

# Leitlinien Diagnostische Validität

## Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“

### Einleitung

Zur Einordnung umweltmedizinischer Diagnosen in kausale umweltbezogene Zusammenhangesannahmen dienen neben dem Umweltmonitoring, d. h. dem chemischen Schadstoffnachweis in einem Umweltmedium wie Luft, Trinkwasser, Nahrung, Hausstaub, das Humanbiomonitoring mit Nachweis von Schadstoffen und deren Metaboliten bzw. Reaktionsprodukten in biologischen Matrices (Haare, Blut, Serum, Urin, Muttermilch, Fettgewebe etc.). Für den Zusammenhang zwischen Exposition und Wirkung sind darüber hinaus einige Parameter des Effekt- und Suszeptibilitätsmonitoring etabliert. Während in der medizinischen Praxis und Klinik eine große Anzahl von diagnostischen Tests der klinischen Chemie/Laboratoriumsmedizin existieren und in der Arbeitsmedizin zahlreiche Biomarker der Belastung und Beanspruchung zur Verfügung stehen, ist das labordiagnostische Repertoire für klinisch-umweltmedizinische Indikationen begrenzt.

Grundsätzlich ist die Erweiterung des Repertoires diagnostischer Parameter in der Umweltmedizin wünschenswert und die Erprobung, Evaluation sowie Einführung neuer Verfahren daher zu begrüßen. Ein neuer diagnostischer Test kann hierbei ein biochemischer/klinisch chemischer Parameter sein, der entweder vollkommen neuartig ist, oder auch ein Testverfahren, das bereits bei anderen klinischen Fragestellungen diagnostisch etabliert ist und basierend auf gewonnenen Erfahrungen neu für umweltmedizinische Indikationen eingesetzt werden soll.

Ein guter diagnostischer Test, mit dem das Vorliegen einer Krankheit beurteilt werden soll, zeichnet sich dadurch aus, dass er zugleich möglichst viele tatsächlich Erkrankte identifiziert und die Gruppe der Nicht-Erkrankten weitgehend ausschließt. Leider werden z.T. aber auch wenig etablierte Parameter aufgrund theoretischer Modelle ohne vorausgehende Validierung für die Diagnostik herangezogen und propagiert. Der umweltmedizinisch Tätige wird somit häufig mit neuen diagnostischen Parametern oder Tests konfrontiert, deren diagnostischer Stellenwert mitunter schwierig zu beurteilen ist. Die RKI-Kommission für Qualitätssicherung in der Umweltmedizin hat daher, unter Einbeziehung externer Sachverständiger, bereits in mehreren im Bundesgesundheitsblatt publizierten Mitteilungen zur Anwendung einzelner diagnostischer Parameter Stellung bezogen (u.a. zu Melatonin [1], Pyrrolurie [2], Parametern des roten Blutbildes bei Mobilfunkexpositionen [3], zum Lymphozytentransformationstest [4] und anderen Immunparametern und zu Parametern des oxidativen Stresses [5]). In der Kommissionsbewertung der meisten bisher betrachteten Parameter ist der fehlende Nachweis der diagnostischen Validierung kritisiert und infolgedessen sind die entsprechenden Testverfahren für die praktische Umweltmedizin als nicht geeignet eingestuft worden.

Die Kommission möchte mit der vorliegenden Stellungnahme die Grundanforderungen benennen, die erfüllt sein müssen, bevor ein neuer Parameter in der individuellen umweltmedizinischen Diag-

nostik eingeführt wird. Nach Auffassung der Kommission sollten etablierte methodische Grundanforderungen an die Validität von demjenigen fachlich belegt werden, der neue diagnostische Parameter in die klinische Umweltmedizin einführt oder solche als diagnostisches Instrument empfiehlt. Damit werden für die Umweltmedizin vergleichbare Qualitätskriterien eingefordert, wie sie in anderen Medizinbereichen, nicht zuletzt in Zusammenhang mit evidenzbasierter Medizin eingefordert werden [6, 7, 8] und insbesondere in der Labormedizin selbstverständlich sind [9, 10, 11]. Die Kommission nimmt jedoch zur Kenntnis, dass allgemeingültige Kriterien der diagnostischen Qualitätssicherung auch in anderen Bereichen nicht immer befriedigend eingehalten wurden [12].

