

### Zusammenfassung

**Zielsetzung:** Entwicklung eines geeigneten Scores zur Evaluation der Qualität von Lungenfunktionsmessungen auf Basis aktueller Empfehlungen. **Material und Methode:** Retrospektive Auswertung randomisiert ausgewählter Messprotokolle von Spirometrien. Die Beurteilungskriterien wurden aus Empfehlungen zur Lungenfunktionsmessung gewonnen. Zum Vergleich erfolgte eine Übersetzung von Einzelergebnissen in einen Score, der zuvor anhand intraindividuelle Verläufe von Lungenfunktionsparametern in einer Gruppe ehemaliger Uranerzbergleute validiert werden konnte. **Ergebnisse:** 417 Messprotokolle von 93 Ärzten unterschiedlicher Fachrichtungen wurden ausgewertet. Von 111 dokumentierten Kalibrierungen war in 10 Fällen diese am Tag der Messung durchgeführt worden. 147 Lungenfunktionskurven wurden als „gut“ eingestuft. Nur auf 25 Messprotokollen war die Durchführung von drei oder mehr Messmanövern dokumentiert. Die Reproduzierbarkeit von Messungen ist daher eingeschränkt. Unter Verwendung des Scores ergab sich über alle Ärzte ein gemittelter Wert von 0,58, was einem mittleren Qualitätsniveau entspricht. **Schlussfolgerungen:** Die Studie generiert und validiert erstmals einen Score, mit dem die Qualität von Spirometrien retrospektiv evaluiert werden kann. Die Anwendung des Scores hat eine inkonsistente Umsetzung von Empfehlungen zur Lungenfunktionsmessung ergeben. In Übereinstimmung mit Ergebnissen internationaler Studien kann auch in Deutschland ein Qualitätsdefizit festgestellt werden.

### Abstract

**Background:** The measurement of quality will gain in importance in the future. Suitable instruments for the measurement of the quality of lung function tests have not yet been developed. In order to evaluate the quality of lung function measurements, a score is introduced, based on current guideline recommendations for lung function test. **Materials and methods:** Records of spirometries were selected at random and evaluated retrospectively. The evaluation criteria were deduced from guideline recommendations. The evaluation criteria were used to calculate a score in order to compare the quality of single records. The individual time course of lung function parameters in a group of former uranium ore miners was used to validate the score. **Results:** 417 spirometry records from 93 physicians of different specialisation were evaluated. From 111 documented calibrations only 10 were accomplished on the same day on which the measurement was made. 147 lung function curves were classified as “good”. Only in 25 records the execution of three or more measuring manoeuvres had been documented. The mean score over all physicians was 0.58, which corresponds to a middle quality level. **Conclusions:** The study generates and validates for the first time a score with which the quality of spirometries can be evaluated. The scoring system retrospectively analyses lung function records and it can therefore be used in the framework of quality management. The application of the score showed an inconsistent application of guideline recommendations. In agreement with studies from other countries a lack of quality was also found in Germany.

### Institutsangaben

Hauptverband der gewerblichen Berufsgenossenschaften

### Korrespondenzadresse

Dr. med. Olaf Hagemeyer · Hauptverband der gewerblichen Berufsgenossenschaften ·  
Alte Heerstr. 111 · 53754 Sankt Augustin · E-mail: olaf.hagemeyer@hvbg.de

**Eingang:** 12. Juni 2003 · **Nach Revision angenommen:** 21. September 2003

### Bibliografie

Pneumologie 2004; 58: 147–154 © Georg Thieme Verlag Stuttgart · New York ·  
ISSN 0934-8387 · DOI 10.1055/s-2004-818376

## Einleitung

Die Messung der Lungenfunktion mittels der Spirometrie gilt als einfach und sensitiv, daher hat sie als Screeninguntersuchung weite Verbreitung gefunden. Diese Methode zur Messung der Lungenfunktion wird aufgrund zunehmender Prävalenz obstruktiver Atemwegserkrankungen zukünftig weiter an Bedeutung gewinnen [1].

Um aussagefähige Spirometriemessungen zu erhalten, wurden seit Anfang der 90er-Jahre von den wissenschaftlichen Fachgremien Empfehlungen zur Durchführung von Lungenfunktionsmessungen unter Qualitätsaspekten publiziert. In Deutschland erarbeiteten 1994 die Deutsche Atemwegsliga und der Hauptverband der gewerblichen Berufsgenossenschaften Empfehlungen [2, 7]. Eine Leitlinie zur Lungenfunktionsmessung mit dem Bearbeitungsstand März 2002 stellt die Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin auf ihrer Homepage vor [3]. Leitlinien sind systematisch entwickelte Darstellungen und Empfehlungen, deren Anwendung rechtlich jedoch der ärztlichen Entscheidungsfreiheit unterliegt. Die Bedeutung von Leitlinien resultiert daraus, dass die Handlungsanleitungen auf dem aktuellen Wissensstand des Fachgebietes basieren.

Alle Empfehlungen zur Lungenfunktionsmessung wurden von Experten entwickelt und repräsentieren das Stadium S1 nach der Empfehlung der Arbeitsgemeinschaft der medizinischen Fachgesellschaften (AMWF). Ob und in welchem Umfang Empfehlungen zur Lungenfunktionsmessung in der praktischen ärztlichen Tätigkeit umgesetzt werden, ist bislang nur vereinzelt untersucht worden. In diesen wenigen Studien aus dem Ausland musste übereinstimmend eine ungenügende Berücksichtigung von Qualitätskriterien festgestellt werden. Den Untersuchungen lagen unterschiedliche Studienansätze zugrunde, wobei neben alleinigen Befragungen [5, 14, 15] auch Messungen an Testpersonen durchgeführt wurden [4, 8]. Eine Interventionsstudie konnte darüber hinaus den positiven Effekt von Trainingsmaßnahmen belegen [6].

