationen gänzlich verzichtet.

Desweiteren können jedoch Missverständnisse bei der Auswahl von Symptomen und der Beantwortung der Fragen nicht ausgeschlossen werden. Auch Komorbidität ist ein Störfaktor für die Genauigkeit der Verdachtsgenerierung. Der Einfluss von Komorbidität auf das Gesamtergebnis lässt sich jedoch schwierig verallgemeinernd beschreiben und würde eine zusätzliche Studie benötigen, welche sich speziell diesem Thema annimmt.

Während der Online-Studie könnten Probanden die Evaluation mehrmals durchgeführt und dadurch zu einer verzerrten Auswertung beigetragen haben. Daneben könnten die Patienten der Pilot- und der eigentlichen Studie in der Medizinischen Klinik I von dem persönlichen Kontakt durch den sie über die Studie informierenden Doktoranden hinsichtlich ihrer Gesamtbewertung der Benutzerfreundlichkeit beeinflusst worden sein. Dieser Abschnitt sollte die relevantesten Störfaktoren aufzeigen, jedoch kann dadurch nicht ausgeschlossen werden, dass es noch weitere Confounder geben kann.

## **15 Stärken von *SymptomCheck***

Neben den eben aufgezeigten Mängeln und Verbesserungsanregungen bietet *SymptomCheck* jedoch einige innovative Aspekte, welche dieses Programm oftmals positiv von bereits bestehenden Selbstdiagnoseprogrammen unterscheidet.

### **Keinerlei kommerzielle Absichten**

Anders als viele der im Kapitel „Apps zur Selbstdiagnose“ beschriebenen Programme, ist *SymptomCheck* frei von jeglicher Werbung. Es existieren weder Links auf kommerzielle Seiten noch wird dem Benutzer zum Aufsuchen namentlich genannter Ärzte geraten. Zudem wird auch nicht – anders als bei vielen anderen Programmen – zu Seiten anderer Thematiken, (z. B. Unterhaltungsseiten mit Gesundheitstipps oder Sudoku) verlinkt. *SymptomCheck* ist für den

Benutzer kostenlos. Die Benutzung des Programms ist an keinerlei Bedingungen, wie beispielsweise das Abonnieren eines Newsletters, geknüpft.

### **Verdachtsüberprüfungskomponente**

Eine große Neuheit unter der bisher verfügbaren Selbstdiagnose Software ist die Verdachtsüberprüfungskomponente von *SymptomCheck*. Damit können die Nutzer nach Generierung einer Liste aus Verdachtsdiagnosen eine mögliche Diagnose auswählen. Im Anschluss daran, wird der Benutzer explizit nach Beschwerden gefragt, welche für diese Diagnose zutreffen müssten. Somit wird die Wahrscheinlichkeit für das Zutreffen einer Diagnose berechnet. Diese Komponente wurde jedoch von einer anderen Doktorandin entwickelt und in dieser Dissertation nicht behandelt<sup>74</sup>. Eine kombinierte Evaluation von Verdachtsgenerierung und Verdachtsüberprüfung ist sehr attraktiv und künftig durchzuführen.

### **Nur temporäre Datensammlung im Rahmen der Evaluation**

Ein wichtiger Punkt der *SymptomCheck* von den anderen Selbstdiagnoseprogrammen unterscheidet ist der Datenschutz. Während viele Apps, wie schon beschrieben, Informationen zum aktuellen Standort während der Benutzung oder allgemein Informationen zum Nutzerverhalten speichern und teilweise sogar an Dritte weitergeben, ist dies bei *SymptomCheck* nicht der Fall<sup>13,52</sup>. Die einzige Datenspeicherung erfolgte im Rahmen der Evaluation. Sobald die Evaluation abgeschlossen war, wurden keine Daten mehr gesammelt und sollen es auch künftig nicht werden.

### **Alle Symptome auf 1 Blick**

Ein Ziel von uns war es, alle Leitsymptome übersichtlich auf einer Startseite darzustellen. Damit erhofften wir uns, dass dieser Aspekt unter anderem eine gute Benutzerfreundlichkeit gewährleistet. Bei anderen Programmen, z. B. auf der Webseite „Onmeda“<sup>75</sup>, kann man zwar oft ein bestimmtes Körperteil direkt anklicken, muss dann aber aus einer langen Liste aus Beschwerden die richtige heraussuchen. Dies stellt sich als zeitaufwendig heraus und es werden nur

schwer alle zutreffenden Beschwerden gefunden. Die Startseite von *SymptomCheck* ist in Oberkategorien aufgegliedert, die jeweils eine kleine Gruppe an Symptomen enthalten. Nur in wenigen Ausnahmefällen sind Symptome zusammengefasst und müssen erst über beispielsweise „Sonstige Beschwerden“ aufgerufen werden.

Die Übersichtlichkeit und der logische Aufbau der Startseite wurden ebenfalls evaluiert. Hierbei gaben in der Online-Evaluation 73,8% an, ihre Symptome gut finden zu können. 1,5 % verhielten sich zu der Aussage neutral und 24,6 % hatten Schwierigkeiten bei der Symptomfindung [siehe [Tabelle 19](#), [Tabelle 30](#)]. Betrachtet man jedoch nur diejenigen, welche befürchteten krank zu sein und demzufolge in Zukunft exakt das Benutzerklientel für *SymptomCheck* darstellen, so waren es bei dieser Gruppe sogar 84,2%, die keine Schwierigkeiten beim Finden ihrer Symptome hatten. Hierbei fanden lediglich 15,8% die Symptomfindung problematisch [siehe [Tabelle 20](#)].

### **Benutzer entscheidet über Tiefe der Befragung**

Ein wichtiger Vorteil von *SymptomCheck* ist, dass der Benutzer frei darüber entscheiden kann, wie genau er zu seinen Symptomen befragt werden möchte. Das Programm benötigt keine Mindestanzahl an Antworten, um eine Verdachtsdiagnoseliste zu generieren. Es kann den Benutzer genauer zu den Symptomen befragen, jedoch müssen die Fragen nicht zwingend beantwortet werden. Sie ermöglichen lediglich ein genaueres Ergebnis. Dementsprechend sind die Verdachtsdiagnosen bei nur sehr oberflächlicher Beantwortung weniger wahrscheinlich. Dennoch ist es für den Benutzer von Vorteil, dass er nicht einen starren Fragenkatalog beantworten muss, um eine Antwort zu erhalten.

Um diesen Aspekt zu validieren, wurden die Probanden nach der Testung zur Dauer der Bearbeitung befragt. Wie im Abschnitt [Ergebnisse](#) bereits beschrieben wurde, hielten erfreulicherweise 81,4% [siehe [Tabelle 19](#), [Tabelle 30](#)] der Teilnehmer an der Online-Evaluation die Beantwortungsdauer für „genau richtig“. Lediglich 14,3% hielten die Dauer der Befragung für „zu kurz“ und 4,3% für „zu lang“. In der Pilotstudie II, bei welcher die Patienten das identische Programm bewerteten, wurde die Beantwortungsdauer sogar mit 93,3% für „genau richtig“

eingestuft. Lediglich jeweils 3,3% empfanden die Dauer als „zu lang“ oder „zu kurz“ [siehe [Tabelle 24](#)].

## Transparenz

Es wird - anders als bei den bereits vorhandenen Selbstdiagnoseprogrammen oder Apps - viel mehr auf Transparenz geachtet. Dies wirkt sich auf verschiedene im Folgenden erläuterte Aspekte aus:

Wie im Kapitel [3.2 Systeme für das Patientenlientel](#) bereits angesprochen, konnten bei vielen Programmen die konkreten Entwickler oftmals nicht zurückverfolgt werden. Es wurde dafür manchmal auf große Namen wie die Harvard Medical School verwiesen.<sup>13,51</sup> Dies ist in unserem Programm anders. Auf der Website von *SymptomCheck* ist vermerkt, dass dies ein von Doktoranden und Mitarbeitern des Lehrstuhls für Informatik an der Universität Würzburg entwickeltes Programm ist. Bei Fragen oder Anmerkungen können sich die Benutzer per Mail jederzeit mit uns in Verbindung setzen.

Ein genereller Vorteil solcher Programme ist, dass deren Entscheidung - im Gegensatz zu der ärztlichen von Vorurteilen beeinflussbaren Entscheidung – auf neutralen mathematischen Gesetzmäßigkeiten beruhen<sup>13</sup>. Auch hierbei wurde stark auf Transparenz geachtet.

Denn ein weiterer positiv hervorzuhebender Aspekt von *SymptomCheck* ist, dass man die Veränderung der Liste an Verdachtsdiagnosen aktiv mitverfolgen kann. Somit sieht der Benutzer, ob eine Verdachtsdiagnose nach Beantwortung einer neuen Frage im Ranking weiter nach oben oder unten rückt oder sogar völlig aus der Liste der 20 wahrscheinlichsten Diagnosen herausfällt. Außerdem kann der Benutzer nach Beantwortung der Anamnesefragen die als wahrscheinlich angezeigten Verdachtsdiagnosen einzeln auswählen. Daraufhin öffnet sich ein Fenster, in welchem genau aufgelistet ist, welche Antwort wie stark für oder gegen die jeweilige Diagnose spricht.

In [Abbildung 32](#) kann das transparente Bewertungssystem anhand des Beispiels der COPD nachvollzogen werden: Wählt der Nutzer „Husten“ als eines der möglichen Symptome aus, so erhält dieses Symptom bezüglich der

chronischen Bronchitis eine positive Wertigkeit von 2. Wird auf dieses sehr oberflächliche Leitsymptom dann genauer nachgefragt, und wird die Frage „Wie würden Sie den Husten beschreiben?“ mit „produktiv“ geantwortet, dann erhält dieses spezifischere Symptom eine zusätzliche Wertung von 10. Hätte man auf die Frage „Seit wann haben Sie Husten“ mit „seit kurzem“ geantwortet, dann würde die Wahrscheinlichkeit für eine COPD um 10 Werteinheiten sinken.

Auch kann die Entwicklung der Wissensbasis anhand der Dissertation nachvollzogen werden. Die genaue Erläuterung des Aufbaus des Programms mitsamt seiner Symptom- und Diagnoseanzahl, sowie das Bewertungssystem ist im Methodenkapitel erläutert und frei zugänglich.

## **16 Anregungen für die Zukunft**

Ein wichtiges Anwendungsgebiet von *SymptomCheck* neben seines Online-Klientels könnte auch der Einsatz bei fremdsprachigen Patienten darstellen. Im Klinikalltag mangelt es oftmals an Dolmetschern für diverse Sprachen. Vor allem bezüglich der aktuellen Flüchtlingsthematik wird dieses Defizit besonders deutlich. So stellt sich eine Anamneseerhebung aufgrund der Sprachbarrieren bereits als oftmals schwierig heraus<sup>76,77</sup>. Für eine zufriedenstellende ärztliche Versorgung ist jedoch auf eine fundierte Anamnese Wertzulegen. Hierbei könnte *SymptomCheck* als ubiquitär verfügbares Hilfsmittel zum Einsatz kommen. Man müsste die Wissensbasis lediglich in die verschiedenen Sprachen übersetzen lassen. Die Patienten könnten somit bereits vor dem Arztkontakt von *SymptomCheck* zu deren Beschwerden befragt werden. Hierbei kann es selbstverständlich vorkommen, dass nicht alle Symptome vom Programm erfasst werden. Dennoch könnte es vor allem in medizinischen Versorgungszentren ohne Dolmetscher von großem Vorteil sein und eine Entlastung sowohl für Ärzte als auch für die Flüchtlingshelfer vor Ort darstellen.

## Zusammenfassung

Das Thema dieser Dissertation lautete „*SymptomCheck* - ein Programm zur symptomorientierten Ableitung eines Diagnosevorschlags“. Der Rahmen dieser Dissertation umfasste die ausführliche, eigenständige Konzeption von *SymptomCheck* (frei zugänglich unter dem Link: <https://go.uni-wue.de/symptomcheck>), dessen Untersuchung auf Benutzerfreundlichkeit und eine Überprüfung auf allgemeine Validität bezüglich der Verdachtsgenerierung. Hierbei wurde die Entwicklung der Wissensbasis mit dem dafür eigens konzipierten Bewertungsschema entwickelt. Desweiteren wurde eine alltagstaugliche Benutzeroberfläche generiert, die ein breites Personenspektrum adressiert. In einer Pilotstudie I wurde das Programm auf die oben erwähnten Aspekte hin untersucht und dessen Mängel überarbeitet. Die endgültige Version von *SymptomCheck* wurde sodann sowohl in einer Online-Evaluation wie auch einer weiteren klinischen Anwendung getestet (Pilotstudie II). Während bei der Online-Evaluation der Fokus auf der Untersuchung der Benutzungsqualität lag, wurde in der Pilotstudie II mit stationären Patienten vor allem die Präzision der Verdachtsgenerierung untersucht.

Verglichen mit den in Kapitel **3.2.4 Evaluation von Selbstdiagnoseprogrammen** vorgestellten Ergebnissen einer Studie der Harvard Medical School, war *SymptomCheck* demnach bezüglich der Stärke der Verdachtsgenerierung anderen Programmen durchaus ebenbürtig. Ein direkter Vergleich ist jedoch schwierig, da die Programme der amerikanischen Studie nur anhand von Fallbeispielen, nicht jedoch an echten Patienten getestet wurden. *SymptomCheck* sticht jedoch aufgrund einiger oben genannter Vorteile aus den anderen online verfügbaren Selbstdiagnoseprogrammen hervor. In puncto Benutzerfreundlichkeit schnitt *SymptomCheck* sowohl in der Online-Evaluation, als auch in der Pilotstudie II zufriedenstellend ab.

Es müssten in Zukunft jedoch noch Studien mit einem größeren Patientenkollektiv und einer breiteren Altersverteilung erfolgen, um altersbezogene Unterschiede bezüglich der Genauigkeit der Verdachtsgenerierung und der Benutzerfreundlichkeit zu eruieren. Desweiteren sind einzelne internistische Fachbereiche auf signifikante Unterschiede und der Einfluss von

Störfaktoren an einem größeren Kollektiv zu untersuchen. Zur Verbesserung der Sensitivität von *SymptomCheck* ist fachärztliche Kooperation ratsam.