Wenn nichtvalidierte Methoden eingesetzt werden, oder Ergebnisse nicht sachgemäß interpretiert werden, werden dem Patienten Belastungen, z.T. Eingriffe und erhebliche Kosten, aufgebürdet, die nicht zu einer sachdienlichen Information oder nicht zu geeigneten Maßnahmen zur Minderung von Beschwerden führen, eventuell sogar unzutreffende Diagnosen und Behandlungsansätze nach sich ziehen.

Die Kommission empfiehlt für die Absicherung der diagnostischen Qualität von neuen Parametern in der umweltmedizinischen Diagnostik ausdrücklich die Durchführung von wissenschaftlichen Studien und Vergleichsuntersuchungen, in denen zwischen exponierten Kollektiven und Kontrollgruppen unterschieden wird.

**Anforderungen an die Validierung**

In den letzten Jahren sind Vorschläge und Checklisten publiziert worden, die bei der Validierung einer diagnostischen Methode zu beachten sind [7, 13, 14]. Aus methodischer Sicht sind vor allem 2 Validierungsschritte wichtig: der analytische und der diagnostische Schritt.

Die analytische Validierung muss die Messeigenschaften und den bestimmungsgemäßen Gebrauch des analytischen Verfahrens darstellen (Leistungsdaten: Präzision, Richtigkeit, Linearität, Nachweisgrenze, Spezifität, Messunsicherheit). Eine ausreichende analytische Qualität wird hier als gegeben vorausgesetzt.

Zur diagnostischen Validierung des klinischen Parameters zählen darüber hinaus vor allem Aussagen zu Normbereichen und Normabweichungen sowie zum Expositionsbezug bzw. zur (umweltmedizinischen) Krankheit. Diese sind mit den Kenngrößen der diagnostischen Sensitivität, Spezifität sowie des prädiktiven Wertes (Vorhersagewertes) des Testverfahrens [10] zu dokumentieren.

Eine eindeutige Unterscheidung zwischen der analytischen und der diagnostischen Stufe ist sehr wichtig. Diese wird dadurch erschwert, dass die Begriffe „Sensitivität“ und „Spezifität“ jeweils unterschiedliche Bedeutung haben [11, 15]:

Analytische Sensitivität beschreibt die Nachweisstärke der Labormethode.

Analytische Spezifität beschreibt, inwieweit die Labormethode nur das misst, was sie vorgibt zu messen.

Diagnostische Sensitivität ist das Vermögen der Untersuchungsmethode, möglichst alle Erkrankten zu erfassen.

Diagnostische Spezifität ist das Vermögen der Untersuchungsmethode, gezielt ein Krankheitsbild zu erfassen und somit Fehlzuordnungen gering zu halten.

Mit der Monographie „Norm und Normabweichung klinischer Daten“ haben Galen und Gambino 1975 ein auf dem prädiktiven Wert aufbauendes Modell zur statistischen Auswertung von klinischen und Labordaten vorgelegt, welches heute noch Gültigkeit besitzt. Einzelheiten und Verfahren sind in dieser Monographie

Tabelle 1

4-Feldertafel			
	Exposition/Gesundheitszustand des Patienten		Summe
	Exponiert/krank	Nicht exponiert/nicht krank	
Testergebnis positiv	richtig positiv (rp)	falsch positiv (fp)	rp+fp
Testergebnis negativ	falsch negativ (fn)	richtig negativ (rn)	fn+rn
Summe	rp+fn	fp+rn	rp+fp+fn+rn

[16] und in verschiedenen Übersichtsarbeiten [9, 10, 11] ausführlich beschrieben. Gute Beschreibungen der anzuwendenden statistischen Verfahren sind zusammen mit anschaulichen Beispielen im Internet einsehbar [17]. Als Grundkriterien zur Objektivierung der Eignung eines klinisch-chemischen/labordiagnostischen Parameters sind die Kennwerte der Spezifität und die Sensitivität üblich (siehe weiter unten). Ohne diese Angaben ist der mögliche Nutzen für den klinischen Einsatz nicht geklärt.