Die Qualität von Lungenfunktionsmessungen wurde nach unserer Kenntnis erst einmal anhand einer Beurteilung von Messprotokollen evaluiert [16]. Stoller u. Mitarb. (1997) haben Messungen ausgewertet, die durch speziell trainiertes Personal bei lungenfunktionsanalytisch erfahrenen Probanden gewonnen worden waren und denen die spätere Qualitätsevaluation bekannt war. Die Qualität in der alltäglichen Praxisroutine angefertigter Spirogramme wurde bislang nicht untersucht.

Eine denkbare Ursache für das Fehlen entsprechender Evaluationen könnte die Komplexität der Beschreibung von „Qualität eines Spirogrammes“ sein. Da Fehler bei der Durchführung einer Spirometrie auf verschiedenen Stufen des Handlungsprozesses auftreten können, kann die aufgezeichnete Messkurve auch unterschiedliche Fehler aufweisen, was die Gesamtbeurteilung erschwert.

Als mathematische Hilfsmittel zur Beschreibung derartiger komplexer „Qualität“ eignen sich Scores, die verschiedene Beurteilungskriterien berücksichtigen und in Form eines zusammenfassenden Zahlenwertes beschreiben können [10]. Aussagekräftig

sind Scores allerdings nur dann, wenn sie hinreichend validiert worden sind.

Im Folgenden werden die Ergebnisse einer Studie vorgestellt, welche die Umsetzung von Empfehlungen zur Lungenfunktionsmessung anhand von Messprotokollen auswertet und unter Verwendung eines Scores eine Aussage über die Qualität von Lungenfunktionsmessungen macht.

## Material und Methode

### Studienansatz

Es handelt sich um eine retrospektive Studie, bei der in der alltäglichen Praxis angefertigte Lungenfunktionsmessungen mit Qualitätskriterien bewertet wurden. Die Erfüllung einzelner Qualitätskriterien wurde in einem zuvor validierten Score erfasst. Die Summe der erreichten Scorepunkte ermöglicht eine Aussage über die Qualität der Lungenfunktionsmessung.

### Probanden

Aktuelle Lungenfunktionsmessprotokolle und die für die Auswertung erforderlichen Ergebnisse früherer Messungen konnten aus einem Kollektiv ehemaliger Bergleute des DDR-Uranerzbergbaus erhalten werden. Für diese Bergleute besteht seit 1992 die Möglichkeit, an regelmäßigen Untersuchungen zur Krankheitsfrüherkennung teilzunehmen [13]. Die Ergebnisse der Untersuchungen werden auf vorgegebenen Untersuchungsbogen durch den Arzt schriftlich fixiert und in anonymisierter Form zentral für wissenschaftliche Zwecke erfasst.

Die medizinischen Untersuchungen beinhalten immer eine Spirometrie. In den Untersuchungsbogen werden die Werte der inspiratorischen Vitalkapazität (IVC), der 1-Sekunden-Kapazität (FEV<sub>1</sub>) und des prozentualen Verhältnisses der beiden Messwerte (FEV<sub>1</sub> % IVC) notiert. Zusätzlich werden auch die anamnestischen Angaben zu früheren Erkrankungen, aktuellen Beschwerden, die Befunde der körperlichen Untersuchung, Ergebnisse von Laboruntersuchungen und die Befunde einer Röntgenuntersuchung des Thorax kodiert. Frühere sowie aktuelle Erkrankungen des Probanden klassifiziert der Arzt nach ICD-9.

Originalaufzeichnungen von diagnostischen Untersuchungen wie z.B. der Lungenfunktionsmessung oder auch die angefertigten Röntgenbilder verbleiben dagegen beim untersuchenden Arzt und werden nicht zentral erfasst.

### Ärzte

Mit den Untersuchungen beauftragt sind Ärzte verschiedener Fachrichtungen (Pneumologen, Arbeitsmediziner, Internisten/Betriebsärzte), die an unterschiedlichen Institutionen arbeiten (Ärzte an Instituten, Krankenhäusern, im öffentlichen Gesundheitsdienst und bei überbetrieblichen arbeitsmedizinischen Diensten sowie niedergelassene Ärzte).

### Lungenfunktionsprotokolle

Für diese Studie wurden von Ärzten ergänzend zu den Untersuchungsbogen Kopien der Aufzeichnungen der Lungenfunktionsmessprotokolle erbeten. Von jedem Arzt wurden zwischen drei und fünf Messkurven erbeten. Die Auswahl der Lungenfunk-

Tab. 1 Ableitung der Qualitätskriterien zur Beurteilung der Lungenfunktionsmessungen aus verschiedenen Leitlinien. „Dokumentierter“ Messwert: Messwert, der bei der nachgehenden Untersuchung an die zentral erfassenden Stellen mitgeteilt wurde

	HVBG (1994)	DAL (1994)	DGAUM (2002)	Score-Parameter
Messparameter: IVC, FVC, FEV <sub>1</sub> , PEF, MEF <sub>72-25</sub>	+	+	+	
berücksichtigt werden die Maximalwerte	+	+	+	dokumentierter Messwert IVC ist der Maximalwert von VC dokumentierter Messwert der IVC > dokumentierter Messwert FEV <sub>1</sub>
Ausdruck von drei Messkurven	+	+	+	Anzahl der Messkurven >= 3
Darstellung der „Hüllkurve“			wird nicht empfohlen	
Referenzwertformel muss erkennbar sein (volumetrische) Kalibrierung	+	+	+	Zeitpunkt der Kalibrierung in den letzten 7 Tagen vor Messzeitpunkt
Kurvenform	+	+	+	Kurvenform ist gut
Beginn des Messmanövers			+	rückextrapoliertes Volumen < 5% FEV <sub>1</sub> oder 0,15 L oder Zeit < 120 ms
Messwertabweichung muss erkennbar sein (VC, FEV <sub>1</sub> < 5%)	+			Messwertabweichung VC < 5% (nicht berücksichtigt)
				Messwertabweichung FEV <sub>1</sub> < 5% (nicht berücksichtigt)
Messwertabweichung muss erkennbar sein (FVC, FEV <sub>1</sub> < 0,2 L)			+	Messwertabweichung FEV <sub>1</sub> < 0,2 L
			+	Messwertabweichung VC < 0,2 L
Ende des Atemmanövers			+	Dauer des Atemmanövers mindestens 6 s oder bis Plateau
Alter				keine Abweichung bei Altersangabe
Größe				keine Abweichung bei Größenangabe
Gewicht				keine Abweichung bei Gewichtsangabe
Obstruktion				aufgrund der Messdokumentation (Kurve, Messwerte) ist Aussage über Atemwegsobstruktion möglich
Beurteilung des Messung				Messung enthält keine groben Auffälligkeiten, weshalb sie für eine wissenschaftliche Auswertung geeignet wäre