Die vorliegende Arbeit ist demnach vermutlich die erste, die ein selbstentwickeltes Online-Programm im Rahmen einer Online-Evaluation an Probanden und im Rahmen einer klinischen Studie an echten Patienten testete. *SymptomCheck* ist somit der erste vielversprechende Schritt, ein für den Alltagsgebrauch entwickeltes, breit gefächertes Selbstdiagnoseprogramm klinisch zu validieren.

## Literaturverzeichnis

1. Internet als Medizinratgeber; Panikmache und Fehlinformation: So gefährlich ist "Dr. Google". FOCUS Online, 2014.  
[http://www.focus.de/gesundheit/gesundleben/vorsorge/risiko/internet-als-medizinratgeber-schwerpunkt-parallele-gesundheitswelt-diagnose-von-dr-google-und-apps-zur-selbstoptimierung\\_id\\_4221893.html](http://www.focus.de/gesundheit/gesundleben/vorsorge/risiko/internet-als-medizinratgeber-schwerpunkt-parallele-gesundheitswelt-diagnose-von-dr-google-und-apps-zur-selbstoptimierung_id_4221893.html) (letzter Zugriff 02. Oktober 2017).
2. Singh K, Brown RJ. From headache to tumour: An examination of health anxiety, health-related Internet use and 'query escalation'. *J Health Psychol* 2016;21(9):2008-20.
3. Cyberchondrie; Krank durch Doktor Google. FOCUS Online, 2008.  
[http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-krank-durch-doktor-google\\_aid\\_268018.html](http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-krank-durch-doktor-google_aid_268018.html) (letzter Zugriff 02. Oktober 2017).
4. Muse K, McManus F, Leung C, Meghreblian B, Williams JMG. Cyberchondriasis: Fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. *J Anxiety Disord* 2012;26(1):189-96.
5. Cyberchondrie; Süchtig nach Symptomsuche. FOCUS Online, 2008.  
[http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-suechtig-nach-symptomsuche\\_aid\\_268019.html](http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-suechtig-nach-symptomsuche_aid_268019.html) (letzter Zugriff 02. Oktober 2017).
6. Puppe B, Riecker G. CardioConsult: Kardiologisches Second-Opinion-Programm. Wiesbaden: Ullstein-Medical; 1998 (mittlerweile vergriffen, aber gebraucht noch bei Amazon erhältlich Stand Oktober 2017).
7. Jutel A. Sociology of diagnosis: a preliminary review. *Sociol Health Illn* 2009;31(2):278-99.
8. Blaxter M. Diagnosis as category and process: The case of alcoholism. *Soc Sci Med* 1978;12:9-17.
9. Dahmer J. Anamnese und Befund: die symptom-orientierte Patientenuntersuchung als Grundlage klinischer Diagnostik; ein interaktives Taschenbuch für Studium und Praxis. 10. Auflage. Stuttgart: Thieme; 2006:1-5.
10. Goyder C, McPherson A, Glasziou P. Diagnosis in general practice. *Self diagnosis. BMJ* 2009;339:b4418.
11. Gupta K, Hooton TM, Roberts PL, Stamm WE. Patient-initiated treatment of uncomplicated recurrent urinary tract infections in young women. *Ann Intern Med* 2001;135(1):9-16.

12. Fox S, Duggan M. Health online 2013. Washington, DC: Pew Internet & American Life Project 2013. <http://www.pewinternet.org/2013/01/15/health-online-2013/> (letzter Zugriff 02. Oktober 2017).
13. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Soc Sci Med* 2015;133:128-35.
14. Puppe F. Wissensbasierte Systeme in der Medizin. *Künstliche Intelligenz* 2008;22(1):53-8.
15. Machine learning. Wikipedia. [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning) (letzter Zugriff 09. November 2017).
16. k-nearest neighbors algorithm. Wikipedia. [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm) (letzter Zugriff 09. November 2017).
17. Bayes' theorem. Wikipedia. [https://en.wikipedia.org/wiki/Bayes%27\\_theorem](https://en.wikipedia.org/wiki/Bayes%27_theorem) (letzter Zugriff 09. November 2017).
18. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. 4. Auflage: Morgan Kaufmann Publishers Inc.; 2016:7-9.
19. Berner ES, La Lande TJ. Overview of clinical decision support systems. In: Berner ES, ed. *Clinical Decision Support Systems: Theory and Practice*. 2. Auflage. New York: Springer Science & Business Media; 2007:3-22.
20. Spreckelsen C, Spitzer K. *Wissensbasen und Expertensysteme in der Medizin*. Wiesbaden: Vieweg+Teubner; 2008:11.
21. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. *Br Med J* 1972;2(5804):9-13.
22. Adams ID, Chan M, Clifford PC, et al. Computer aided diagnosis of acute abdominal pain: a multicentre study. *Br Med J (Clin Res Ed)* 1986;293(6550):800-4.
23. Musen MA. Stanford Medical Informatics: uncommon research, common goals. *MD Comput* 1999;16(1):47-8, 50.
24. Shortliffe EH. *Computer-based medical consultations, MYCIN*. New York: Elsevier; 1976:1-61, 233-41.
25. Schreiber S, Schäfer H, Junker R, Bruhn HD. *LaborMedizin. Indikationen, Methodik und Laborwerte. Pathophysiologie und Klinik*. Stuttgart: Schattauer; 2011:280-1.
26. Prävention von Infektionen, die von Gefäßkathetern ausgehen. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2017;60(2):216-30.
27. Suttorp N, Berg T. *Infektionskrankheiten: verstehen, erkennen, behandeln*. Stuttgart: Thieme; 2004:37-8.

28. Beierle C, Kern-Isberner G. Methoden wissensbasierter Systeme: Grundlagen, Algorithmen, Anwendungen. Wiesbaden: Springer; 2014:16.
29. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994;330(25):1792-6.
30. DXplain. Boston: Massachusetts General Hospital; 1992.
31. Iliad. Salt Lake City: Applied Informatics; 1992.
32. Meditel, Computer Assisted Diagnosis. Devon, PA: Meditel, Inc.; 1991.
33. QMR (Quick Medical Reference). Pittsburgh: Camdat; 1992.
34. Ramnarayan P, Tomlinson A, Rao A, Coren M, Winrow A, Britto J. ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation. Arch Dis Child 2003;88(5):408-13.
35. Ramnarayan P, Winrow A, Coren M, et al. Diagnostic omission errors in acute paediatric practice: impact of a reminder system on decision-making. BMC Med Inform Decis Mak 2006;6:37.
36. Kimura M, Sakamoto M, Adachi T, Sagara H. Diagnosis of febrile illnesses in returned travelers using the PC software GIDEON. Travel Med Infect Dis 2005;3(3):157-60.
37. Edberg SC. Global Infectious Diseases and Epidemiology Network (GIDEON): A World Wide Web-Based Program for Diagnosis and Informatics in Infectious Diseases. Clin Infect Dis 2005;40(1):123-6.
38. Ramnarayan P, Kapoor RR, Coren M, et al. Measuring the Impact of Diagnostic Decision Support on the Quality of Clinical Decision Making: Development of a Reliable and Valid Composite Score. J Am Med Inform Assoc 2003;10(6):563-72.
39. Elstein AS, Friedman CP, Wolf FM, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. J Am Med Inform Assoc 1996;3(6):422-8.
40. Gesundheits-Apps in der Diskussion. ZDF heute, 2015. <http://www.zdf.de/ZDFmediathek/beitrag/video/2368914/Gesundheits-Apps-in-der-Diskussion#/beitrag/video/2368914/Gesundheits-Apps-in-der-Diskussion> (letzter Zugriff 28. November 2015).
41. Smartphones und Tablets: Worauf man bei Medizin-Apps achten sollte. SPIEGEL Online, 2015. <http://www.spiegel.de/gesundheit/diagnose/medizin-apps-smartphone-ersetzt-nicht-den-arzt-a-1026323.html> (letzter Zugriff 02. Oktober 2017).
42. Number of apps available in leading app stores as of July 2015. Statista, 2015. <http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/> (letzter Zugriff 28. Oktober 2015).

43. Andreassen HK, Trondsen M. The empowered patient and the sociologist. *Soc Theory Health* 2010;8(3):280-7.
44. Bury M, Taylor D. Towards a Theory of Care Transition: From Medical Dominance to Managed Consumerism. *Soc Theory Health* 2008;6(3):201-19.
45. WebMD. WebMD LLC. <https://www.webmd.com/webmdapp3> (letzter Zugriff 17. Oktober 2017).
46. iTriage. iTriage LLC. <https://itunes.apple.com/de/app/itriage-health-doctor-and-symptoms-search/id304696939?mt=8> (letzter Zugriff 19. Oktober 2017).
47. Jutel A, Lupton D. Digitizing diagnosis: a review of mobile applications in the diagnostic process. *Diagnosis* 2015;2(2):89-96.
48. Isabel Symptom Checker. Isabel Healthcare, 2015. <http://symptomchecker.isabelhealthcare.com/> (letzter Zugriff 19. November 2015).
49. Dr. Moms - Treatment Guide. Mobogenie, 2015. <http://www.mobogenie.com/download-dr-moms-treatment-guide-7576.html> (letzter Zugriff 02. Oktober 2017).
50. Hamilton AD, Brady RR. Medical professional involvement in smartphone 'apps' in dermatology. *Br J Dermatol* 2012;167(1):220-1.
51. Virtual Nurse. The Best Android Symptom Checker. Senstore. <http://www.senstore.com/> (letzter Zugriff 21. November 2015).
52. Gesundheits-Apps als Arzt-Ersatz. ARD, 2015. <http://www.ardmediathek.de/tv/Mittagsmagazin/Gesundheits-Apps-als-Arzt-Ersatz/Das-Erste/Video?documentId=27256916&bcastId=314636> (letzter Zugriff 28. November 2015).
53. Your Rapid Diagnosis. WWW Machealth Pty Ltd, 2011. <http://www.yourrapiddiagnosis.com/WhatisRapidDiagnosis/tabid/621/Default.aspx> (letzter Zugriff 17. November 2015).
54. Symptify. 2013. <https://symptify.com> (letzter Zugriff 17. November 2015).
55. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
56. Symptom-Checker. Mehr Zufall als Diagnose. *Süddeutsche Zeitung*, 2015. <http://www.sueddeutsche.de/gesundheit/symptom-checker-mehr-zufall-als-diagnose-1.2556320> (letzter Zugriff 20. November 2015).
57. MEDoctor. MEDoctor Systems, Inc. [www.medoctor.com](http://www.medoctor.com) (letzter Zugriff 19. Oktober 2017).
58. DocResponse. DocResponse, Inc. <https://www.docresponse.com/> (letzter Zugriff 19. Oktober 2017).

59. Herold G. Innere Medizin 2013 : eine vorlesungsorientierte Darstellung ; unter Berücksichtigung des Gegenstandskataloges für die Ärztliche Prüfung ; mit ICD 10-Schlüssel im Text und Stichwortverzeichnis. Köln: Herold; 2013.
60. Thomas G. Basiswissen Urologie. 5. Auflage. Heidelberg: Springer; 2011:50.
61. Sarah S. Last Minute Innere Medizin. 1. Auflage. München: Elsevier, Urban & Fischer; 2012:160.
62. Battegay E, Siegenthaler W, Aeschlimann AG, et al. Siegenthalers Differenzialdiagnose: Innere Krankheiten - vom Symptom zur Diagnose. 20. Auflage. Stuttgart: Thieme; 2012.
63. Baenkler HW, Goldschmidt H, Hahn JM, et al. Kurzlehrbuch Innere Medizin. 2. Auflage. Stuttgart, New York: Thieme; 2010.
64. Horny J. Differentialdiagnostisches Kompendium: Symptome, Befunde, Krankheiten, Laborwerte, EKG-Befunde, Röntgenbefunde. 4. Auflage. Basel u.a.: Karger; 1995.
65. Jachmann-Jahn U, Breusch S. Klinikleitfaden Leitsymptome, Differenzialdiagnosen. München, Jena: Elsevier, Urban & Fischer; 2009.
66. Furger P. Innere Medizin quick. 3. Auflage. Stuttgart, New York: Thieme; 2011.
67. Strauß VE. Selbstdiagnose. Das große Handbuch der Gesundheit. Das Standardwerk zur Erhaltung der Gesundheit für die ganze Familie. München: Weltbild; 1994.
68. Herold G. Innere Medizin 2014 : eine vorlesungsorientierte Darstellung ; unter Berücksichtigung des Gegenstandskataloges für die Ärztliche Prüfung ; mit ICD 10-Schlüssel im Text und Stichwortverzeichnis. Köln: Herold; 2014.
69. Laugwitz B, Schrepp M, Held T. Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: Heinecke AM, Paul H, eds. Mensch und Computer 2006: Mensch und Computer im StrukturWandel. München: Oldenbourg Verlag; 2006:125-34.
70. Schrepp M. User Experience Questionnaire Handbook. All you need to know to apply the UEQ successfully in your projects. 2015. [www.ueq-online.org](http://www.ueq-online.org) (letzter Zugriff 29. Oktober 2015).
71. Hellmich B. Fallbuch Innere Medizin: 150 Fälle aktiv bearbeiten. 4. Auflage. Stuttgart: Thieme; 2012.
72. Gaus W, Muche R. Medizinische Statistik: Angewandte Biometrie für Ärzte und Gesundheitsberufe. Stuttgart: Schattauer; 2014:281-91.
73. Anteil der Internetnutzer nach Altersgruppen in Deutschland in den Jahren 2014 und 2015. Statista, 2015. <http://de.statista.com/statistik/daten/studie/3101/umfrage/internetnutzung-in-deutschland-nach-altersgruppen/> (letzter Zugriff 17. November 2015).

74. Azar I. Konzeption und Evaluation eines webbasierten Patienteninformationsprogramms zur Überprüfung internistischer Verdachtsdiagnosen (medizinische Dissertation der Universität Würzburg; eingereicht).
75. Symptom-Check. Beschwerden und Symptome lokalisieren. Onmeda, 2015. <http://www.onmeda.de/symptome/anatomicus.html> (letzter Zugriff 11. November 2015).
76. Heer IM. Integrierende Versorgung von Flüchtlingen. Notfall Rettungsmed 2016;19(5):332-8.
77. Szajna A, Ward J. Access to Health Care by Refugees: A Dimensional Analysis. Nurs Forum 2015;50(2):83-9.