Umweltmedizinische Diagnostik beruht auf dichotomen Entscheidungen, z.B. „krank/nicht krank“ oder auch „exponiert/nicht exponiert“. Ein diagnostischer Test/Biomonitoring soll die richtige Zuordnung zu einem dieser Zustände unterstützen [7]. Eine Laboruntersuchung kann dabei als ein prognostisches Verfahren (Test) aufgefasst werden, das eine „richtige Vorhersage“ machen soll, oder aber auch als erklärendes/retrospektives Verfahren, das die Auswahl einer angemessenen Ursachenzuweisung unterstützt. Letzteres kann, übertragen auf die Umweltmedizin, einer Entscheidung zwischen „exponiert/nicht exponiert“ entsprechen. Demzufolge wird aus einem Testresultat eine Ja-/Nein-Aussage: „krank“ bzw. „exponiert“ (positiv) und „nicht krank“ bzw. „nicht exponiert“ (negativ). Laborwerte des Human-Biomonitorings, ebenso wie andere Tests mit kontinuierlichen Skalen (Laborwerte), werden in dichotome Aussagen umgewandelt, indem sie zu einem Referenzwert bzw. zu einem diagnostischem Entscheidungswert (Clinical Decision Level (DL), Cut-Off-Wert) in Relation gestellt werden. In jedem Fall müssen für die Klassifizierung und Beurteilung der Messungen und Tests einheitliche und begründete Ver-

fahren herangezogen werden. Die empfohlenen Entscheidungswerte sind ggf. nach Alter und Geschlecht zu differenzieren und, sofern vorhanden, ist der Einfluss weiterer Faktoren bei der Durchführung der Tests und Messungen anzugeben.

Im Idealfall würde jedem positiven Testergebnis auch eine klare Krankheits- bzw. Expositionsdiagnose gegenüberstehen und ebenso jedem negativen Testergebnis ein belegter Ausschluss der Krankheit bzw. der Exposition. In der Realität werden sich jedoch weniger eindeutige Verteilungen finden. Um die Übereinstimmung von Testergebnissen mit den tatsächlichen Gegebenheiten zu beschreiben, können die Ergebnisse aus einer Validierungsstudie mit vielen empirischen Test-Ergebnis-Paaren von „Ja/Nein“-Aussagen in einer 4-Feldertafel zusammengeführt werden (■ **Tabelle 1**).

Ein diagnostisches Verfahren bzw. ein diagnostischer Test würde insbesondere dann in seiner Qualität als besonders geeignet eingestuft, wenn sich die Werte der Einzeltests allein auf die Diagonale (Zelle: „richtig positiv“ und Zelle: richtig negativ (rn)) verteilen würden; ein solcher Test wäre akkurat ((rp+rn)/(rp+fp+fn+rn)), sensitiv und spezifisch. Die Güte von umweltmedizinischen Tests wird jedoch nur selten dieses Kriterium erfüllen können. Die diagnostische Sensitivität beschreibt den Anteil der durch den Test richtig erkannten Personen mit der fraglichen Erkrankung bzw. einer vorliegenden Exposition zu allen Kranken/Exponierten (rp/(rp+fn)), d.h. die Güte des Tests innerhalb der Gruppe der Erkrankten/Exponierten. Die Spezifität gibt den Anteil der durch den Test richtig negativ erkannten Personen an den Nicht-Erkrankten (rn/(rn+fp)) wieder, sie beschreibt die Güte

des Tests in Bezug zum Ausschluss von Erkrankung/Exposition.