HVBG: Hauptverband der gewerblichen Berufsgenossenschaften; DAL: Deutsche Atemwegsliga; DGAUM: Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin

tionsmessprotokolle erfolgte dabei randomisiert. Einschlusskriterium war, dass die Untersuchung in den letzten sechs Monaten vor Beginn dieser Studie im Mai 2002 erfolgt war.

Insgesamt wurden 485 Messprotokolle von 103 Ärzten erbeten.

### Qualitätskriterien

Zur Beurteilung der Messqualität der Spirometrien wurden Kriterien aus drei Empfehlungen [2, 3, 7] zur Lungenfunktionsmessung hergeleitet. Die Kriterien sind in Tab. 1 dargestellt. Die Benutzung dieser Empfehlungen erfolgte vor dem Hintergrund, dass die Anforderungen an eine gute Lungenfunktionsmessung in den drei Empfehlungen unterschiedlich beschrieben werden.

Alle Protokolle von Lungenfunktionsmessungen wurden dann im Hinblick auf die Umsetzung dieser Kriterien durch einen lungenfunktionsanalytisch erfahrenen Arzt beurteilt und die Umsetzung der einzelnen Kriterien dokumentiert.

Beurteilt wurde auch, ob anhand der Spirogramme eine Aussage über das Vorliegen einer obstruktiven Ventilationsstörung gemacht werden konnte und wie groß die Aussagefähigkeit der Messkurven war. Diesbezüglich wurde insbesondere auf Artefakte, Hinweise auf Husten, die Darstellung eines erkennbaren Peaks bei der Ausatmung sowie auf einen schnellen Beginn der Einatmungs- und auf eine hinreichend lange Ausatmungsphase geachtet. Ergänzend wurden das auf den Messprotokollen verzeichnete Alter, die Größe und das Gewicht erfasst (Tab. 1).

### Score

Um die Qualität von Lungenfunktionsmessungen anschaulich darstellen zu können, wurde ein Score gebildet. Hierzu wurde für die Umsetzung eines jeden Qualitätskriteriums jeweils ein Punkt vergeben, die Summe der Punkte addiert und die Gesamtpunktzahl durch die Zahl der beurteilbaren Kriterien dividiert.

Da das Ziel die Beurteilung der Messqualität von Ärzten war, wurde aus den Scores aller Messungen eines Arztes der durchschnittlich erreichte Wert bestimmt.

## Validierung des Scores

Zur Generierung eines Scores gehört der Nachweis seiner Validität. Diese ist dann anzunehmen, wenn mittels einer statistischen Analyse die Korrelation zwischen dem Score und der Zielgröße „gute Lungenfunktionsmessung“ belegt werden kann [10,18]. Als gute Lungenfunktionsmessung wird eine Messung definiert, wenn sie einen guten Maßstab für die ventilatorische Leistungsfähigkeit des Probanden darstellt.

Als Maßstab für die ventilatorische Leistungsfähigkeit standen die Messwerte von IVC und FEV<sub>1</sub> zur Verfügung. Die Messwerte waren einerseits aus den Spirogrammen ersichtlich, andererseits waren aber auch schon aus vorherigen Untersuchungen Messwerte von IVC und FEV<sub>1</sub> zentral erfasst. Unter der Annahme, dass bei einem lungengesunden Probanden lediglich ein geringer jährlicher Abfall der 1-Sekunden-Kapazität von etwa 30 bis 50 ml im Rahmen des physiologischen Alterungsprozesses zu erwarten ist, sollten bei validen Lungenfunktionsmessungen auch im zeitlichen Verlauf nur geringe Schwankungen der Messwerte zu erkennen sein [3].

Es bot sich daher an, die schon vorhandenen Messwerte miteinander zu vergleichen und die größten Werte zu bestimmen. Diese maximal gemessenen Werte von IVC und FEV<sub>1</sub> dienten als Referenzwert. Die Differenz zwischen den Referenzwerten und den Werten auf den übersandten spirometrischen Messprotokollen ergab dann den gesuchten „Qualitätsindikator“. Der Wert wurde als „D-max“ bezeichnet.

Eine gute Lungenfunktionsmessung entspricht damit einem kleinen Wert von D-max.

Zum Ausschluss krankheitsbedingter Schwankungen der Messwerte von IVC und FEV<sub>1</sub> wurden für die Validierung des Scores anhand der nach ICD-9 verschlüsselten Diagnosen alle Probanden ausgeschlossen, bei denen jemals eine Atemwegs- oder Lungenerkrankung diagnostiziert worden war (keine ICD-9: 160–165, 460–519). Die Berechnung von D-max wurde nur bei Probanden durchgeführt, bei denen mindestens drei Lungenfunktionsmessungen zentral gespeichert waren. Weiterhin wurden in die Validierung nur Messungen einbezogen, bei denen mindestens fünf Scorekriterien beurteilt werden konnten. Dadurch wurden Messungen von der Validierung ausgeschlossen, für die der Score aufgrund schlechter Dokumentation des Messvorganges nicht zuverlässig berechnet werden konnte.