## Tabellenverzeichnis

Tabelle 1: Anzahl der den Fachgebieten zugeordneten Diagnosen .....	19
Tabelle 2: Endgültiges Bewertungsschema .....	24
Tabelle 3: Bewertungsschema "Bauchschmerzen" .....	25
Tabelle 4: Ursprüngliches Bewertungsschema .....	28
Tabelle 5: Erklärung der doppelten Negation anhand des Beispiels "Übelkeit, Erbrechen bei Appendizitis" .....	29
Tabelle 6: Erklärung der Problematik des ursprünglichen Bewertungsschemas.....	30
Tabelle 7: Wertebereiche des UEQ-Fragebogens.....	41
Tabelle 8: Fiktives Beispiel zur Bewertung der 6 Skalen des UEQ-Fragebogens.....	42
Tabelle 9: Bewertung der 6 Skalen des UEQ-Fragebogens (Pilotstudie I) .....	54
Tabelle 10: Geschlechtsabhängige Bewertung der 6 Skalen des UEQ-Fragebogens (Pilotstudie I).....	55
Tabelle 11: Diagnoseliste (Pilotstudie I) .....	56
Tabelle 12: Platzierung der Diagnosen (Pilotstudie I).....	58
Tabelle 13: Medien zur Information über Symptome und Krankheiten (Online- Evaluation) .....	61
Tabelle 14: Internetseiten als medizinische Informationsquelle (Online-Evaluation)....	62
Tabelle 15: Bewertung der 6 Skalen (Online-Evaluation) .....	64
Tabelle 16: Geschlechtsabhängige Bewertung der 6 Skalen (Online-Evaluation) .....	64
Tabelle 17: Bewertung der Aussage "Ich befürchte krank zu sein." (Online- Evaluation) .....	65
Tabelle 18: Unterschiede bezüglich der Bewertung der 6 Skalen hinsichtlich der unterschiedlichen Bewertung der Aussage "Ich befürchte krank zu sein." (Online-Evaluation) .....	66
Tabelle 19: Spezifische Bewertungen zu SymptomCheck (Online-Evaluation) .....	68
Tabelle 20: Vergleich der Gruppen "Ich befürchte krank zu sein." und "Ich befürchte nicht, krank zu sein." (Online-Evaluation).....	69
Tabelle 21: Medien zur Information über Symptome und Krankheiten (Pilotstudie II) ..	73
Tabelle 22: Internetseiten als medizinische Informationsquelle (Pilotstudie II) .....	74
Tabelle 23: Bewertung der 6 Skalen (Pilotstudie II).....	75
Tabelle 24: Spezifische Bewertungen zu SymptomCheck (Pilotstudie II) .....	77
Tabelle 25: Diagnoseliste (Pilotstudie II) .....	78
Tabelle 26: Diagnoseeinteilung nach medizinischem Fachgebiet (Pilotstudie II) .....	79

Tabelle 27: Platzierung der Diagnosen (Pilotstudie II).....	79
Tabelle 28: Mittelwert und Median der Diagnoseplatzierung (Pilotstudie II).....	81
Tabelle 29: Stratum „Ich befürchte krank zu sein“ vs „Ich befürchte nicht krank zu sein“ hinsichtlich der 26 Items des UEQ-Fragebogens (Online-Evaluation) .....	106
Tabelle 30: Genaue Übersicht über die Bewertung der spezifischen Fragen zu SymptomCheck (Online-Evaluation) .....	110
Tabelle 31: Stratum „Ich befürchte krank zu sein“ vs „Ich befürchte nicht krank zu sein“ hinsichtlich der Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation) .....	111
Tabelle 32: Stratum „männlich“ vs „weiblich“ hinsichtlich der Aussage „Ich befürchte krank zu sein“ und der Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation) .....	112
Tabelle 33: Stratum „Altersgruppe 18 bis 30 Jahre“ vs. „Altersgruppe 31 Jahre und älter“ Unterschiede hinsichtlich der Aussage „Ich befürchte krank zu sein“ und der Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation) .....	114
Tabelle 34: Ausführliche Liste der Diagnosevorschläge von SymptomCheck .....	115
Tabelle 35: Zusammengefasste Liste der Diagnosevorschläge von SymptomCheck .....	121
Tabelle 36: Beispiele nicht-pathologischer Erklärungsvorschläge von SymptomCheck für die Beschwerden .....	125
Tabelle 37: Vergleich der Geschlechterverteilung (Pilotstudie I, Pilotstudie II und Online-Evaluation) .....	126
Tabelle 38: Vergleich der Altersverteilung (Pilotstudie I, Pilotstudie II und Online-Evaluation) .....	126
Tabelle 39: Vergleich der Schulabschlüsse (Pilotstudie I, Pilotstudie II und Online-Evaluation) .....	126
Tabelle 40: Bewertung der 26 Items des UEQ-Fragebogens (Pilotstudie I).....	127
Tabelle 41: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation) .....	128
Tabelle 42: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation: Ich befürchte krank zu sein; Antworten berücksichtigt: trifft völlig zu & trifft eher zu) .....	129
Tabelle 43: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation: Ich befürchte krank zu sein; Antworten berücksichtigt: trifft gar nicht zu & trifft eher nicht zu) .....	130

Tabelle 44: Bewertung der 26 Items des UEQ-Fragebogens (Pilotstudie II).....	131
Tabelle 45: Skalenvergleich geschlechtsabhängig (Pilotstudie II) .....	131
Tabelle 46: Skalenvergleich altersabhängig (Pilotstudie II) .....	132
Tabelle 47: Detaillierte Diagnoseliste (Pilotstudie I) .....	133

## Abbildungsverzeichnis

Abbildung 1: Detaillierte Symptomabfrage zum Beispiel "akuter Bauchschmerz" .....	22
Abbildung 2: Startseite von SymptomCheck .....	33
Abbildung 3: Beispiele für zusätzliche Symptommöglichkeiten .....	34
Abbildung 4: Frauenbeschwerden.....	35
Abbildung 5: 6 Skalen des UEQ-Fragebogens; übersetzt und verändert aus dem Englischen <sup>69</sup> .....	41
Abbildung 6: Fiktives Beispiel zur Bewertung der 6 Skalen (UEQ-Fragebogen).....	42
Abbildung 7: Korrekte Diagnose unter den ersten 3 Plätzen (SymptomCheck; n=26) .	48
Abbildung 8: Korrekte Diagnose unter den ersten 3 Plätzen (Arzt; n=26).....	49
Abbildung 9: Korrekte Diagnose auf erstem Platz (SymptomCheck; n=26) .....	50
Abbildung 10: Korrekte Diagnose auf erstem Platz (Arzt; n=26).....	50
Abbildung 11: Korrekte Diagnose auf erstem Platz (Arzt; n=13).....	51
Abbildung 12: Geschlechterverteilung (Pilotstudie I; n=16) .....	52
Abbildung 13: Altersverteilung (Pilotstudie I; n=16) .....	52
Abbildung 14: Aufteilung der Patienten nach Schulabschlüssen (Pilotstudie I; n=16)..	53
Abbildung 15: Bewertung der 6 Skalen (Pilotstudie I).....	54
Abbildung 16: Geschlechtsabhängige Bewertung der 6 Skalen (Pilotstudie I).....	55
Abbildung 17: Überprüfung der Verdachtsgenerierung: Überblick über die Platzierung der Diagnosen (Pilotstudie I) .....	58
Abbildung 18: Geschlechterverteilung (Online-Evaluation; n=75).....	60
Abbildung 19: Altersverteilung (Online-Evaluation; n=75) .....	60
Abbildung 20: Aufteilung der Probanden nach Schulabschlüssen (Online-Evaluation; n=75).....	61
Abbildung 21: Medien zur Information über Symptome und Krankheiten (Online-Evaluation; n=75) .....	62
Abbildung 22: Bewertung der 6 Skalen (Online-Evaluation) .....	64
Abbildung 23: Geschlechtsabhängige Bewertung der 6 Skalen (Online-Evaluation) ...	65
Abbildung 24: Unterschiede bezüglich der Bewertung der 6 Skalen hinsichtlich der unterschiedlichen Bewertung der Aussage "Ich befürchte krank zu sein." (Online-Evaluation).....	66
Abbildung 25: Geschlechterverteilung (Pilotstudie II; n=33) .....	72
Abbildung 26: Altersverteilung (Pilotstudie II; n=33) .....	72
Abbildung 27: Aufteilung der Patienten nach Schulabschlüssen (Pilotstudie II; n=33).	73

Abbildung 28: Bewertung der 6 Skalen (Pilotstudie II).....	75
Abbildung 29: Überprüfung der Verdachtsgenerierung: Überblick über die Platzierung der Diagnosen (Pilotstudie II) .....	80
Abbildung 30: Probandeninformation der Online-Evaluation .....	135
Abbildung 31: Online-Evaluation.....	136
Abbildung 32: Transparentes Bewertungssystem .....	136

# Anhang

## 1 Tabellen

Nachstehend sind ergänzende Analysen aufgeführt, die Gruppenunterschiede im Hinblick auf die 26 Fragen des Online-Fragebogens zur Benutzerfreundlichkeit untersuchten.

**Tabelle 29: Stratum „Ich befürchte krank zu sein“ vs „Ich befürchte nicht krank zu sein“ hinsichtlich der 26 Items des UEQ-Fragebogens (Online-Evaluation)**

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
erfreulich	trifft nicht zu	3	8	11	0,7
	trifft zu	13	22	35	
Gesamt		16	30	46	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
verständlich	trifft nicht zu	3	11	14	0,2
	trifft zu	16	21	37	
Gesamt		19	32	51	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
kreativ	trifft nicht zu	8	11	19	1,0
	trifft zu	9	12	21	
Gesamt		17	23	40	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
leicht zu lernen	trifft nicht zu	4	6	10	1,0
	trifft zu	15	26	41	
Gesamt		19	32	51	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
erfrischend	trifft nicht zu	5	7	12	0,7
	trifft zu	8	16	24	
Gesamt		13	23	36	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
spannend	trifft nicht zu	2	8	10	0,7
	trifft zu	10	19	29	
Gesamt		12	27	39	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
interessant	trifft nicht zu	1	6	7	0,2
	trifft zu	17	24	41	
Gesamt		18	30	48	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
voraussagbar	trifft nicht zu	1	1	2	1,0
	trifft zu	13	26	39	
Gesamt		14	27	41	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
schnell	trifft nicht zu	3	9	12	0,7
	trifft zu	14	26	40	
Gesamt		17	35	52	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
neu	trifft nicht zu	1	3	4	1,0
	trifft zu	16	25	41	
Gesamt		17	28	45	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
bedienbar	trifft nicht zu	3	5	8	1,0
	trifft zu	15	29	44	
Gesamt		18	34	52	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
gut	trifft nicht zu	2	10	12	0,2
	trifft zu	16	24	40	
Gesamt		18	34	52	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
kompliziert	trifft nicht zu	1	6	7	0,2
	trifft zu	18	26	44	
Gesamt		19	32	51	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
anziehend	trifft nicht zu	1	7	8	0,2
	trifft zu	14	19	33	
Gesamt		15	26	41	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
modern	trifft nicht zu	4	5	9	1,0
	trifft zu	14	20	34	
Gesamt		18	25	43	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
angenehm	trifft nicht zu	2	5	7	0,7
	trifft zu	16	21	37	
Gesamt		18	26	44	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
vorhersagbar	trifft nicht zu	3	5	8	0,7
	trifft zu	10	23	33	
Gesamt		13	28	41	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
abwechslungsreich	trifft nicht zu	5	8	13	1,0
	trifft zu	10	14	24	
Gesamt		15	22	37	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
zuverlässig	trifft nicht zu	4	9	13	0,3
	trifft zu	12	12	24	
Gesamt		16	21	37	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
effizient	trifft nicht zu	4	8	12	0,7
	trifft zu	14	20	34	
Gesamt		18	28	46	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
übersichtlich	trifft nicht zu	10	16	26	1,0
	trifft zu	9	15	24	
Gesamt		19	31	50	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
flüssig	trifft nicht zu	6	8	14	0,7
	trifft zu	10	20	30	
Gesamt		16	28	44	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
aufgeräumt	trifft nicht zu	6	19	25	0,5
	trifft zu	8	14	22	
Gesamt		14	33	47	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
schön	trifft nicht zu	10	16	26	1,0
	trifft zu	8	11	19	
Gesamt		18	27	45	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
sympathisch	trifft nicht zu	4	7	11	1,0
	trifft zu	11	18	29	
Gesamt		15	25	40	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
auffällig	trifft nicht zu	6	8	14	0,7
	trifft zu	9	8	17	
Gesamt		15	16	31	

**Tabelle 30: Genauer Übersicht über die Bewertung der spezifischen Fragen zu SymptomCheck (Online-Evaluation)**

		Anzahl (n)	Anteil (%)	Gültige Prozente
<b>Ich konnte meine Symptome gut finden.</b>				
Gültig	Ja	48	64,0	73,8
	Neutral	1	1,3	1,5
	Nein	16	21,1	24,6
	Gesamt	65	86,7	100,0
Fehlend		10	13,3	
Gesamt		75	100,0	
<b>Unklare Begriffe wurden mir gut erklärt.</b>				
Gültig	Ja	40	53,3	65,6
	Neutral	9	12,0	14,8
	Nein	12	16,0	19,7
	Gesamt	61	81,3	100,0
Fehlend		14	18,7	
Gesamt		75	100,0	
<b>Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.</b>				
Gültig	Ja	63	84,0	96,9
	Neutral	1	1,3	1,5
	Nein	1	1,3	1,5
	Gesamt	65	86,7	100,0
Fehlend		10	13,3	
Gesamt		75	100,0	
<b>Dieses Programm ist im Alltag sehr nützlich.</b>				
Gültig	Ja	42	56,0	64,6
	Neutral	10	13,3	15,4
	Nein	13	17,3	20,0
	Gesamt	65	86,7	100,0
Fehlend		10	13,3	
Gesamt		75	100,0	
<b>Das Programm hat mir besser geholfen als andere Informationsplattformen im Internet.</b>				
Gültig	Ja	25	33,3	38,5
	Neutral	20	26,7	30,8
	Nein	20	26,7	30,8
	Gesamt	65	86,7	100,0
Fehlend		10	13,3	
Gesamt		75	100,0	