Es sei angemerkt, dass zur Berechnung der Sensitivität und Spezifität eine anerkannte Diagnostik („Goldstandard“) für die Erkrankung bzw. das Vorliegen einer Exposition erforderlich ist [8], da die Verteilung der Testwerte, ausgehend von der Kenntnis um die „Ja-/Nein“-Entscheidung der Diagnostik, rückblickend (retrospektiv) betrachtet wird. Die Kenngrößen der diagnostischen Sensitivität und der Spezifität sind die beiden Größen, die vom Labor in jedem Fall zur Bewertung ihrer diagnostischen Tests angegeben werden müssen. Für die Beurteilung der Gesamtgüte des Testverfahrens wäre es notwendig, jeweils den genauen Anteil der Erkrankungen/Expositionen unter den diagnostizierten Personen zu kennen.

Wenn – wie es in der Umweltmedizin oft der Fall ist – kein echter „Goldstandard“ zur Verfügung steht, sind Sensitivität und Spezifität nicht absolute, sondern nur relative Größen.

In der medizinischen Praxis hat allerdings nicht nur der erklärende Anteil der Ursachenzuweisung Bedeutung. Für den klinischen Bezug ist insbesondere der prognostische Wert von Test- und Diagnoseverfahren von Bedeutung. Der Vorhersagewert (prädiktiver Wert) eines diagnostischen Verfahrens beschreibt die Wahrscheinlichkeit, dass der Patient tatsächlich den Zustand aufweist, den der Test anzeigt (positiver Vorhersagewert:  $rp/(rp+fp)$ , negativer Vorhersagewert:  $rn/(rn+fn)$ ).

Als Maß der diagnostischen Information eines Tests dient auch die Likelihood Ratio (LR), die das Chancenverhältnis beschreibt, um wie viel häufiger ein Testergebnis bei einem Kranken im Vergleich mit einem nicht erkrankten Patienten auftritt [8]:

$$LR_{\text{positiv}} = \text{Sensitivität}/(1-\text{Spezifität})$$
$$LR_{\text{negativ}} = (1-\text{Sensitivität})/\text{Spezifität}$$

Die Höhe des positiv prädiktiven Wertes ist immer im Zusammenhang mit der Prävalenz einer Krankheit zu interpretieren [18, 19]. Ob ein Test sinnvoll eingesetzt wird, hängt insofern von der A-priori-Wahrscheinlichkeit für die Krankheit im Einzelfall ab. Da oft zuverlässige Angaben

zur Prävalenz fehlen, gewinnt die Vortestwahrscheinlichkeit auf der Grundlage der Anamnese und subjektiven Einschätzung des Arztes an Bedeutung.

Eine dokumentiert hohe prognostische Güte eines Testverfahrens gibt dem Praktiker die Sicherheit, dass sich die aus dem Test abgeleitete Information auch mit hoher Wahrscheinlichkeit als richtig erweist. Diagnostische Testverfahren mit einer hohen Vorhersagekraft (hohe positive und negative prädiktive Werte) haben für den Arzt bei der Einordnung und Wertung der Befunde eine höhere Relevanz.

### Vorgehensweise bei der Validierung

Die Kommission empfiehlt folgendes mehrstufiges Vorgehen zur Validierung diagnostischer Tests unter Heranziehung biostatistischer Sachverständiger:

1. Validierung der analytischen Methode (z.B. Richtigkeit und Präzision) und analytische Qualitätssicherung des Parameters einschließlich der Präanalytik (z.B. [11, 20, 21]).
2. Validierung der diagnostischen Aussagekraft durch Untersuchungen zur Verteilung des Parameters (Häufigkeitsverteilung) bei spezifisch Erkrankten und Gesunden unter Berücksichtigung von demographischen Faktoren (Alter, Geschlecht etc.) und unter Einschluss von Patienten mit unterschiedlicher Exposition und verschiedene Begleiterkrankungen. Die Erfassung von möglichen Confoundern (z.B. Beruf, Lebensstil, Rauchen etc.) ist unerlässlich. Anhand dieser Daten kann eine Aussage zur Sensitivität und Spezifität getroffen werden.
3. In einer kontrollierten prospektiven Studie wird der Test bei Patienten mit der Verdachtsdiagnose getestet und das Ergebnis der durch einen definierten und anerkannten „Goldstandard“ gesicherten spezifischen Diagnose gegenübergestellt. Anhand der Ergebnisse einer solchen Studie können Entscheidungs-Werte (Cut-Off-Werte) festgelegt werden.

Das einfachste Verfahren ist die Bestimmung des Referenzwertes bei dem Kollektiv der Gesunden bzw. Nicht-Exponierten

als 95. Perzentil (s. a. [21, 22, 23, 24, 25]). Die hierbei allgemein empfohlene Mindest-Fallzahl liegt bei 120 Probanden.

In der klinischen Chemie/Laboratoriumsmedizin wird der Cut-Off-Wert meistens durch eine ROC-Analyse (Receiver Operating Characteristic) ermittelt, indem für verschiedene Bezugspunkte („Cut-off-Punkte“) jeweils die Sensitivität und Spezifität berechnet und gegeneinander aufgetragen werden [26]. In der Expositionsabschätzung werden ROC-basierte Verfahren zur Validierung von Expositionsprognosen angewandt [27]. Dabei ergibt sich jeweils eine tabellarische bzw. graphische Übersicht, wie viele Patientendaten je nach Lage des Cut-Off-Punktes (bzw. des gewählten Referenzwertes) als test-positiv oder test-negativ eingeordnet werden.

Bei der Festlegung des für das Verfahren empfohlenen Cut-Off-Wertes ist sorgfältig abzuwägen, ob ein Fehler erster oder zweiter Art (falsch positiv oder falsch negativ) schwerwiegender ist. Für Suchtests (Screening) wird man eher Wert auf eine hohe Empfindlichkeit (Sensitivität) legen, für Beweistests dagegen eine hohe Spezifität wählen.

Zur Methodik und zum Vergleich unterschiedlicher Verfahren wird auf die entsprechenden Übersichtsarbeiten verwiesen [9, 28, 29, 30, 31, 32, 33].

### Zusammenfassung

Die Kommission begrüßt die Einführung neuer Parameter mit umweltmedizinischer Indikation, wenn diese den üblichen Qualitätsansprüchen der Laboratoriumsmedizin entsprechen. Auch für neu eingeführte Parameter in der Umweltmedizin muss das Prinzip der Evidenz gelten, die vorher nachzuweisen ist. Ein Grundprinzip des ärztlichen Handelns ist das *nil nocere*. Fehlerhafte, nicht valide Diagnostik mit der Folge zweifelhafter Diagnosen erfüllt diesen Anspruch nicht und ist deshalb abzulehnen.

### Federführung

Dr. med. Birger Heinzow, Landesamt für soziale Dienste Schleswig-Holstein, Umweltbezogener Gesundheitsschutz, Brunswikerstr. 4, 24105 Kiel

Prof. Dr. med. Dr. phil. nat. Andreas D. Kappos (Frankfurt/Main)

Prof. Dr. rer. nat. Michael Schwenk (Tübingen)

Dr. Michael Schümann (Behörde für Soziales, Gesundheit und Verbraucherschutz, Freie und Hansestadt Hamburg, Billstraße 80a, 20539 Hamburg)

### Externer Sachverständiger

Prof. Dr. med. Jürgen Windeler, Medizinischer Dienst der Spitzenverbände der Krankenkassen, Lützowstraße 53, 45141 Essen