Die Summe der Scorepunkte wurde mit D-max korreliert.

## Statistische Analyse

Die Korrelation von D-max und Score wurde mittels Pearson-Test auf Signifikanz getestet.

## Ethik

Die Vorschriften des Datenschutzes sind bei der Durchführung der Studie eingehalten worden.

Die Untersuchung wurde durch den Hauptverband der gewerblichen Berufsgenossenschaften finanziert.

## Ergebnisse

### Beachtung von Qualitätskriterien

Ausgewertet wurden 417 spirometrische Messkurven von 93 Ärzten, das bedeutet, dass 90% der angeschriebenen Ärzte Messkurven übersandt haben. Wegen unterschiedlicher Kurvendarstellung (Fluss-Volumen- oder Volumen-Zeit-Kurven) und unterschiedlicher Dokumentation von weiteren Parametern (Kalibrierungsdatum, Nennung der Referenzwertformel, von Alter, Größe und Gewicht) konnten nicht immer alle Qualitätskriterien Berücksichtigung finden. In 9 Fällen lag keine dokumentierte grafische Aufzeichnung der Messkurve vor, hier waren die Messungen lediglich in Form von Zahlenwerten dokumentiert.

Bei Betrachtung der einzelnen Qualitätskriterien fällt insbesondere auf, dass nur bei 25 Messungen wenigstens drei Messmanöver dokumentiert wurden. Aussagen über die Reproduzierbarkeit waren daher nur in wenigen Fällen möglich (Tab. 2).

Zu selten wurde auch eine am Messtag durchgeführte Kalibrierung dokumentiert (2,5%). Bei 111 Fällen ausgewerteten Messungen erfolgte die Kalibrierung am gleichen Tag in 10 Fällen und innerhalb eines Monats vor der Messung in 30 Fällen. Der maximale Abstand zum Messdatum betrug 465 Tage.

Die Kurvenform wurde retrospektiv nur in 147 Fällen als „gut“ bewertet.

Die Umsetzung aller in den Empfehlungen geforderten Qualitätskriterien konnte bei Benutzung der Empfehlungen der Deutschen Atemwegsliga bei 25 von 417 Lungenfunktionsmessungen, bei Benutzung der Empfehlungen des HVBG bei 17 Messungen und bei Benutzung der Leitlinie der DGAUM bei einer Messung bestätigt werden.

Wesentliche Folge der Nichtbeachtung von Qualitätskriterien ist eine nur sehr begrenzte Validität der Messungen. Hieraus resultiert eine erhebliche Schwankungsbreite von Lungenfunktionsmesswerten im zeitlichen Verlauf bei Lungengesunden (Abb. 1).

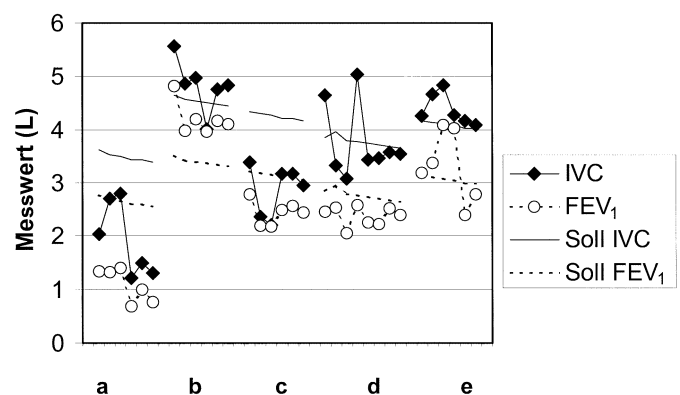


Abb. 1 Ergebnisse von Lungenfunktionsmessungen bei 5 Probanden (a–e) im zeitlichen Verlauf. Aufgetragen wurden die einzelnen Messwerte von IVC und FEV<sub>1</sub> sowie die zugehörigen mittleren VC und FEV<sub>1</sub>-Sollwerte (Ordinate: Mess- und Sollwerte in Litern, Abszisse: Zeit).

Tab. 2 Grad der Erfüllung des Kriteriums. „Dokumentierter“ Messwert: Messwert, der bei der nachgehenden Untersuchung an die zentral erfassende Stelle mitgeteilt wurde.