		Anzahl (n)	Anteil (%)	Gültige Prozente
Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.				
Gültig	Ja	49	65,3	77,8
	Neutral	4	5,3	6,3
	Nein	10	13,3	15,9
	Gesamt	63	84,0	100,0
Fehlend		12	16,0	
Gesamt		75	100,0	
Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel)				
Gültig	Ja	59	78,7	90,8
	Neutral	2	2,7	3,1
	Nein	4	5,3	6,2
	Gesamt	65	86,7	100,0
Fehlend	System	10	13,3	
Gesamt		75	100,0	
Die Dauer für die Beantwortung der Fragen war...				
Gültig	zu lang	3	4,0	4,3
	genau richtig	57	76,0	81,4
	zu kurz	10	13,3	14,3
	Gesamt	70	93,3	100,0
Fehlend	System	5	6,7	
Gesamt		75	100,0	

**Tabelle 31: Stratum „Ich befürchte krank zu sein“ vs „Ich befürchte nicht krank zu sein“ hinsichtlich der Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation)**

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Ich konnte meine Symptome gut finden.	Ja	16	25	41	0,3
	Nein	3	11	14	
Gesamt		19	36	55	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Unklare Begriffe wurden mir gut erklärt.	Ja	11	21	32	0,7
	Nein	3	9	12	
Gesamt		14	30	44	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.	Ja	19	35	54	1,0
	Nein	0	1	1	
Gesamt		19	36	55	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Dieses Programm ist im Alltag sehr nützlich.	Ja	15	20	35	0,3
	Nein	3	9	12	
Gesamt		18	29	47	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Das Programm hat mir besser geholfen, als andere Informationsplattformen im Internet.	Ja	10	13	23	0,2
	Nein	4	14	18	
Gesamt		14	27	41	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	Ja	15	27	42	0,7
	Nein	2	7	9	
Gesamt		17	34	51	

		Ich befürchte krank zu sein.		Gesamt	P
		Trifft zu.	Trifft nicht zu.		
Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel)	Ja	17	34	51	0,6
	Nein	2	2	4	
Gesamt		19	36	55	

**Tabelle 32: Stratum „männlich“ vs „weiblich“ hinsichtlich der Aussage „Ich befürchte krank zu sein“ und der Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation)**

		Geschlecht		Gesamt	P
		männlich	weiblich		
Ich befürchte krank zu sein.	Ja	11	10	21	0,1
	Nein	13	30	43	
Gesamt		24	40	64	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Ich konnte meine Symptome gut finden.	Ja	17	31	48	0,5
	Nein	4	12	16	
Gesamt		21	43	64	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Unklare Begriffe wurden mir gut erklärt.	Ja	11	29	40	0,5
	Nein	5	7	12	
Gesamt		16	36	52	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.	Ja	21	42	63	1,0
	Nein	0	1	1	
Gesamt		21	43	64	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Dieses Programm ist im Alltag sehr nützlich.	Ja	14	28	42	0,7
	Nein	3	10	13	
Gesamt		17	38	55	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Das Programm hat mir besser geholfen als andere Informationsplattformen im Internet.	Ja	8	17	25	0,7
	Nein	5	15	20	
Gesamt		13	32	45	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	Ja	16	33	49	1,0
	Nein	3	7	10	
Gesamt		19	40	59	

		Geschlecht		Gesamt	P
		männlich	weiblich		
Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel).	Ja	19	40	59	0,6
	Nein	2	2	4	
Gesamt		21	42	63	

**Tabelle 33: Stratum „Altersgruppe 18 bis 30 Jahre“ vs. „Altersgruppe 31 Jahre und älter“  
Unterschiede hinsichtlich der Aussage „Ich befürchte krank zu sein“ und der  
Bewertung spezifischer Aussagen zu SymptomCheck (Online-Evaluation)**

		Alter		Gesamt	P
		18 bis 30	ab 31		
Ich befürchte krank zu sein.	Ja	15	6	21	0,3
	Nein	36	7	43	
Gesamt		51	13	64	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Ich konnte meine Symptome gut finden.	Ja	38	10	48	1,0
	Nein	13	3	16	
Gesamt		51	13	64	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Unklare Begriffe wurden mir gut erklärt.	Ja	31	9	40	0,4
	Nein	11	1	12	
Gesamt		42	10	52	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Es ist sinnvoll, dass genauere Fragen zu meinen Symptomen gestellt werden.	Ja	51	12	63	0,2
	Nein	0	1	1	
Gesamt		51	13	64	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Dieses Programm ist im Alltag sehr nützlich.	Ja	34	8	42	0,7
	Nein	12	1	13	
Gesamt		46	9	55	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Das Programm hat mir besser geholfen als andere Informationsplattformen im Internet.	Ja	20	5	25	1,0
	Nein	16	4	20	
Gesamt		36	9	45	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	Ja	37	12	49	0,1
	Nein	10	0	10	
Gesamt		47	12	59	

		Alter		Gesamt	P
		18 bis 30	ab 31		
Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen werden (z. B. Stress, Bewegungsmangel).	Ja	46	13	59	0,572
	Nein	4	0	4	
Gesamt		50	13	63	

**Tabelle 34: Ausführliche Liste der Diagnosevorschläge von SymptomCheck**

Fachgebiet	Diagnosen
<b>Kardiologie</b>	<b>1 akutes Koronarsyndrom</b>
	<b>2 stabile Angina pectoris</b>
	<b>3 Vorhofmyxom</b>
	<b>4 Perikardtamponade</b>
	<b>5 Herzvitium</b>
	<b>6 bradykarde Herzrhythmusstörung</b> <i>darunter fallen:</i> Karotis-Sinus-Syndrom Bradyarrhythmia absoluta Sinuatrialer Block AV-Block 1,2. Grades bradykardes Sick-Sinus-Syndrom
	<b>7 tachykarde Herzrhythmusstörung</b> <i>darunter fallen:</i> tachykardes Sick-Sinus-Syndrom AV-Knoten-Reentrytachykardie Supraventrikuläre Extrasystole Sinustachykardie Torsade-de-Pointes-Tachykardie Atrioventrikuläre Reentrytachykardie (WPW-Syndrom) Fokale atriale Tachykardie Junktionale ektope Tachykardie
	<b>8 Vorhofflattern</b>
	<b>9 Vorhofflimmern</b>
	<b>10 Hypotonie</b>
	<b>11 Orthostaseintoleranz</b>
	<b>12 Stresskardiomyopathie</b>
	<b>13 Globale Herzinsuffizienz</b> <i>darunter fallen:</i>
	<b>14 Rechtsherzinsuffizienz</b>
	<b>15 Linksherzinsuffizienz</b> <i>darunter fallen:</i> leichte Linksherzinsuffizienz mäßige Linksherzinsuffizienz starke Linksherzinsuffizienz

**Pulmologie**

- rausgenommen:
  - akute Linksherzinsuffizienz
  - chronische Rechtsherzinsuffizienz
  - chronische Herzinsuffizienz
- 16 Perikarditis**
  - rausgenommen: Pericarditis constrictiva ("Panzerherz")*
- 17 Endokarditis**
  - darunter fallen:*
    - akute Endokarditis
    - subakute Endokarditis
- 18 Myokarditis**
- 19 Aortenklappenstenose**
  - darunter fallen:*
    - mittelgradige Aortenklappenstenose
    - hochgradige Aortenklappenstenose
- 20 akute Aortenklappeninsuffizienz**
- 21 chronische Mitralklappeninsuffizienz**
- 22 akute Mitralklappeninsuffizienz**
- 23 chronische Mitralklappeninsuffizienz**
- 24 Mitralklappenstenose**
- 25 Lungenembolie**
- 26 akute Bronchitis**
- 27 chronische Bronchitis**
- 28 Lungenentzündung**
  - darunter fallen:*
    - typische Lobärpneumonie
    - atypische Bronchopneumonie
- 29 Pleuritis**
- 30 Pleuraerguss**
- 31 COPD**
- 32 Asthma bronchiale**
- 33 pulmonale Hypertonie**
- 34 Lungenödem**
  - darunter fallen:*
    - Interstitielles Lungenödem
    - Alveoläres Lungenödem
- 35 Lungenfibrose**
  - darunter fallen:*
    - leichte Lungenfibrose**
    - progrediente Lungenfibrose**
- 36 Lungenkarzinom**
  - darunter fallen:*
    - Bronchialkarzinom
    - Pancoast-Syndrom
    - Bronchioloalveoläres Adenokarzinom
- 37 Bronchiektasen**
- 38 Pneumothorax**

**Gastroenterologie**

- 39 Lungenemphysem**  
*darunter fallen:*  
Lungenemphysem (bronchitischer Typ)  
Lungenemphysem (dyspnoisch-kachektischer Typ)
- 40 akutes Lungenversagen**
- 41 Cholelithiasis**
- 42 Gallensäureverlustsyndrom**
- 43 akute Gastritis**
- 44 Ulcus ventriculi**
- 45 Ulcus duodeni**
- 46 Morbus Crohn**
- 47 Colitis ulcerosa**
- 48 Gastroösophageale Refluxkrankheit**
- 49 akute Gastroenteritis**
- 50 akute Pankreatitis**
- 51 Appendizitis**
- 52 Ileus**
- 53 Peritonitis**
- 54 gastrointestinale Blutung**
- 55 Divertikulitis**
- 56 Zwerchfellhernie**
- 57 Magenkarzinom**
- 58 Ösophagusdivertikel**
- 59 Achalasie**
- 60 Kolorektales Karzinom**
- 61 Ileumtumor**
- 62 Reizdarmsyndrom**
- 63 Ösophagusruptur**
- 64 chronische Pankreatitis**
- 65 Pankreaskarzinom**
- 66 Gluten-sensitive Enteropathie**
- 67 Laktoseintoleranz**
- 68 funktionelle Dyspepsie**
- 69 Hodgkin-Lymphom**
- 70 Non-Hodgkin-Lymphom**

**Hämatonkologie**

- darunter fallen:*  
Multiples Myelom  
Immunozytom (Morbus Waldenström)  
Haarzelleukämie  
Sézary-Lymphom  
Mycosis fungoides  
follikuläres Lymphom  
Mantelzelllymphom  
Diffuses großzelliges B-Zell-Lymphom  
Mediastinales großzelliges B-Zell-Lymphom  
Burkitt-Lymphom  
CLL

	<b>71</b>	<b>Leukämie</b>
		<i>darunter fallen:</i>
		CLL
		AML
		ALL
		CML
		Polycythaemia vera
		Osteomyelofibrose
	<b>72</b>	<b>Myelodysplastisches Syndrom</b>
	<b>73</b>	<b>Koagulopathie</b>
	<b>74</b>	<b>Anämie</b>
<b>Endokrinologie</b>	<b>75</b>	<b>Hypoglykämie<sup>1</sup></b>
	<b>76</b>	<b>euthyreote Struma</b>
	<b>77</b>	<b>Hypothyreose</b>
	<b>78</b>	<b>Hyperthyreose</b>
	<b>79</b>	<b>endokrine Orbitopathie</b>
	<b>80</b>	<b>akute eitrige Thyreoiditis</b>
	<b>81</b>	<b>De-Quervain-Thyreoiditis</b>
	<b>82</b>	<b>Schilddrüsenkarzinom</b>
	<b>83</b>	<b>Hyperparathyreoidismus</b>
	<b>84</b>	<b>Diabetes mellitus</b>
	<b>85</b>	<b>Hypoparathyreoidismus</b>
	<b>86</b>	<b>hormonelle Ursache<sup>1</sup></b>
		<i>rausgenommen:</i>
		Akromegalie
		primärer Hyperaldosteronismus
		Hypoaldosteronismus
		Hyperkortisolismus = Cushing-Syndrom
		Inzidentalome der Nebennieren
		Phäochromozytom
		chronische Hypophysenvorderlappeninsuffizienz
		akute Hypophysenvorderlappeninsuffizienz
		Nebennierenrindeninsuffizienz = Hypokortisolismus
		endokrin inaktive Hypophysentumoren
		Prolaktinom
		Schwartz-Bartter-Syndrom
	<b>87</b>	<b>Androgenitales Syndrom</b>
	<b>88</b>	<b>Hirsutismus</b>
	<b>89</b>	<b>Gynäkomastie</b>
	<b>90</b>	<b>Diabetes insipidus</b>
<b>Hepatologie</b>	<b>91</b>	<b>Leberzirrhose</b>
	<b>92</b>	<b>virusbedingte Hepatitis</b>
<b>Angiologie</b>	<b>93</b>	<b>Aortendissektion</b>
	<b>94</b>	<b>abdominelle arterielle Verschlusskrankheit</b>
	<b>95</b>	<b>periphere arterielle Verschlusskrankheit</b>
	<b>96</b>	<b>tiefe Beinvenenthrombose</b>
	<b>97</b>	<b>tiefe Armvenenthrombose</b>

	98	Raynoud Phänomen
	99	Aortenisthmusstenose
	100	arterielle Hypertonie
	101	Lymphödem <i>rausgenommen:</i> renovaskuläre Hypertonie
<b>Infektiologie</b>	102	Infektion <sup>1</sup>
	103	Tuberkulose
	104	akute unspezifische Lymphadenitis
	105	Zytomegalie
	106	Meningitis
	107	Borreliose
	108	sexuell übertragbare Erkrankung
	109	Herpes labialis
	110	Keuchhusten
	111	Virus-Grippe
	112	Erkältung <sup>1</sup>
<b>Orthopädie</b>	113	Osteomalazie
	114	Osteoporose
	115	Arthritis
	116	Morbus Paget
	117	Wirbelsäulensyndrom
	118	Kyphoskoliose / Brustkorbdeformität
	119	Fraktur oder Prellung der Rippen
	120	orthopädische Erkrankung <sup>1</sup>
<b>Psychiatrie</b>	121	psychische Ursache <sup>1</sup>
	122	Herzneurose
	123	Angststörung / Panikattacke
	124	Depression
<b>Urologie</b>	125	Orchitis
	126	Hodentorsion
	127	Harninkontinenz
	128	urologische Erkrankung <sup>1</sup>
<b>Neurologie</b>	129	neurologische Erkrankung <sup>1</sup>
	130	Apoplex
	131	Polyneuropathie
	132	Spannungskopfschmerz
	133	Migräne
<b>Gynäkologie</b>	134	Gravidität
	135	Zystitis
<b>Dermatologie</b>	136	Hauterkrankung <sup>1</sup>
<b>HNO</b>	137	Sinusitis
	138	Rhinitis
	139	Rhinosinusitis
	140	Schlafapnoesyndrom
	141	Morbus Menière
	142	benigner paroxysmaler Lagerungsschwindel