### Kommissionsmitglieder

Dr. A. Beyer (Umweltmed. Ambulanz Berlin-Steglitz/Zehlendorf), Prof. Dr. W. Dott (Universitätsklinikum Aachen, Institut für Hygiene und Umweltmedizin), Prof. Dr. H. Drexler (Friedrich-Alexander-Universität Erlangen-Nürnberg, Institut für Arbeits-, Sozial- und Umweltmedizin), Prof. Dr. H. Dunkelberg (Universität Göttingen, Abt. Allg. Hygiene u. Umweltmedizin), Prof. Dr. Th. Eikmann (Universität Gießen, Institut f. Hygiene u. Umweltmedizin), Dr. B. Heinzow (Landesamt für soziale Dienste Schleswig-Holstein, Dezernat Umweltbezogener Gesundheitsschutz, Kiel), Prof. Dr. C. Hornberg (Universität Bielefeld, Fakultät für Gesundheitswissenschaften), Prof. Dr. Dr. A.D. Kappos (Frankfurt/Main), Prof. Dr. K.E. von Mühlendahl (Kinderhospital Osnabrück, Gemeinnützige Kinderumwelt GmbH), Prof. Dr. D. Nowak (LMU München, Klinikum Innenstadt, Institut u. Poliklinik für Arbeits- und Umweltmedizin), PD Dr. F.-A. Pitten (Institut für Krankenhaushygiene und Infektionskontrolle GbR, Gießen), Dr. W. Stück (Ökologischer Ärztenbund/ISDE, Koblenz), Prof. Dr. M. Schwenk (Tübingen), Dr. R. Suchenwirth (Niedersächsisches Landesgesundheitsamt, Abt. Umweltmedizin/Epidemiologie, Hannover), Prof. Dr. M. Wilhelm (Universität Bochum, Hygiene, Sozial- und Umweltmedizin).

### Ständige Gäste

Dr. N. Englert (Umweltbundesamt, Berlin), Dr. A. Hahn (Bundesinstitut für Risikobewertung, Berlin), Dr. U. Winkler (Bundesministerium für Gesundheit, Referat 332).

### Geschäftsstelle im RKI

Dr. D. Eis, Dr. U. Wolf.

### Literatur

1. RKI-Kommission (2005) Melatonin in der umweltmedizinischen Diagnostik im Zusammenhang mit elektromagnetischen Feldern (EMF). Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 48:1406–1408
2. RKI-Kommission (2007) Die (Krypto-)Pyrrolurie in der Umweltmedizin: Eine valide Diagnose? Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:1324–1330
3. RKI-Kommission (2006) Parameter des roten Blutbildes bei Exposition durch Mobilfunkanlagen. Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 49:833–835
4. RKI-Kommission (2008) Qualitätssicherung beim Lymphozytentransformationstest – Addendum zum LTT-Papier der RKI-Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz (im Druck)
5. RKI-Kommission (2008) Oxidativer Stress und Möglichkeiten seiner Messung aus umweltmedizinischer Sicht. Mitteilung der Kommission „Methoden und Qualitätssicherung in der Umweltmedizin“. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz (im Druck)
6. Price CP (2000) Evidence-based laboratory medicine: supporting decision-making. *Clinical chemistry* 46:1041–1050
7. Bossuyt PM, Reitsma JB, Bruns DE, et al. (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine* 138:W1–12
8. Hawkins RC (2005) The evidence based medicine approach to diagnostic testing: practicalities and limitations. *The Clinical biochemist* 26:7–18
9. Köbberling J, Richter K, Trampisch HJ, Windeler J (1991) Methodologie der medizinischen Diagnostik. Entwicklung, Beurteilung und Anwendung von Diagnoseverfahren in der Medizin. Springer-Verlag, Berlin, Heidelberg, New York
10. Richter K, Abel U, Klar R, et al. (2005) Die Grundlagen der Validierung einfacher diagnostischer Tests. *Journal of Molecular medicine* 66: 1432–1440
11. Hallbach J (2006) Klinische Chemie und Hämatologie für den Einstieg, 2. überarbeitete Auflage. Qualitätssicherung; Methoden- und Geräteevaluierung. Thieme, Stuttgart, S. 368–392
12. Oosterhuis WP, Niessen RW, Bossuyt PM (2000) The science of systematic reviewing studies of diagnostic tests. *Clin Chem Lab Med* 38:577–588
13. Van den Bruel A, Cleemput I, Aertgeerts B, et al. (2007) The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *Journal of clinical epidemiology* 60:1116–1122
14. Sackett DL, Haynes RB (2002) The architecture of diagnostic research. *BMJ (Clinical research ed.)* 324:539–541