	<b>Scoreparameter</b>	<b>Anzahl ausgewerteter Messungen</b>	<b>Erfüllung (n = Anzahl; Prozent Angaben beziehen sich auf alle auswertbaren Kurven)</b>
<i>Anwendung von ATS-Kriterien</i>	Anwendung ist auf dem Ausdruck erkennbar: 1 Punkt	417	4 (1,0%)
<i>Messparameter: IVC, FVC, FEV<sub>1</sub>, PEF, MEF<sub>72-25</sub></i>			nicht ausgewertet
<i>berücksichtigt werden die Maximalwerte</i>	dokumentierte Messwerte, IVC ist der Maximalwert von VC: 1 Punkt	417	270 (64,7%)
	dokumentierter Messwert der IVC > dokumentierter Messwert FEV <sub>1</sub> : 1 Punkt	417	407 (97,6%)
<i>Ausdruck von drei Messkurven</i>	Anzahl der Messkurven >= 3: 1 Punkt	417	25 (6,0%)
<i>Darstellung der „Hüllkurve“</i>			nicht ausgewertet
<i>Referenzwertformel muss erkennbar sein (volumetrische) Kalibrierung</i>		417	239 (58,7%)
	Zeitpunkt der Kalibrierung in den letzten 7 Tagen vor Messzeitpunkt: 1 Punkt	113	30 (26,5%)
<i>Kurvenform</i>	Kurvenform ist gut 1 Punkt	408 <sup>1)</sup>	147 (36,0)
<i>Beginn des Messmanövers</i>	rückextrapoliertes Volumen < 5% FEV <sub>1</sub> oder 0,15 L oder Zeit < 120 ms: 1 Punkt	261	187 (71,6%)
<i>Messwertabweichung muss erkennbar sein (VC, FEV<sub>1</sub> &lt; 5%)</i>	Messwertabweichung VC < 5% (nicht berücksichtigt)	54	35 (64,8%)
	Messwertabweichung FEV <sub>1</sub> < 5% (nicht berücksichtigt)		
<i>Messwertabweichung muss erkennbar sein (FVC, FEV<sub>1</sub> &lt; 0,2 L)</i>	Messwertabweichung FEV <sub>1</sub> < 0,2 L: 1 Punkt	23	14 (61,9%)
	Messwertabweichung VC < 0,2 L: 1 Punkt	23	10 (43,5%)
<i>Ende des Atemmanövers</i>	Dauer des Atemmanövers mindestens 6 Sek. oder bis Plateau: 1 Punkt	253	119 (47%)
<i>Alter</i>	keine Abweichung bei Altersangabe: 1 Punkt	325	321 (98,8%)
<i>Größe</i>	keine Abweichung bei Größenangabe: 1 Punkt	318	306 (96,2%)
<i>Gewicht</i>	keine Abweichung bei Gewichtsangabe: 1 Punkt	161	152 (94,4%)
<i>Obstruktion</i>	aufgrund der Messdokumentation (Kurve + Messwerte) ist Aussage über Atemwegsobstruktion möglich: 1 Punkt	417	268 (64,4%)
<i>Beurteilung der Messung</i>	Messung enthält keine groben Auffälligkeiten, weshalb sie für eine wissenschaftliche Auswertung geeignet wäre: 1 Punkt	417	38 (9,1%)

<sup>1)</sup> In den übrigen Fällen (n = 9) lag keine dokumentierte analoge Aufzeichnung der Lungenfunktionsmesskurve vor  
 HVBG: Hauptverband der gewerblichen Berufsgenossenschaften  
 DAL: Deutsche Atemwegsliga  
 DGAUM: Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin

## Score

Abb. 2 zeigt die Häufigkeitsverteilung des Scores bei den Ärzten, welche Messprotokolle zur Auswertung zur Verfügung gestellt haben. Über alle Ärzte gemittelt errechnet sich ein Score von 0,52 (Standardabweichung: 0,10). Der maximal erreichbare Wert wäre 1,0.

Die Auswertung des Scores zeigt somit einen mittleren Wert, was mit der Beurteilung der Umsetzung einzelner Qualitätskriterien übereinstimmt.

## Validierung des Scores

Zur Validierung des Scores konnten unter den im Methodikteil beschriebenen Voraussetzungen lediglich 18 Einzelmessungen von Lungengesunden herangezogen werden. Die Berechnung der Korrelation zwischen dem Score und D-max ergab einen Wert von  $-0,629$  ( $p = 0,01$ ) (Abb. 3). Hohe Punktzahlen im Score korrelieren somit statistisch signifikant mit einer hohen Reproduzierbarkeit von Messergebnissen im zeitlichen Verlauf.

## Diskussion

Die Verbesserung der Effizienz medizinischer Leistungen setzt die Bestimmung der „Qualität“ medizinischer Leistungen voraus [18]. Scores sind zwar vornehmlich aus der prognostischen Bewertung von klinischen Befunden bekannt [11,12], sie können aber auch im Rahmen der Qualitätssicherung eingesetzt werden, wenn deskriptive Daten zusammengefasst und verglichen werden sollen [10].

Stoller u. Mitarb. 1997 haben Scores zur Beurteilung der Qualität von Lungenfunktionsmesskurven eingesetzt [16]. Diese Arbeitsgruppe hatte Scores aus Leitlinien hergeleitet. In dieser prospektiven Studie waren den beteiligten Untersuchern die spätere Qualitätsevaluation vor der Durchführung der Messungen bekannt. Die Untersucher waren darüber hinaus speziell geschult worden. Bei den Probanden handelte es sich um Emphysepatienten mit gesichertem  $\alpha$ 1-Antitrypsinmangel, bei denen gute Erfahrung mit der Lungenfunktionsanalyse unterstellt werden kann. Im Gegensatz zu dieser Arbeitsgruppe wurden von uns Spirogramme retrospektiv ausgewertet, wobei den Untersuchern zum Zeitpunkt der Durchführung der Untersuchung die spätere Verwendung der Messprotokolle im Rahmen einer Evaluation nicht bekannt war. Das hier vorgelegte zweischrittige Studiendesign (Score-Generierung, Score-Anwendung) beschreibt daher die Qualität der Lungenfunktionsmessung in der alltäglichen Routine und nicht unter besonders gestalteten Studienbedingungen.

Zur Generierung von Scores sind zwei Ansätze möglich: Einerseits können die Kriterien durch Experten ausgewählt und hinsichtlich ihrer Bedeutung gewichtet werden, andererseits kann eine multivariate Analyse zur Erkennung der einzelnen Scorebestandteile und ihrer Gewichtung erfolgen. Bei beiden Ansätzen kann auf eine Validierung des entwickelten Scores mit einem möglichst unabhängigen Datensatz nicht verzichtet werden [9].