	143	Otitis media
	144	Tonsillitis
<b>Stoffwechsel</b>	145	Porphyrie
<b>Nicht pathologisch</b>	146	altersbedingte Potenzstörungen <sup>1</sup>
	147	Übermüdung <sup>1</sup>
	148	laute Umgebung <sup>1</sup>
	149	altersbedingter Haarausfall <sup>1</sup>
	150	von Natur aus trockene oder dünne Haare <sup>1</sup>
	151	beanspruchte Stimme <sup>1</sup>
	152	Ödeme durch langes Stehen <sup>1</sup>
	153	von Natur aus kalte Hände <sup>1</sup>
	154	von Natur aus blasser Hauttyp <sup>1</sup>
	155	von Natur aus trockene Haut <sup>1</sup>
	156	Mückenstich <sup>1</sup>
	157	natürliche Bräune <sup>1</sup>
	158	von Natur aus starke Unterleibsschmerzen <sup>1</sup>
	159	von Natur aus starke Körperbehaarung <sup>1</sup>
	160	von Natur aus kleine Brust <sup>1</sup>
	161	von Natur aus muskulöser Typ <sup>1</sup>
	162	von Natur aus ungleich große Brust <sup>1</sup>
	163	Dehydratation (zu wenig getrunken) <sup>1</sup>
	164	schlechte Kondition <sup>1</sup>
	165	Bewegungsmangel <sup>1</sup>
	166	ernährungsbedingte Symptome <sup>1</sup>
	167	Reisekrankheit /Kinetose
		<i>rausgenommen:</i>
		Schichtarbeit, unbequeme Schlafgelegenheit
<b>Sonstiges</b>	168	Vitamin K Mangel
	169	Elektrolytstörung
	170	Nikotinabusus <sup>1</sup>
	171	Synkope
	172	unfallsbedingte Verletzung <sup>1</sup>
	173	akute Hyperventilation
	174	chronische Hyperventilation
	175	Vergiftung
	176	Allergie
	177	Medikamentennebenwirkungen <sup>1</sup>
	178	Adipositas

<sup>1</sup> Keine ICD-Diagnose

Nicht nummerierte Diagnosen sind in der ursprünglichen, nicht aber in der endgültigen Version enthalten.

**Tabelle 35: Zusammengefasste Liste der Diagnosevorschläge von SymptomCheck**

Fachgebiet	Diagnosen	Insgesamt
Kardiologie	1 akutes Koronarsyndrom	1
	2 stabile Angina pectoris	2
	3 Vorhofmyxom	3
	4 Perikardtamponade	4
	5 Herzvitium	5
	6 bradykarde Herzrhythmusstörung	6
	7 tachykarde Herzrhythmusstörung	7
	8 Vorhofflattern	8
	9 Vorhofflimmern	9
	10 Hypotonie	10
	11 Orthostaseintoleranz	11
	12 Stresskardiomyopathie	12
	13 Globale Herzinsuffizienz	13
	14 Rechtsherzinsuffizienz	14
	15 Linksherzinsuffizienz	15
	16 Perikarditis	16
	17 Endokarditis	17
	18 Myokarditis	18
	19 Aortenklappenstenose	19
	20 akute Aortenklappeninsuffizienz	20
	21 chronische Mitralklappeninsuffizienz	21
	22 akute Mitralklappeninsuffizienz	22
	23 chronische Mitralklappeninsuffizienz	23
	24 Mitralklappenstenose	24
Pulmologie	1 Lungenembolie	25
	2 akute Bronchitis	26
	3 chronische Bronchitis	27
	4 Lungenentzündung	28
	5 Pleuritis	29
	6 Pleuraerguss	30
	7 COPD	31
	8 Asthma bronchiale	32
	9 pulmonale Hypertonie	33
	10 Lungenödem	34
	11 Lungenfibrose	35
	12 Lungenkarzinom	36
	13 Bronchiektasen	37
	14 Pneumothorax	38
	15 Lungenemphysem	39
	16 akutes Lungenversagen	40

Gastroenterologie	1	Cholelithiasis	41
	2	Gallensäureverlustsyndrom	42
	3	akute Gastritis	43
	4	Ulcus ventriculi	44
	5	Ulcus duodeni	45
	6	Morbus Crohn	46
	7	Colitis ulcerosa	47
	8	Gastroösophageale Refluxkrankheit	48
	9	akute Gastroenteritis	49
	10	akute Pankreatitis	50
	11	Appendizitis	51
	12	Ileus	52
	13	Peritonitis	53
	14	gastrointestinale Blutung	54
	15	Divertikulitis	55
	16	Zwerchfellhernie	56
	17	Magenkarzinom	57
	18	Ösophagusdivertikel	58
	19	Achalasie	59
	20	Kolorektales Karzinom	60
	21	Ileumtumor	61
	22	Reizdarmsyndrom	62
	23	Ösophagusruptur	63
	24	chronische Pankreatitis	64
	25	Pankreaskarzinom	65
	26	Gluten-sensitive Enteropathie	66
	27	Laktoseintoleranz	67
	28	funktionelle Dyspepsie	68
Hämatonkologie	1	Hodgkin-Lymphom	69
	2	Non-Hodgkin-Lymphom	70
	3	Leukämie	71
	4	Myelodysplastisches Syndrom	72
	5	Koagulopathie	73
	6	Anämie	74
Endokrinologie	1	Hypoglykämie <sup>1</sup>	75
	2	euthyreote Struma	76
	3	Hypothyreose	77
	4	Hyperthyreose	78
	5	endokrine Orbitopathie	79
	6	akute eitrige Thyreoiditis	80
	7	De-Quervain-Thyreoiditis	81
	8	Schilddrüsenkarzinom	82

	9	Hyperparathyreoidismus	83
	10	Diabetes mellitus	84
	11	Hypoparathyreoidismus	85
	12	hormonelle Ursache <sup>1</sup>	86
	13	Androgenitales Syndrom	87
	14	Hirsutismus	88
	15	Gynäkomastie	89
	16	Diabetes insipidus	90
Hepatology	1	Leberzirrhose	91
	2	virusbedingte Hepatitis	92
Angiologie	1	Aortendissektion	93
	2	abdominelle arterielle Verschlusskrankheit	94
	3	periphere arterielle Verschlusskrankheit	95
	4	tiefe Beinvenenthrombose	96
	5	tiefe Armvenenthrombose	97
	6	Raynaud Phänomen	98
	7	Aortenisthmusstenose	99
	8	Hypertonie	100
	9	Lymphödem	101
Infektiologie	1	Infektion <sup>1</sup>	102
	2	Tuberkulose	103
	3	akute unspezifische Lymphadenitis	104
	4	Zytomegalie	105
	5	Meningitis	106
	6	Borreliose	107
	7	sexuell übertragbare Erkrankung	108
	8	Herpes labialis	109
	9	Keuchhusten	110
	10	Virus-Grippe	111
	11	Erkältung <sup>1</sup>	112
Orthopädie	1	Osteomalazie	113
	2	Osteoporose	114
	3	Arthritis	115
	4	Morbus Paget	116
	5	Wirbelsäulensyndrom	117
	6	Kyphoskoliose / Brustkorbdeformität	118
	7	Fraktur oder Prellung der Rippen	119
	8	orthopädische Erkrankung <sup>1</sup>	120
Psychiatrie	1	psychische Ursache <sup>1</sup>	121

	2	Herzneurose	122
	3	Angststörung / Panikattacke	123
	4	Depression	124
Urologie	1	Orchitis	125
	2	Hodentorsion	126
	3	Harninkontinenz	127
	4	urologische Erkrankung <sup>1</sup>	128
Neurologie	1	neurologische Erkrankung <sup>1</sup>	129
	2	Apoplex	130
	3	Polyneuropathie	131
	4	Spannungskopfschmerz	132
	5	Migräne	133
Gynäkologie	1	Gravidität	134
	2	Zystitis	135
Dermatologie	1	Hauterkrankung <sup>1</sup>	136
HNO	1	Sinusitis	137
	2	Rhinitis	138
	3	Rhinosinusitis	139
	4	Schlafapnoesyndrom	140
	5	Morbus Menière	141
	6	benigner paroxysmaler Lagerungsschwindel	142
	7	Otitis media	143
	8	Tonsillitis	144
Stoffwechsel	1	Porphyrie	145
Nicht pathologisch	1	altersbedingte Potenzstörungen <sup>1</sup>	146
	2	Übermüdung <sup>1</sup>	147
	3	laute Umgebung <sup>1</sup>	148
	4	altersbedingter Haarausfall <sup>1</sup>	149
	5	von Natur aus trockene oder dünne Haare <sup>1</sup>	150
	6	beanspruchte Stimme <sup>1</sup>	151
	7	Ödeme durch langes Stehen <sup>1</sup>	152
	8	von Natur aus kalte Hände <sup>1</sup>	153
	9	von Natur aus blasser Hauttyp <sup>1</sup>	154
	10	von Natur aus trockene Haut <sup>1</sup>	155
	11	Mückenstich <sup>1</sup>	156
	12	natürliche Bräune <sup>1</sup>	157
	13	von Natur aus starke Unterleibsschmerzen <sup>1</sup>	158

	14	von Natur aus starke Körperbehaarung <sup>1</sup>	159
	15	von Natur aus kleine Brust <sup>1</sup>	160
	16	von Natur aus muskulöser Typ <sup>1</sup>	161
	17	von Natur aus ungleich große Brust <sup>1</sup>	162
	18	Dehydratation (zu wenig getrunken) <sup>1</sup>	163
	19	schlechte Kondition <sup>1</sup>	164
	20	Bewegungsmangel <sup>1</sup>	165
	21	ernährungsbedingte Symptome <sup>1</sup>	166
	22	Reisekrankheit /Kinetose	167
Sonstiges	1	Vitamin K Mangel	168
	2	Elektrolytstörung	169
	3	Nikotinabusus <sup>1</sup>	170
	4	Synkope	171
	5	unfallsbedingte Verletzung <sup>1</sup>	172
	6	akute Hyperventilation	173
	7	chronische Hyperventilation	174
	8	Vergiftung	175
	9	Allergie	176
	10	Medikamentennebenwirkungen <sup>1</sup>	177
	11	Adipositas	178

<sup>1</sup> Keine ICD-Diagnose

**Tabelle 36: Beispiele nicht-pathologischer Erklärungsvorschläge von SymptomCheck für die Beschwerden**

Fragen-Antworten	H	H	H	H
	psychische Ursache	Dehydratation (zu wenig getrunken)	Bewegungs-mangel / zu wenig Sport	Symptome sind ernährungs-bedingt (schlechte Ernährung, etc.)
Kopf-/ Gesichtsschmerzen	P1	P4		
Rückenschmerzen	P1		P4	
Verdauungsprobleme & Stuhlgangs-besonderheiten	P1	P4	P1	P4
starke Anlagerung von Körperfett am Bauch			P4	P4
Schwindel	P2	P4		
Haarausfall	P4			P4
Verstopfung	P1	P4	P4	P4
Seit wann leiden Sie unter Verstopfung?				
erst seit kurzem, plötzlich aufgetreten	P1	P4	P4	P4
schon seit längerer Zeit	P1	P4	P4	P4
ungewöhnliche Farbe des Stuhlgangs				P4

**Tabelle 37: Vergleich der Geschlechterverteilung (Pilotstudie I, Pilotstudie II und Online-Evaluation)**

Geschlecht	Pilotstudie I		Pilotstudie II		Online-Evaluation	
	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]
männlich	10	62,5	22	66,7	25	32,9
weiblich	6	37,5	11	33,3	50	65,8
Gesamt	16	100,0	33	100,0	75	100,0

**Tabelle 38: Vergleich der Altersverteilung (Pilotstudie I, Pilotstudie II und Online-Evaluation)**

Altersgruppe	Pilotstudie I		Pilotstudie II		Online-Evaluation	
	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]
18 bis 20	0	0,0	0	0,0	1	1,3
21 bis 30	0	0,0	1	3,0	57	76,0
31 bis 40	1	6,3	3	9,1	4	5,3
41 bis 50	3	18,8	6	18,2	3	4,0
51 bis 60	5	31,3	8	24,2	9	12,0
61 bis 70	5	31,3	11	33,3	1	1,3
71 bis 80	2	12,5	4	12,1	0	0,0
Gesamt	16	100,0	33	100,0	75	100,0

**Tabelle 39: Vergleich der Schulabschlüsse (Pilotstudie I, Pilotstudie II und Online-Evaluation)**

Schulabschluss	Pilotstudie I		Pilotstudie II		Online-Evaluation	
	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]	Anzahl [n]	Anteil [%]
allgemeine oder fachgebundene Hochschulreife	5	31,3	12	36,4	67	89,3
mittlere Reife	5	31,3	7	21,2	7	9,3
Hauptschulabschluss	6	37,5	11	33,3	1	1,3
keinen Schulabschluss / Andere	0	0,0	2	6,1	0	0,0
keine Aussage	0	0	1	3,0	0	0
Gesamt	16	100,0	33	100,0	75	100,0

Tabelle 40: Bewertung der 26 Items des UEQ-Fragebogens (Pilotstudie I)