15. Saah AJ, Hoover DR (1997) “Sensitivity” and “specificity” reconsidered: the meaning of these terms in analytical and diagnostic settings. *Annals of internal medicine* 126:91–94
16. Galen RS, Gambino SR (1975) Beyond normality. The predictive value and efficiency of medical diagnoses. Wiley, New York
17. Biorama Virtuell medizinisch analytisches Labor. <http://www.biorama.ch/>.
18. Hrudye SE, Leiss W (2003) Risk management and precaution: insights on the cautious use of evidence. *Environmental health perspectives* 111:1577–1581
19. Loong TW (2003) Understanding sensitivity and specificity with the right side of the brain. *BMJ (Clinical research ed.)* 327:716–719
20. Eurachem Methodenvalidierung der Eurachem: <http://www.eurachem.org/>
21. NCCLS (2000) C28-A2: How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline – Second Edition ISBN 1-56238-406-6. <http://www.clsi.org/>
22. HBM-Kommission (1996) Konzept der Referenz- und Human-Biomonitoring-(HBM)-Werte in der Umweltmedizin. Bundesgesundheitsbl 39:221–224
23. Solberg HE (2006) Establishment and use of reference values. In: Burtis CA, Ashwood ER, Bruns DE (eds.) Tietz textbook of clinical chemistry and molecular diagnostics, 4th ed. Saunders, St. Louis, pp 425–448
24. Henny J, Petersen PH (2004) Reference values: from philosophy to a tool for laboratory medicine. *Clin Chem Lab Med* 42:686–691
25. Poulsen OM, Holst E, Christensen JM (1997) A supplement to the approved IFCC Recommendation and the theory of reference values. *Pure & Appl Chem* 69:1601–1611
26. Greiner M, Pfeiffer D, Smith RD (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine* 45:23–41
27. Schümann M (2007) Methoden der Bewertung von Expositionsprognosen. In: Mekel O, Mosbach-Schulz O, Schümann M, et al. (eds.) Evaluation von Standards und Modellen der Expositionsabschätzung. Teil 1: Grundlagen der bevölkerungsbezogenen Expositionsmodellierung. WaBoLu-Hefte 02/07, S. 161–179. <http://www.umweltdaten.de/publikationen/fpdf-f/3357.pdf>. Umweltbundesamt, Berlin
28. DeBari VA (2006) Computation of decision levels from differentiated logistic regression probability curves. *Annals of clinical and laboratory science* 36:194–200
29. Formann AK, Kohlmann T (1996) Latent class analysis in medical research. *Statistical methods in medical research* 5:179–211
30. Schäfer H (1989) Constructing a cut-off point for a quantitative diagnostic test. *Statistics in medicine* 8:1381–1391
31. Magder LS, Fix AD (2003) Optimal choice of a cut point for a quantitative diagnostic test performed for research purposes. *Journal of clinical epidemiology* 56:956–962
32. Obuchowski NA, Lieber ML, Wians FH Jr (2004) ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical chemistry* 50: 1118–1125
33. Gallop RJ, Crits-Christoph P, Muenz LR, Tu XM (2003) Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. *Understanding statistics* 2:219–242