In unserem Fall lag Expertenwissen in Form verschiedener Empfehlungen zur Durchführung von Lungenfunktionsmessungen vor [2,3,7], woraus die einzelnen Qualitätskriterien abgeleitet

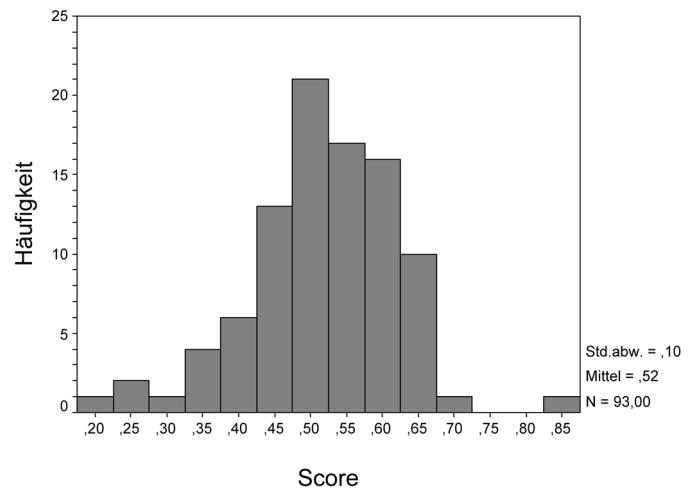


Abb. 2 Häufigkeitsverteilung des Scores.

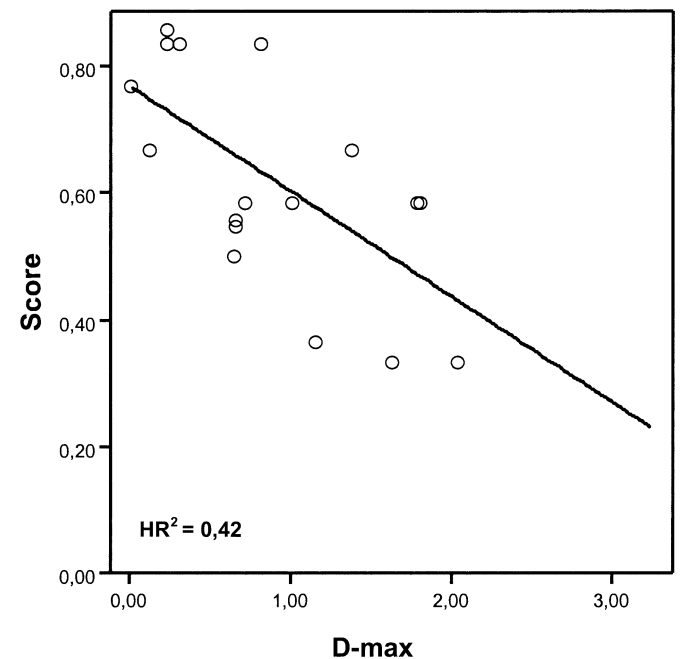


Abb. 3 Korrelation zwischen Score und D-max.

werden konnten. Ergänzend wurden weitere Faktoren berücksichtigt, die indirekt einen Hinweis auf die Struktur- und Prozessqualität bei der Durchführung der Lungenfunktionsmessung geben. Interessant erschien z.B. die Frage, ob die Größe aktuell vor der Lungenfunktionsmessung gemessen oder der Proband nur nach seiner Größe gefragt wird, da die korrekte Berücksichtigung von Alter und Größe für die Berechnung der Sollwerte von Bedeutung ist. In einem geringen Prozentsatz mussten bei diesen Daten Unterschiede zwischen den Befunden der körperlichen Untersuchung und den für die Berechnung der Lungenfunktionsreferenzwerte festgestellt werden, was den Rückschluss auf eine nicht optimale Prozessqualität in den entsprechenden Untersuchungszentren zulässt.

Probleme bereitete die Definition des Zielparameters „gute Lungenfunktionsmessung“. Während vielfach Scores eine Aussage über einen eindeutig definierten Zielparameter ermöglichen



Tab. 3 Übersicht über Studien, in denen die Qualität von Lungenfunktionsmessungen evaluiert worden ist