Item	Skala	Item		n	MW	SD
		Negative Polung	Positive Polung			
1	Attraktivität	unerfreulich	erfreulich	15	↑ 1,7	1,4
2	Durchschaubarkeit	unverständlich	verständlich	15	↑ 2,0	1,2
3	Originalität	phantasielos	kreativ	15	↑ 1,3	1,4
4	Durchschaubarkeit	schwer zu lernen	leicht zu lernen	15	↑ 2,3	1,4
5	Stimulation	minderwertig	wertvoll	15	↑ 1,2	1,0
6	Stimulation	langweilig	spannend	15	↑ 0,9	1,6
7	Stimulation	uninteressant	interessant	15	↑ 1,7	1,2
8	Steuerbarkeit	unberechenbar	voraussagbar	15	↔ 0,7	1,9
9	Effizienz	langsam	schnell	15	↑ 1,7	1,6
10	Originalität	konventionell	originell	15	↑ 2,1	1,3
11	Steuerbarkeit	behindernd	unterstützend	15	↑ 1,7	1,9
12	Attraktivität	schlecht	gut	15	↑ 2,0	1,3
13	Durchschaubarkeit	kompliziert	einfach	15	↑ 1,7	1,7
14	Attraktivität	abstoßend	anziehend	15	↑ 1,2	1,1
15	Originalität	herkömmlich	neuartig	15	↑ 1,7	1,1
16	Attraktivität	unangenehm	angenehm	15	↑ 2,1	0,9
17	Steuerbarkeit	unsicher	sicher	15	↔ 0,5	1,6
18	Stimulation	einschläfernd	aktivierend	15	↑ 1,3	1,1
19	Steuerbarkeit	nicht erwartungskonform	erwartungskonform	15	↑ 1,1	1,2
20	Effizienz	ineffizient	effizient	15	↑ 1,7	1,2
21	Durchschaubarkeit	verwirrend	übersichtlich	15	↑ 2,2	1,4
22	Effizienz	unpragmatisch	pragmatisch	15	↑ 2,0	1,1
23	Effizienz	überladen	aufgeräumt	15	↑ 1,6	1,6
24	Attraktivität	unattraktiv	attraktiv	15	↑ 1,7	1,3
25	Attraktivität	unsympathisch	sympathisch	15	↑ 1,9	1,2
26	Originalität	konservativ	innovativ	15	↑ 1,3	1,6

Tabelle 41: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation)

Item	Skala	Item		n	MW	SD
		Negative Polung	Positive Polung			
1	Attraktivität	unerfreulich	erfreulich	65	↑ 0,9	1,6
2	Durchschaubarkeit	unverständlich	verständlich	66	↑ 1,1	1,6
3	Originalität	phantasielos	kreativ	65	↔ 0,1	1,5
4	Durchschaubarkeit	schwer zu lernen	leicht zu lernen	66	↑ 1,3	1,6
5	Stimulation	minderwertig	wertvoll	66	↔ 0,3	1,4
6	Stimulation	langweilig	spannend	65	↔ 0,5	1,5
7	Stimulation	uninteressant	interessant	66	↑ 1,0	1,4
8	Steuerbarkeit	unberechenbar	voraussagbar	66	↑ 1,1	1,2
9	Effizienz	langsam	schnell	66	↑ 1,2	1,6
10	Originalität	konventionell	originell	66	↑ 1,1	1,4
11	Steuerbarkeit	behindernd	unterstützend	65	↑ 1,6	1,5
12	Attraktivität	schlecht	gut	64	↑ 0,8 <sup>1</sup>	1,6
13	Durchschaubarkeit	kompliziert	einfach	64	↑ 1,5	1,5
14	Attraktivität	abstoßend	anziehend	64	↔ 0,6	1,5
15	Originalität	herkömmlich	neuartig	64	↔ 0,7	1,5
16	Attraktivität	unangenehm	angenehm	62	↑ 0,8 <sup>1</sup>	1,4
17	Steuerbarkeit	unsicher	sicher	62	↑ 0,8 <sup>1</sup>	1,3
18	Stimulation	einschläfernd	aktivierend	62	↔ 0,3	1,4
19	Steuerbarkeit	nicht erwartungskonform	erwartungskonform	63	↔ 0,2	1,6
20	Effizienz	ineffizient	effizient	63	↔ 0,6	1,5
21	Durchschaubarkeit	verwirrend	übersichtlich	63	↔ 0,0	1,9
22	Effizienz	unpragmatisch	pragmatisch	62	↔ 0,7	1,8
23	Effizienz	überladen	aufgeräumt	61	↔ -0,4	1,8
24	Attraktivität	unattraktiv	attraktiv	62	↔ -0,5	1,6
25	Attraktivität	unsympathisch	sympathisch	60	↔ 0,3	1,6
26	Originalität	konservativ	innovativ	62	↔ 0,3	1,2

<sup>1</sup> Mittelwert >0,8, durch die Darstellung mit nur einer Dezimalen jedoch in der Tabelle nicht darstellbar

**Tabelle 42: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation: Ich befürchte krank zu sein; Antworten berücksichtigt: trifft völlig zu & trifft eher zu)**

Item	Skala	Item		n	MW	SD
		Negative Polung	Positive Polung			
1	Attraktivität	unerfreulich	erfreulich	18	↑ 1,2	1,7
2	Durchschaubarkeit	unverständlich	verständlich	19	↑ 1,4	1,6
3	Originalität	phantasielos	kreativ	19	↔ 0,3	1,6
4	Durchschaubarkeit	schwer zu lernen	leicht zu lernen	19	↑ 1,5	1,8
5	Stimulation	minderwertig	wertvoll	19	↔ 0,3	1,1
6	Stimulation	langweilig	spannend	19	↔ 0,8	1,5
7	Stimulation	uninteressant	interessant	19	↑ 1,7	1,0
8	Steuerbarkeit	unberechenbar	voraussagbar	19	↑ 0,9	1,1
9	Effizienz	langsam	schnell	19	↑ 1,3	1,6
10	Originalität	konventionell	originell	19	↑ 1,6	1,2
11	Steuerbarkeit	behindernd	unterstützend	19	↑ 1,6	1,6
12	Attraktivität	schlecht	gut	19	↑ 1,1	1,5
13	Durchschaubarkeit	kompliziert	einfach	19	↑ 2,1	1,1
14	Attraktivität	abstoßend	anziehend	19	↑ 1,3	1,3
15	Originalität	herkömmlich	neuartig	19	↑ 1,3	1,6
16	Attraktivität	unangenehm	angenehm	19	↑ 1,3	1,2
17	Steuerbarkeit	unsicher	sicher	19	↔ 0,6	1,2
18	Stimulation	einschläfernd	aktivierend	18	↔ 0,6	1,7
19	Steuerbarkeit	nicht erwartungskonform	erwartungskonform	19	↔ 0,6	1,9
20	Effizienz	ineffizient	effizient	19	↑ 0,9	1,4
21	Durchschaubarkeit	verwirrend	übersichtlich	19	↔ 0,1	1,9
22	Effizienz	unpragmatisch	pragmatisch	19	↔ 0,7	1,9
23	Effizienz	überladen	aufgeräumt	18	↔ 0,2	1,6
24	Attraktivität	unattraktiv	attraktiv	19	↔ -0,3	1,6
25	Attraktivität	unsympathisch	sympathisch	19	↔ 0,5	1,7
26	Originalität	konservativ	innovativ	19	↔ 0,6	1,7

**Tabelle 43: Bewertung der 26 Items des UEQ-Fragebogens (Online-Evaluation: Ich befürchte krank zu sein; Antworten berücksichtigt: trifft gar nicht zu & trifft eher nicht zu)**

Item	Skala	Item		n	MW	SD
		Negative Polung	Positive Polung			
1	Attraktivität	unerfreulich	erfreulich	38	↑ 0,9	1,5
2	Durchschaubarkeit	unverständlich	verständlich	38	↑ 0,8 <sup>1</sup>	1,7
3	Originalität	phantasielos	kreativ	37	↔ 0,2	1,5
4	Durchschaubarkeit	schwer zu lernen	leicht zu lernen	38	↑ 1,2	1,6
5	Stimulation	minderwertig	wertvoll	38	↔ 0,4	1,5
6	Stimulation	langweilig	spannend	37	↔ 0,6	1,6
7	Stimulation	uninteressant	interessant	38	↑ 0,9	1,5
8	Steuerbarkeit	unberechenbar	voraussagbar	38	↑ 1,2	1,1
9	Effizienz	langsam	schnell	38	↑ 1,2	1,7
10	Originalität	konventionell	originell	38	↑ 1,0	1,4
11	Steuerbarkeit	behindernd	unterstützend	37	↑ 1,6	1,4
12	Attraktivität	schlecht	gut	37	↔ 0,8	1,6
13	Durchschaubarkeit	kompliziert	einfach	37	↑ 1,2	1,6
14	Attraktivität	abstoßend	anziehend	37	↔ 0,5	1,5
15	Originalität	herkömmlich	neuartig	37	↔ 0,7	1,5
16	Attraktivität	unangenehm	angenehm	35	↑ 0,8 <sup>1</sup>	1,4
17	Steuerbarkeit	unsicher	sicher	35	↑ 1,0	1,3
18	Stimulation	einschläfernd	aktivierend	36	↔ 0,3	1,3
19	Steuerbarkeit	nicht erwartungskonform	erwartungskonform	36	↔ 0,0	1,6
20	Effizienz	ineffizient	effizient	36	↔ 0,5	1,6
21	Durchschaubarkeit	verwirrend	übersichtlich	36	↔ -0,2	2,0
22	Effizienz	unpragmatisch	pragmatisch	35	↔ 0,8	1,9
23	Effizienz	überladen	aufgeräumt	35	↔ -0,6	2,0
24	Attraktivität	unattraktiv	attraktiv	35	↔ -0,5	1,7
25	Attraktivität	unsympathisch	sympathisch	33	↔ 0,4	1,5
26	Originalität	konservativ	innovativ	35	↔ 0,1	1,0

<sup>1</sup> Mittelwert >0,8, durch die Darstellung mit nur einer Dezimalen jedoch in der Tabelle nicht darstellbar

**Tabelle 44: Bewertung der 26 Items des UEQ-Fragebogens (Pilotstudie II)**

Item	Skala	Item		n	MW	SD
		Negative Polung	Positive Polung			
1	Attraktivität	unerfreulich	erfreulich	30	↑ 2,1	1,2
2	Durchschaubarkeit	unverständlich	verständlich	30	↑ 2,3	1,1
3	Originalität	phantasielos	kreativ	30	↑ 1,9	1,3
4	Durchschaubarkeit	schwer zu lernen	leicht zu lernen	30	↑ 2,4	1,0
5	Stimulation	minderwertig	wertvoll	30	↑ 1,8	1,1
6	Stimulation	langweilig	spannend	30	↑ 1,8	1,0
7	Stimulation	uninteressant	interessant	29	↑ 2,3	0,9
8	Steuerbarkeit	unberechenbar	voraussagbar	28	↑ 0,8 <sup>1</sup>	1,2
9	Effizienz	langsam	schnell	30	↑ 1,9	1,3
10	Originalität	konventionell	originell	29	↑ 2,2	0,9
11	Steuerbarkeit	behindernd	unterstützend	30	↑ 2,3	1,2
12	Attraktivität	schlecht	gut	30	↑ 2,5	0,9
13	Durchschaubarkeit	kompliziert	einfach	30	↑ 2,1	1,4
14	Attraktivität	abstoßend	anziehend	30	↑ 1,7	1,1
15	Originalität	herkömmlich	neuartig	30	↑ 2,1	1,0
16	Attraktivität	unangenehm	angenehm	30	↑ 2,1	1,0
17	Steuerbarkeit	unsicher	sicher	30	↑ 0,9	1,4
18	Stimulation	einschläfernd	aktivierend	30	↑ 1,8	1,1
19	Steuerbarkeit	nicht erwartungskonform	erwartungskonform	30	↑ 1,3	1,2
20	Effizienz	ineffizient	effizient	30	↑ 1,5	1,4
21	Durchschaubarkeit	verwirrend	übersichtlich	30	↑ 2,1	1,3
22	Effizienz	unpragmatisch	pragmatisch	30	↑ 2,2	1,0
23	Effizienz	überladen	aufgeräumt	30	↑ 1,8	1,4
24	Attraktivität	unattraktiv	attraktiv	30	↑ 1,4	1,4
25	Attraktivität	unsympathisch	sympathisch	30	↑ 1,8	1,4
26	Originalität	konservativ	innovativ	30	↔ -0,2	1,9

<sup>1</sup> Mittelwert >0,8, durch die Darstellung mit nur einer Dezimalen jedoch in der Tabelle nicht darstellbar

**Tabelle 45: Skalenvergleich geschlechtsabhängig (Pilotstudie II)**

Geschlecht		n	Min.	Max.	MW	SD
männlich	Attraktivität	21	-0,3	3,0	1,9	0,9
	Durchschaubarkeit	21	-0,8	3,0	2,2	1,1
	Effizienz	21	0,0	3,0	1,9	1,0
	Verlässlichkeit	21	0,0	3,0	1,4	0,8
	Stimulation	21	0,0	3,0	1,9	0,9
	Originalität	21	0,0	3,0	1,6	1,0
weiblich	Attraktivität	9	0,7	3,0	2,1	0,8
	Durchschaubarkeit	9	1,0	3,0	2,4	0,7
	Effizienz	9	0,0	2,8	1,7	0,9
	Verlässlichkeit	9	0,0	2,3	1,3	0,8
	Stimulation	9	0,7	3,0	2,0	0,7
	Originalität	9	0,5	2,5	1,5	0,6

**Tabelle 46: Skalenvergleich altersabhängig (Pilotstudie II)**

Altersklasse	n	Min.	Max.	MW	SD	
21 bis 30	Attraktivität	1	2,67	2,67	2,67	.
	Durchschaubarkeit	1	3,00	3,00	3,00	.
	Effizienz	1	2,50	2,50	2,50	.
	Verlässlichkeit	1	1,25	1,25	1,25	.
	Stimulation	1	2,25	2,25	2,25	.
	Originalität	1	2,50	2,50	2,50	.
	Personenanzahl	1				
31 bis 40	Attraktivität	3	1,00	3,00	2,33	1,15
	Durchschaubarkeit	3	1,00	3,00	2,33	1,15
	Effizienz	3	0,50	3,00	2,17	1,44
	Verlässlichkeit	3	1,25	2,75	1,83	0,80
	Stimulation	3	0,50	3,00	2,17	1,44
	Originalität	3	0,75	3,00	2,25	1,30
	Personenanzahl	3				
41 bis 50	Attraktivität	4	1,00	3,00	1,96	0,95
	Durchschaubarkeit	4	1,25	3,00	2,44	0,80
	Effizienz	4	1,00	2,75	2,00	0,74
	Verlässlichkeit	4	0,75	2,25	1,31	0,66
	Stimulation	4	0,50	2,50	1,56	0,88
	Originalität	4	0,00	2,25	1,19	0,94
	Personenanzahl	4				
51 bis 60	Attraktivität	8	0,83	2,83	1,92	0,77
	Durchschaubarkeit	8	-0,25	3,00	2,22	1,06
	Effizienz	8	0,75	3,00	1,94	0,74
	Verlässlichkeit	8	0,67	3,00	1,66	0,84
	Stimulation	8	1,50	3,00	1,97	0,62
	Originalität	8	0,75	2,25	1,56	0,48
	Personenanzahl	8				
61 bis 70	Attraktivität	11	-0,33	3,00	1,79	1,02
	Durchschaubarkeit	11	-0,75	3,00	2,00	1,13
	Effizienz	11	0,00	3,00	1,52	1,11
	Verlässlichkeit	11	0,00	2,25	1,18	0,81
	Stimulation	11	0,00	3,00	1,88	0,96
	Originalität	11	0,00	3,00	1,59	0,86
	Personenanzahl	11				
71 bis 80	Attraktivität	3	1,50	2,17	1,89	0,35
	Durchschaubarkeit	3	2,25	3,00	2,50	0,43
	Effizienz	3	1,25	3,00	2,08	0,88
	Verlässlichkeit	3	0,50	1,25	0,83	0,38
	Stimulation	3	1,25	2,25	1,92	0,58
	Originalität	3	0,00	0,75	0,50	0,43
	Personenanzahl	3				