Quelle	Methode	Ergebnis	Anmerkung
Mushtaq u. Mitarb. 1995	<ul style="list-style-type: none"> <li>– Messung von 3 Probanden in 22 Lungenfunktionslaboratorien</li> <li>– ergänzender Fragebogen</li> <li>– Berechnung von Referenzwerten aus vorgegebenen Daten in allen 22 Laboratorien</li> </ul>	<ul style="list-style-type: none"> <li>– signifikante Unterschiede der Messergebnisse der drei Probanden (<math>p &lt; 0,05</math>)</li> <li>– signifikante Unterschiede bei der Berechnung der Referenzwerte (<math>p &lt; 0,05</math>)</li> <li>– Unterschiede in der Durchführung der Messung</li> </ul>	
Dowson u. Mitarb. 1998	<ul style="list-style-type: none"> <li>– Messung der gleichen 3 Probanden in 22 Lungenfunktionslaboratorien</li> <li>– Berechnung von Referenzwerten aus vorgegebenen Daten in allen 22 Laboratorien</li> </ul>	<ul style="list-style-type: none"> <li>– bessere Messung von FEV<sub>1</sub> und VC</li> <li>– weiterhin signifikante Messwertunterschiede</li> </ul>	<ul style="list-style-type: none"> <li>– erneute Evaluation der 22 Laboratorien nach Beratung und Training der Untersucher</li> <li>– wegen des unzufriedenstellenden Ergebnisses geben die Autoren Empfehlung zum regelmäßigen Audit</li> </ul>
Stoller u. Mitarb. 1997	<ul style="list-style-type: none"> <li>– 1129 Patienten mit <math>\alpha</math>1-Antitrypsinmangel</li> <li>– 37 Untersuchungszentren</li> <li>– Spirometrie vor und nach Bronchodilatation</li> <li>– Untersucher und Probanden waren über Teilnahme an Evaluation informiert</li> <li>– Training für Untersucher</li> <li>– ATS-Kriterien wurde eingehalten</li> </ul>	<ul style="list-style-type: none"> <li>– Reproduzierbarkeit von FEV<sub>1</sub> 95 %</li> <li>– 48 % der Zentren konnten in 90 % der Fälle reproduzierbare Messkurven darstellen</li> <li>– Hauptproblem: Erreichen des Kriteriums für Testende</li> </ul>	<ul style="list-style-type: none"> <li>– die Studie wurde unter besonders günstigen Umständen durchgeführt</li> <li>– die Studie belegt die gute Reproduzierbarkeit auch spirometrischer Messungen</li> <li>– keine Aussage über Langzeitverlauf der Lungenfunktion möglich</li> </ul>
Schuldheisz u. Mitarb. 1998	<ul style="list-style-type: none"> <li>– Einsatz eines Fragebogens</li> <li>– 200 zufällig ausgewählte Mitglieder der American Thoracic Society</li> </ul>	<ul style="list-style-type: none"> <li>– Rücklauf: 102 Antworten</li> <li>– eine „signifikante“ Minderheit weicht von den quantitativen Kriterien der Akzeptabilität von Lungenfunktionsmesskurven ab</li> </ul>	
Dowson u. Mitarb. 1999	<ul style="list-style-type: none"> <li>– Telefoninterview</li> <li>– 95 Praxen</li> <li>– 224 Allgemeinärzte</li> </ul>	<ul style="list-style-type: none"> <li>– nur 2 (von 10) Untersuchern waren trainiert worden</li> <li>– von 3 Untersuchern wusste keiner, wie ihr Pneumotachograph zu kalibrieren ist</li> </ul>	<ul style="list-style-type: none"> <li>– Autoren schließen auf eine geringe Validität der Spirometrien</li> </ul>
Eaton u. Mitarb. 1999	<ul style="list-style-type: none"> <li>– randomisierte prospektive Studie</li> <li>– 30 hausärztliche Praxen</li> <li>– Gruppe 1: Arzt und Assistentin wurden trainiert;</li> <li>– Gruppe 2: kein Training</li> <li>– Auswertung von Spirometrien nach ATS-Kriterien</li> </ul>	<ul style="list-style-type: none"> <li>– Akzeptabilität und Reproduzierbarkeit signifikant besser in trainierter Gruppe (<math>p &lt; 0,0001</math>)</li> <li>– Trainingseffekt wurde belegt</li> <li>– insgesamt aber noch Mängel in der Umsetzung der ATS-Kriterien, insbesondere Testende-Kriterien</li> </ul>	
Valentini u. Mitarb. 2001	<ul style="list-style-type: none"> <li>– 734 Lungenfunktionsmessungen</li> <li>– Auswertung anhand von Qualitätskriterien</li> </ul>	<ul style="list-style-type: none"> <li>– zu niedrige Messung von FEV<sub>1</sub> in 64,2% und von FVC in 85,8%</li> <li>– 12,3% der Spirometrien wurden von den Untersuchern fälschlich auffällig bezeichnet</li> </ul>	
Piirilä u. Mitarb. 2002	<ul style="list-style-type: none"> <li>– Fragebogen</li> <li>– Umsetzung in Score</li> </ul>	<ul style="list-style-type: none"> <li>– n = 637 Bogen; davon 203 vollständig ausgefüllt</li> <li>– mittlerer Score = 67 (maximal: 100)</li> <li>– Hauptprobleme: Vorbereitung des Patienten, Durchführung der Messung, Erkennen einer auswertbaren Messung, Kontrolle von Akzeptabilität und Reproduzierbarkeit</li> </ul>	<ul style="list-style-type: none"> <li>– Autoren stellen eine Qualitätssteigerung im Vergleich zu einer 10 Jahre früheren Studie fest</li> <li>– sie empfehlen regelmäßiges Training für Ärzte und Assistenzpersonal</li> </ul>

(z. B.: Mortalität unter Behandlung], ist die Definition einer „guten“ Lungenfunktionsmessung vergleichsweise schwierig. Die Anwendung von Referenzwertformeln und der Vergleich zwischen Sollwerten und Messwerten kommen wegen individuell hochnormaler Lungenfunktionsmesswerte nicht in Betracht. In unserem Fall konnte stattdessen auf mehrere Lungenfunktionsmessungen aus vergangenen Jahren zurückgegriffen werden, weshalb sich ein intraindividueller Vergleich der Messdaten anbot. Aus dem Vergleich zwischen den maximal gemessenen Werten von IVC und FEV<sub>1</sub> und den Messwerten aus dem aktuell vorliegenden Spirogramm konnte der Zielparameter „D-max“ gewonnen werden.

Dieses Vorgehen impliziert als Fehler die mit dem physiologischen Alterungsprozess erklärable Reduktion der Volumina von bis zu 50 ml pro Jahr. Bei unseren Auswertungen waren die Messschwankungen aber so groß, dass die physiologische Volumenreduktion demgegenüber nicht ins Gewicht fiel.

Als wesentliches Ergebnis der Studie konnte ein Score auf Grundlage der Leitlinien generiert werden. Die Validität des Scores wurde statistisch nachgewiesen.

Die Auswertung des Scores, aber auch von einzelnen Qualitätskriterien, zeigt, dass Verbesserungsmöglichkeiten der Messqualität gegeben sind. Diese betreffen insbesondere die Zahl durch-

geführten und dokumentierter Messmanöver, aus der eine fehlende Kenntnis über die tatsächlichen Maximalwerte der gemessenen Parameter resultiert. Die im zeitlichen Verlauf deutlichen Schwankungen der Lungenfunktionsmesswerte werden hierdurch erklärt (Abb. 1). Unbefriedigend ist auch die sehr geringe Zahl von nur 10 Messungen, die mit Lungenfunktionsmessgeräten durchgeführt wurden, die auch am gleichen Tag geeicht worden waren.

Untersuchungen zur Qualität von Lungenfunktionsmessungen liegen aus Deutschland bisher nicht vor. Ausländische Studien wurden nur vereinzelt durchgeführt [4–6, 8, 14–17]. Diese Studien basieren auf unterschiedlichen methodische Ansätzen (u. a. Messung von Testpersonen, Befragungen der Ärzte). Trotz unterschiedlicher Vorgehensweise mussten in allen uns bekannten Studien jedoch übereinstimmend Qualitätsdefizite festgestellt werden (Tab. 3).

Der Vorteil eines Scores liegt in der Möglichkeit einer Vergleichbarkeit mit zukünftigen Untersuchungen. Aufgrund der Literatur wurde ein ähnlicher Evaluationsansatz unter Verwendung eines Scores von Piirilä u. Mitarb. (2002) gewählt [14]. Abweichend stützt sich diese Arbeitsgruppe aber auf eine Befragung der Ärzte und zieht andere Parameter mit in den Score ein. Lungenfunktionsmessprotokolle beurteilen sie in der hier präsentierten Weise im Sinne einer Ergebnisqualität nicht, vielmehr beurteilen sie Struktur- und Prozessqualität. Daher ist auch das Ergebnis nicht vergleichbar.

Insgesamt wird somit ein Ansatz für eine Evaluation von Lungenfunktionsmesskurven vorgestellt. Es kann aufgrund dieser erstmaligen Evaluation zur Qualität von spirometrischen Lungenfunktionsmessungen ein Handlungsbedarf zur Verbesserung der Qualität abgeleitet werden. Vordringlich scheint dabei eine intensiviertere Sensibilisierung der Ärzte. Unzureichende Qualität bedeutet, dass Erkrankungen falsch oder zu spät diagnostiziert werden. Hieraus resultieren nicht erforderliche Behandlungsmaßnahmen mit ineffektiver Nutzung der Ressourcen des Gesundheitssystems und vermeidbare psychische Belastungen der vermeintlichen Patienten. Auch zukünftig sind daher Beurteilungen der Lungenfunktionsqualität erforderlich. Hierzu erscheint die Anwendung des Scores auch in prospektiven Untersuchungen sinnvoll.

## Literatur

- <sup>1</sup> Chan-Yeung M. Spirometry and tests of bronchial hyperresponsiveness in population studies. *Int J Tuberc Lung Dis* 2000; 4: 633–638
- <sup>2</sup> Deutsche Atemwegsliga. Durchführung von Lungenfunktionsprüfungen in der Praxis. *Pneumologie* 1994; 48: 292–295
- <sup>3</sup> Deutsche Gesellschaft für Arbeits- und Umweltmedizin. Leitlinie: Lungenfunktionsprüfungen in der Arbeitsmedizin. [www.dgaum.de](http://www.dgaum.de) 2002
- <sup>4</sup> Dowson LJ, Mushtaq M, Watts T et al. A re-audit of pulmonary function laboratories in the West Midlands. *Respir Med* 1998; 92: 1155–1162
- <sup>5</sup> Dowson LJ, Yeung A, Allen MB. General practice spirometry in North Staffordshire. *Monaldi Arch Chest Dis* 1999; 54: 186–188
- <sup>6</sup> Eaton T, Withy S, Garrett JE et al. Spirometry in primary care practice: the importance of quality assurance and the impact of spirometry workshops. *Chest* 1999; 116: 416–423
- <sup>7</sup> Hauptverband der gewerblichen Berufsgenossenschaften. Anhang 1: Leitfaden für die LUNGENFUNKTIONSPRÜFUNG bei arbeitsmedizinischen Vorsorgeuntersuchungen nach Berufsgenossenschaftlichen Grundsätzen. In: Hauptverband der gewerblichen Berufsgenossenschaften (Hrsg.): Berufsgenossenschaftliche Grundsätze für spezielle arbeitsmedizinische Untersuchungen. Stuttgart: Gentner Verlag, 1994: 639–651
- <sup>8</sup> Mushtaq M, Hayton R, Watts T et al. An audit of pulmonary function laboratories in the West Midlands. *Respir Med* 1995; 89: 263–270
- <sup>9</sup> Ohmann C. Preoperative prediction of peri- and postoperative complications: criteria for statistical evaluation of predictive systems. *Theor Surg* 1991; 6: 3–18
- <sup>10</sup> Ohmann C. Qualitätssicherung aus theoretischer Sicht. *Chirurg* 1995; 66: 657–664
- <sup>11</sup> Ohmann C, Groß-Weege W. Scoring-Systeme auf der chirurgischen Intensivstation. I. *Chirurg* 1992; 63: 1021–1028
- <sup>12</sup> Ohmann C, Groß-Weege W. Scoring-Systeme auf der chirurgischen Intensivstation. II. *Chirurg* 1993; 64: 21–27
- <sup>13</sup> Otten H, Perlebach E, Piekarski C et al. Die Betreuung der ehemaligen Arbeitnehmer im Uranerzbergbau Wismut. *Berufsgenossenschaft* 1999 [Mai]; 285–290
- <sup>14</sup> Piirilä P, Pietinalho A, Lopenen M et al. The quality of spirometric examinations in Finland: results from a national questionnaire survey. *Clin Physiol Func Im* 2002; 22: 233–239
- <sup>15</sup> Schuldheisz S, Phillips BA, Berry DT. Pulmonary function testing: detection of invalid performance. *J Ky Med Assoc* 1998; 96: 183–185
- <sup>16</sup> Stoller JK, Buist S, Burrows B et al. Quality control of spirometric testing in the registry for patients with severe  $\alpha$ -1-antitrypsin deficiency. *Chest* 1997; 111: 899–909
- <sup>17</sup> Valenti E, Manzari G, De Angelis V et al. Quality control of spirometric tests used in health surveillance for occupational exposure risk in the province of Viterbo. *Med Lav* 2001; 92: 5–11
- <sup>18</sup> Zorn U, Ollenschläger G. Qualitätsbestimmung in der medizinischen Versorgung – ein universelles Entwicklungsinstrument für Qualitätsindikatoren. *Zschr ärztl Fortbild Qualsicherung* 1999; 93: 123–128