Tabelle 47: Detaillierte Diagnoseliste (Pilotstudie I)

Nr.	erfasst als	Rang	Art	Diagnose	Symptome	VG
1	stabile Angina pectoris <sup>1</sup>	72	VG	3-Gefäß-KHK; NYHA II	wenig	VK
	arterielle Hypertonie <sup>1</sup>	2	VG	arterielle Hypertonie		
	Nikotinabusus <sup>1</sup>	15	ND	Nikotinabusus		
2	chronische Herzinsuffizienz	125	VG	ischämische Kardiomyopathie	nein	VK
	Schlafapnoesyndrom		ND	Schlafapnoesyndrom	nein	
3	Übergewicht	9	VG	Adipositas Grad III	wenig	Adipositas- studie
	arterielle Hypertonie	5	ND	arterielle Hypertonie	nein	
			ND	gestörte Glucosetoleranz		
			ND	Sodbrennen		
			ND	chronische Knie-/Hüftschmerzen		
	Migräne	13	ND	Migräne		
	chronische Bronchitis	19	ND	chronische Bronchitis		
	Medikamentennebenwirkung		ND	ACE-Hemmer bedingter Reizhusten	nein	
	nicht erkannt <sup>1</sup>		VG	gastrales Non-Hodgkin-Lymphom	wenig	
	COPD <sup>1</sup>	18	ND	COPD		
	Schlafapnoesyndrom <sup>1</sup>	172	ND	Schlafapnoesyndrom		Staging
	arterielle Hypertonie	155	ND	arterielle Hypertonie	nein	
5	chronische Herzinsuffizienz <sup>1</sup>	24	VG	hochgradige Herzinsuffizienz		VK
6	Perikarderguß		VG	persistenter, hämodynamisch-nicht relevanter Perikarderguß		VK
			ND	Zustand nach kombinierter Herz- und Nierentransplantation		
	Diabetes mellitus Typ 2	53	ND	Diabetes mellitus Typ 2		
7			ND	sekundärer Hyperparathyreoidismus		VK
	chronische Herzinsuffizienz <sup>1</sup>	68	VG	LV-EF=37%		
	stabile Angina pectoris <sup>1</sup>	36	ND	KHK		
	Hypothyreose		ND	Hypothyreose		
	Depression <sup>1</sup>	22	ND	Depression		VK
	psychische Ursache <sup>1</sup>	1	ND			
8	chronische Herzinsuffizienz <sup>1</sup>	89	VG	diastolische Herzinsuffizienz		VK
	arterielle Hypertonie <sup>1</sup>	3	ND	arterielle Hypertonie		

9	arterielle Hypertonie <sup>1</sup>	8	VG	arterielle Hypertonie			EV
10	Vorhofflimmern <sup>1</sup> & Herzrhythmusstörungen <sup>1</sup>	21	VG	Arrhythmia absoluta bei Vorhofflimmern			
	Schlafapnoesyndrom <sup>1</sup>	37	ND	tachyarrhythmische Entgleisung 2014 (beschwerdefrei) Schlafapnoesyndrom			VK
11	Depression <sup>1</sup>	11	ND	chronische Bronchitis (beschwerdefrei) Depression			
	Bronchialcarcinom <sup>1</sup>	41		Bronchialcarcinom			prä-Chemo- therapie
12	chronische Herzinsuffizienz	159		LV-EF=51% (beschwerdefrei)		nein	
	stabile Angina pectoris		ND	KHK (beschwerdefrei)		nein	VK
	arterielle Hypertonie	1	ND	Hypertonie			
13	Asthma <sup>1</sup>	23	ND	Asthma			
	Gravidität <sup>1</sup>	40		Gravidität			post-stationär
14			ND	Hypothyreose			
			ND	Restless-Legs-Syndrom			
	Test abgebrochen!		VG	KHK			
			VG	LV-EF=41%			VK
15			VG	mittelgradige Mitralsuffizienz (Zustand nach Aortenklappenersatz und Mittel- und Trikuspidalregokonstraktion)			
			VG	NYHA II (kardiologisch unauffällig)			EV
16			ND	Hypercholesterinämie			
	chronische Herzinsuffizienz <sup>1</sup>	14	VG	LV-EF=40%			
	akute Aortenklappeninsuffizienz <sup>1</sup>	20	VG	leichte Aortenklappeninsuffizienz			
	chron. Aortenklappeninsuffizienz	26					EV
	chron. Mitralklappeninsuffizienz <sup>1</sup>	55	VG	mittelgradige Mitralklappeninsuffizienz			

VG=Vorstellungsgrund; ND=Nebendiagnose; VK= Verlaufskontrolle; EV=Erstvorstellung  
<sup>1</sup> zur Überbrückung der Verdachtsanerkennung verwendet

## 2 Abbildungen

Abbildung 30: Probandeninformation der Online-Evaluation

### Probandeninformation

Sehr geehrte Dame, sehr geehrter Herr,  
hiermit möchten wir Sie über die folgende klinische Prüfung informieren.  
Ihre Teilnahme daran ist freiwillig. Bitte lesen Sie jedoch den folgenden Text sorgfältig durch.

#### Was ist der Zweck dieser klinischen Prüfung?

In dieser Studie geht es darum, ein medizinisches Programm auf seine Benutzerfreundlichkeit und seinen Nutzen hin zu testen.

#### Wie läuft die Prüfung ab?

Es werden Ihnen eine Reihe von Symptomen (z.B. Schnupfen, Halsschmerzen) aufgezeigt. Ihre Aufgabe ist es die Symptome, die Sie derzeit haben, auszuwählen.

Im Anschluss werden Sie vom Programm automatisch genauer zu Ihren Symptomen befragt. Auf dem Bildschirm werden Ihnen parallel dazu wahrscheinliche Diagnosen angezeigt. Dabei sind die Diagnosen von oben nach unten in absteigender Reihenfolge nach der zutreffenden Wahrscheinlichkeit geordnet. Das bedeutet, dass die Diagnose, die Ihnen an erster Stelle angezeigt wird, die wahrscheinlichste ist. Somit soll Ihnen das Programm anhand Ihrer Symptome einen Überblick über mögliche Verdachtsdiagnosen geben. Wenn Ihnen eine Verdachtsdiagnose unbekannt ist, können Sie sich genauer darüber informieren. Klicken Sie dafür auf das Infokästchen neben der Verdachtsdiagnose.

#### Wichtig: Nach der Testung des Programms müssen Sie auf den Feedback-Button klicken!

Dann wird Ihnen ein Fragebogen angezeigt. Es wäre schön, wenn Sie diesen vollständig ausfüllen könnten. Ihre Daten bleiben dabei völlig anonym. Danach haben Sie den Test erfolgreich beendet!

Im Anschluss an Ihre Testung wird der beantwortete Fragebogen von einem Doktoranden ausgewertet. Danach werden Ihre Daten vollständig gelöscht.

Durch Ihre Teilnahme tragen Sie zur Verbesserung des Programms bei!

[Start](#)

Bei weiteren Fragen klicken Sie [hier](#).

Abbildung 31: Online-Evaluation

Abbildung 32: Transparentes Bewertungssystem

### 3 Studienprotokoll

# Studienprotokoll

Verdachtsgenerierungs- und  
Verdachtsüberprüfungsprogramm  
„**SymptomCheck**“

V 3.0, 04.05.2015

#### **Studienleiter:**

Prof. Dr. Stefan Störk

Deutsches Zentrum für Herzinsuffizienz/Medizinische Klinik und  
Poliklinik I

Universitätsklinikum Würzburg

Straubmühlweg 2a

D-97080 Würzburg

## **Inhalt**

1. Einleitung
2. Studienziele und Fragestellungen/Hypothesen
3. Studienbeschreibung
  - 3.1. Studiendesign
  - 3.2. Studienpopulation
  - 3.3. Einschlusskriterien
  - 3.4. Ausschlusskriterien
4. Studienablauf
  - 4.1. Patienten am DZHI
  - 4.2. Studiendauer für die einzelnen Teilnehmer
5. Intendierte (noch zu evaluierende) Leistung des Programms SymptomCheck
6. Datenmanagement / Datenschutz
7. Fallzahlschätzung / Biometrie
8. Nutzen-Risiko-Abwägung
9. Anlagen

## 1. Einleitung

Heutzutage ist das Internet der alltägliche Begleiter der Mehrzahl unserer Bevölkerung. Dabei dienen Suchmaschinen (wie z. B. Google) oftmals auch zu medizinischen Zwecken. Ratsuchende tauschen sich auf den zahlreich verfügbaren medizinischen Portalen und Foren im Internet aus. Die Erwartungshaltung ist, dass durch Eingabe einiger weniger Informationen (häufig: Symptome) eine auf sie zutreffende Diagnose gestellt oder ausgeschlossen werden kann. Diese Internetangebote sind jedoch umstritten: problematisch ist, dass die meisten webbasierten Programme einzelne Symptome bewerten und keine Symptomenkomplexe; eine gewichtete Diagnosestellung ist somit nicht möglich und die Folge sind hohe Raten an falsch positiven und falsch negativen Diagnosen. Zudem neigen Patienten aber auch gesunde Normalpersonen häufig dazu, sich zu überdiagnostizieren, sich also eine Krankheit zusprechen, die vermutlich gar nicht vorliegt. Die Fehlinterpretation harmloser Befundkonstellationen kann so völlig unbegründetes Unbehagen und Angst verursachen. Forscher aus den USA haben dafür einen neuen Begriff geprägt, die „Cyberchondrie“.<sup>1</sup> Viele Menschen neigen auch grundsätzlich zu einer hypochondrischen Einstellung ihrem Gesundheitszustand gegenüber, die durch unkoordinierte Verdachtsgenerierung (z. B. in Gesundheitsforen des Internets) oft noch verstärkt statt vermindert wird.

Aus diesem Grund wurde nun am Lehrstuhl Informatik gemeinsam mit dem DZHI ein Programm, **SymptomCheck**, entwickelt, das sich zum Ziel gesetzt hat, hier objektiver und wertneutraler zu beraten und damit möglicherweise Angst vermeidend zu wirken. **SymptomCheck** ist internetbasiert und besteht aus zwei Komponenten: mit der Verdachtsgenerierungskomponente kommt man zu einem Diagnosevorschlag; mittels der Verdachtsüberprüfungskomponente wird die Diagnose bekräftigt oder entkräftet.

Das hier skizzierte Projekt will die Performance von **SymptomCheck** in einer ersten Pilotstudie charakterisieren. Als Vorläufer von **SymptomCheck** kann

---

<sup>1</sup> Quelle: [http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-suechtig-nach-symptomsuche\\_aid\\_268019.html](http://www.focus.de/gesundheit/ratgeber/psychologie/news/tid-9411/cyberchondrie-suechtig-nach-symptomsuche_aid_268019.html)

**CardioConsult** (erschienen 1998) angesehen werden.<sup>2</sup> Dieses System richtete sich vornehmlich an Ärzte, und weniger an Patienten. **CardioConsult** sollte als Second-Opinion-Programm Ärzte bei der Diagnostik unterstützen. Im Gegensatz dazu richtet sich **SymptomCheck** an die Allgemeinbevölkerung, die Gesundheitsrat suchen. **SymptomCheck** soll effizienter und schneller bearbeitbar sein, da nicht alle Symptome abgefragt werden. Viele Fragen erscheinen nur, wenn man in vorherigen Fragen bestimmte Antworten gegeben hat. Zudem kann **SymptomCheck** nicht nur Diagnosen vorschlagen, sondern auch ausschließen. Durch die graphische Aufarbeitung soll es benutzerfreundlicher und intuitiv zu bedienen sein. Selbstverständlich ersetzt **SymptomCheck** nicht die ärztliche Konsultation, sondern soll dazu beitragen, die „Ressource Arzt“ besser zu nutzen.

## 2. Studienziele und Fragestellungen/Hypothesen

### **Primäres Studienziel:**

Es soll getestet werden, ob **SymptomCheck** Benutzerfreundlichkeit (und somit Alltagstauglichkeit) mit ausführlicher Symptomabfrage und fundierter Verdachtsgenerierung und Verdachtsüberprüfung vereinen kann.

#### 1. Benutzerfreundlichkeit

Sind die Benutzer grundsätzlich in der Lage, **SymptomCheck** zu bedienen? Finden sich die Nutzer gut mit der Suche von für sie zutreffenden Symptomen zurecht und kommen sie mit dem Frage-Stil zurecht? Führt **SymptomCheck** die Benutzer rasch zu den gewünschten Informationen (mögliche Verdachtsdiagnosen, Entkräftigung von Verdachtsdiagnosen)? Erachten die Probanden **SymptomCheck** zur „Selbstdiagnose“ als nützlich?

#### 2. Überprüfung der Sensitivität und Spezifität der Verdachtsgenerierung und Verdachtsüberprüfung

Hilft SymptomCheck Patienten, mögliche Erkrankungen zu erkennen bzw. auszuschließen? Schlägt SymptomCheck richtige Diagnosen vor ohne auf

---

<sup>2</sup> Puppe, B. und Riecker, G.: CardioConsult: Kardiologisches Second-Opinion-Programm, CD-ROM

der anderen Seite ungefährliche Symptome unnötig zu dramatisieren?  
Verbessert die Verdachtsüberprüfung die Diagnosestellung durch den Arzt?

### **Sekundäre Studienziele:**

Es soll untersucht werden, ob die Probanden befürchten krank zu sein und ob **SymptomCheck** diese Befürchtungen lindern kann.

Es soll untersucht werden, über welche Medien sich die Probanden normalerweise über Ihre Symptome oder Erkrankungen informieren (z. B. Internet, Bücher, Apothekenumschau etc.).

Es soll untersucht werden, ob demographische Parameter die obigen Fragestellungen möglicherweise in relevantem Ausmaß beeinflussen (z. B. Alter, Geschlecht, Bildungsgrad).

### **Hypothesen:**

1. **SymptomCheck** wird als benutzerfreundlich empfunden.
2. Die zutreffende Diagnose wird von **SymptomCheck** „erkannt“, d.h. die korrekte Diagnose wird unter den 10 wahrscheinlichsten Diagnosen aufgelistet.
3. Jüngere Personen und Personen mit einem höheren Bildungsgrad kommen mit **SymptomCheck** besser zurecht.

## **3. Studienbeschreibung**

### **3.1. Studiendesign**

Querschnittstudie ohne Kontrollgruppe. Keine Intervention.

### **3.2. Studienpopulation**

**SymptomCheck** soll an folgenden Probanden getestet werden:

- Probanden (n=60) werden in der Ambulanz des Deutschen Zentrums für Herzinsuffizienz in Würzburg rekrutiert. Diese Gruppe setzt sich aus Patienten zusammen, die die ambulante Sprechstunde besuchen und somit an konkreten Symptomen leiden. Dabei handelt es sich primär um

kardiologische Diagnosen, viele werden jedoch auch unter Begleiterkrankungen (z. B. Diabetes mellitus, Hypertonie etc.) leiden. **SymptomCheck** wird hier mittels Tablet-PC / Computer eingesetzt.

### 3.3. Einschlusskriterien

- Volljährigkeit (18 Jahre)
- Schriftliche Einverständniserklärung

### 3.4. Ausschlusskriterien

- Keine

## 4. Studienablauf

### 4.1. Patienten am DZHI

- Geplant ist, 60 Patienten in der Ambulanz des Herzinsuffizienzentrums (DZHI) zu befragen. Die Testung wird noch vor dem Arztbesuch während der Wartezeit erfolgen. Die Patienten werden vorab über den Ablauf dieser Testung anhand einer Patientenaufklärung informiert. Im Anschluss unterzeichnen die Probanden die Einwilligungserklärung zur Teilnahme an der Datenerhebung.
- Vor Testung des Programms füllen die Patienten einen Evaluationsbogen aus, der ihnen in Papierform gegeben wird (Demographie etc.).
- Dann erhalten sie ein Tablet-PC / Computer, mit dem sie das Programm bedienen. Den Probanden wird eine Startseite aufgezeigt, die thematisch in Gruppen geordnet ist. In jeder Gruppe befinden sich Symptomlisten, aus denen die zutreffenden auszuwählen sind. Sollten bestimmte Symptome unklar formuliert sein, bietet das Programm eine Erklärungshilfe, wenn auf den Erklärungsbutton neben dem jeweiligen Symptom gedrückt wird. Sobald alle zutreffenden Symptome vom Probanden ausgewählt wurden, fragt **SymptomCheck** im Anschluss die einzelnen Symptome bezüglich z. B. Intensität, Dauer oder Lokalisation differenzierter ab. Dem Patienten bleibt es jedoch selbst überlassen, wie genau er die Fragen beantworten möchte. Parallel zur Beantwortung wird der Testperson ein Ranking mit möglichen Diagnosen aufgezeigt. Hierbei werden Diagnosen, die mit hoher Wahrscheinlichkeit zutreffen, visuell hervorgehoben.
- Die einzelnen Diagnosen können (nicht obligat) danach in der Verdachtsüberprüfung genauer bearbeitet werden. Hier wird kurz erklärt, um was es sich bei dieser Erkrankung handelt. Es werden spezifische Fragen zu der Verdachtsdiagnose gestellt, die nicht nur Symptome, sondern auch Untersuchungsbefunde und -möglichkeiten einbezieht. Sowohl bei den Symptomen, als auch bei den Untersuchungen wird wieder eine Erklärungshilfe angeboten.

- Nach Bearbeitung des Programms sollen sie die Probanden abschließend einen zweiten Fragebogen ausfüllen. Damit ist die für den Patienten die Erhebung abgeschlossen.
- Das Ergebnis des Diagnoserankings wird im Anschluss mit den in der Realität zutreffenden Diagnosen abgeglichen; diese Diagnosen werden der Patientenakte entnommen.

#### **4.2. Studiendauer für die einzelnen Teilnehmer**

- Die Studiendauer für jeweils einen Teilnehmer variiert stark von der Schnelligkeit und Gründlichkeit der Beantwortung der Fragebögen und von **SymptomCheck**.
- Geschätzt wird eine Dauer von 10-20 Minuten pro Person.

#### **5. Intendierte (noch zu evaluierende) Leistung des Programms SymptomCheck**

- Wahrscheinlichkeitsberechnung diverser Diagnosen anhand der Patientensymptomatik durch Berücksichtigung multipler Symptome
- Möglichkeit zur differenzierten Symptomabfrage trotz schlichter und auf wesentliche Symptome reduzierter Benutzeroberfläche
- Ausschluss möglicher Erkrankungen anhand spezifischer Fragen, anhand deren eine Erkrankung für eher zutreffend oder unwahrscheinlich eingestuft werden kann
- Erhalt von weiteren Informationen zu bestimmten Diagnosen und Untersuchungen

#### **6. Datenmanagement / Datenschutz**

- Vor Beginn der Testung müssen die Probanden nach angemessener Vorabinformation eine Einwilligungserklärung zur Teilnahme an der Studie unterschreiben, die sie ebenfalls als Kopie erhalten. Dadurch bestätigen die Probanden, dass ihre Daten zu Studienzwecken benutzt werden dürfen.
- Im Anschluss erhält die Testperson den Evaluationsbogen, auf dem eine Kenn-Nummer (Stud-PID) vermerkt ist.
- Auf einer Liste in Papierform wird jede Stud-PID einem Patienten mittels seinem Patientenaufkleber zugeordnet und so pseudonymisiert. Diese Liste verbleibt im DZHI und wird spätestens nach Ablauf von 3 Monaten vernichtet, so dass die Informationen dann ausschließlich anonymisiert prozessiert werden.
- Die zugewiesene Stud-PID gibt der Patient bei Bearbeitung des Programms in den Tablet-PC / Computer ein. Die auf dem Tablet-PC / Computer erhobenen

Informationen werden anonym auf dem Server der Fakultät für Informatik in Würzburg<sup>3</sup> unter der Leitung von Herrn Prof. Frank Puppe (Lehrstuhl Informatik VI) gespeichert. Diese Daten werden nur für die Auswertung dieser Studie gespeichert und benutzt. Der Lehrstuhl für Informatik hat keinen Zugang zur Patientenidentifikationsliste, die im DZHI verbleibt.

- Im Anschluss daran wird (durch Doktoranden) das Diagnoseergebnis des Programms mit den Diagnosen der Patientenakte verglichen.
- Dabei werden außer Angaben zu Geschlecht, Alter, Bildungsstand und Diagnosen keine weiteren Informationen verwendet. Weitere Informationen insbesondere die Namen der Probanden bleiben somit pseudonymisiert, nach Ablauf von drei Monaten werden die Daten anonymisiert.

## **7. Fallzahlschätzung / Biometrie**

Es handelt sich um eine Pilotstudie. Eine formale Fallzahlschätzung erfolgt nicht, da zunächst die Benutzerfreundlichkeit von **SymptomCheck** qualitativ erfasst werden soll. Die Ergebnisse sollen als Basis für eine Folgestudie dienen, die die Diagnosesicherheit und die klinische Wertigkeit des Einsatzes eines solchen Systems quantifiziert. Geplant ist die Rekrutierung von n=60 Patienten am DZHI. Die Daten werden rein deskriptiv ausgewertet. Statistische Kenngrößen zur Diagnosesicherheit werden explorativ erhoben, ebenso wie die Testung eines möglichen Einflusses von demographischen Variablen auf die Testergebnisse (Regressionsmodelle).

## **8. Nutzen-Risiko-Abwägung**

- Die Probanden, die an der Studie teilnehmen, haben selbst keinen Nutzen an der Studie. Allerdings tragen sie durch ihre Teilnahme zur Evaluierung und Verbesserung des Programms **SymptomCheck** bei.
- Da die Patienten lediglich einen Evaluationsbogen ausfüllen und das Programm durcharbeiten sollen, gibt es für sie kein gesundheitliches Risiko. Die Patienteninformationen werden pseudonymisiert, so dass auch die Gefahr, dass diese an unberechtigte Personen gelangen, minimiert wird.

---

<sup>3</sup> Institut für Informatik der Universität Würzburg, Adresse: Am Hubland, 97074 Würzburg

## 9. Anlagen

- Patienteninformationsbogen für das DHZI
- Einwilligungserklärung Fragebögen
- UEQ-Fragebogen (User Experience Questionnaire; standardisierter Fragebogen)

## Unterschrift zum Protokoll

Projektleiter

Prof. Stefan Störk

---

Datum

---

Unterschrift

## 4 Fragebögen

Deutsches Zentrum  
für Herzinsuffizienz  
Würzburg



### Prä-Test-Fragebogen

Probandennummer:

Sehr geehrte Damen und Herren,  
im Folgenden wurden einige Fragen zusammengestellt, die für die Auswertung des Tests wichtig sind. Es wäre schön, wenn Sie die Fragen **vollständig und korrekt** beantworten könnten.

Selbstverständlich werden Ihre Antworten **anonym und vertraulich** behandelt.

<b>Alter</b>				
<b>Geschlecht</b>	<input type="checkbox"/>	männlich	<input type="checkbox"/>	weiblich
<b>Schulabschluss</b>	<input type="checkbox"/>	allgemeine oder fachgebundene Hochschulreife	<input type="checkbox"/>	mittlere Reife
	<input type="checkbox"/>	Hauptschulabschluss	<input type="checkbox"/>	keinen Schulabschluss / Andere

	trifft völlig zu						trifft gar nicht zu
1. Ich befürchte krank zu sein.	<input type="checkbox"/>						
2. Ich will testen, ob ich krank bin.	<input type="checkbox"/>						

<b>3. Sind bei Ihnen Vorerkrankungen bekannt?</b>	<input type="checkbox"/>	<b>JA</b>	<input type="checkbox"/>	<b>Ich weiß nicht</b>	<input type="checkbox"/>	<b>NEIN</b>
	WENN JA:					
<b>→ Welche?</b>						





## Post-Test-Fragebogen

Probandennummer:

Sehr geehrte Damen und Herren,  
vielen Dank für die Nutzung des Programms. Zum Abschluss wäre es schön,  
wenn Sie noch ein paar Fragen beantworten könnten. Dabei ist es wichtig, dass  
Sie die Fragen **vollständig und korrekt** beantworten.

Selbstverständlich werden Ihre Antworten anonym und vertraulich behandelt.

**Wir möchten Sie bitten das Programm zu bewerten.** Bitte kreuzen Sie  
**immer eine Antwort an**, auch wenn Sie bei der Einschätzung zu einem  
Begriffspaar unsicher sind oder finden, dass es nicht so gut zum Produkt passt.

### Das Programm ist...

	1	2	3	4	5	6	7		
Unerfreulich	<input type="checkbox"/>	erfreulich	1						
Unverständlich	<input type="checkbox"/>	verständlich	2						
Kreativ	<input type="checkbox"/>	phantasielos	3						
leicht zu lernen	<input type="checkbox"/>	schwer zu lernen	4						
erfrischend	<input type="checkbox"/>	einschläfernd	5						
Langweilig	<input type="checkbox"/>	spannend	6						
Uninteressant	<input type="checkbox"/>	interessant	7						
Unberechenbar	<input type="checkbox"/>	voraussagbar	8						
Schnell	<input type="checkbox"/>	langsam	9						
Neu	<input type="checkbox"/>	alt	10						
Unbedienbar	<input type="checkbox"/>	bedienbar	11						



4. Dieses Programm ist im Alltag sehr nützlich.	<input type="checkbox"/>						
5. Das Programm hat mir besser geholfen als andere Informationsplattformen im Internet	<input type="checkbox"/>						
6. Es wurde mir eine gute Übersicht über mögliche Diagnosen gegeben.	<input type="checkbox"/>						
7. Es ist nützlich, dass mir auch andere Erklärungen für meine Beschwerden vorgeschlagen wurden (z. B. Stress, Bewegungsmangel).	<input type="checkbox"/>						

<b>Die Dauer für die Beantwortung der Fragen war...</b>	<input type="checkbox"/>	<b>viel zu lang</b>	<input type="checkbox"/>	<b>etwas zu lang</b>	<input type="checkbox"/>	<b>genau richtig</b>	<input type="checkbox"/>	<b>etwas zu kurz</b>	<input type="checkbox"/>	<b>viel zu kurz</b>
---	--------------------------	---------------------	--------------------------	----------------------	--------------------------	----------------------	--------------------------	----------------------	--------------------------	---------------------

Ich befürchte krank zu sein trifft völlig zu     trifft garnicht zu

Das Konzept dieses Programms ist eine gute Ergänzung zu den bereits bestehenden Informationsplattformen im Internet trifft völlig zu     trifft garnicht zu

Durch das Programm wurde ich angeregt, mich weiter mit meiner Erkrankung auseinanderzusetzen trifft völlig zu     trifft garnicht zu

**Vielen Dank!**

## **Danksagung**

Ich möchte Herrn Prof. Dr. Störk, und Herrn Prof. Dr. Puppe für die Überlassung des Themas, die kollegiale Zusammenarbeit und die Betreuung meiner Dissertation durch deren Diskussionsbereitschaft und zahlreiche Anregungen herzlich danken.

Mein Dank gilt auch allen, die mich in den vergangenen Jahren unterstützt und diese Arbeit ermöglicht haben.

Ganz besonders danke ich aber meiner Familie, die mir in jeder Lebenssituation zur Seite gestanden hat und mich stets unterstützt.