

RiliBÄK – ein neues Berechnungskonzept

Im Rahmen der BNLD-Informationen wurde bereits mehrfach aufgezeigt, welche Folgen die kommende „Richtlinie der Bundesärztekammer zur Qualitätssicherung quantitativer laboratoriumsmedizinischer Untersuchungen“ (RiLiBÄK) haben wird, wenn sie demnächst in Kraft gesetzt wird (1-3). Mit dem Versand eines neuen Entwurfes (Stand: August 2006) durch die Bundesärztekammer wird deutlich, dass der Teil A unverändert in Kraft gesetzt werden soll. Dieser Teil zwingt alle Laboratorien zur Einhaltung zahlreicher "Soll"-Vorschriften nach DIN 15189, die im Teil A als "Muss"-Vorschriften formuliert worden sind. Da die Überprüfung der Laboratorien durch Staatsbeamte statt findet und nicht, wie bei der Akkreditierung üblich, durch einen Laborleiter eines anderen Labors (also einen Fachmann, dem die Sachverhalte aus eigener Erfahrung bekannt sind), darf man davon ausgehen, dass die neue RiliBÄK einer verschärften Akkreditierung gleich kommt und keinesfalls als "Minimalakkreditierung" oder "Akkreditierung light" zu betrachten ist. Dem BNLD-Vorschlag, Teil A als Empfehlung und nicht als Richtlinie zu veröffentlichen, ist die Bundesärztekammer nicht entgegen gekommen.

Um nicht erneut missverstanden zu werden: Eine Akkreditierung auf freiwilliger Basis kann z.B. für eine Beteiligung an Studien zwingend notwendig sein, weil durch die umfangreiche Dokumentation auch die Rückverfolgung über größere Zeiträume hinweg gesichert ist. Vielfach hat sich auch die Akkreditierung einzelner Verfahren oder die Akkreditierung von Teilbereichen in einigen Laboratorien als grundsätzlich sinnvoll erwiesen. Im Rahmen der RiliBÄK würde es sich aber um eine staatliche Zwangsakkreditierung aller Labors auf dem Verordnungsweg handeln ohne dass bisher bewiesen wurde, dass sich hierdurch die Qualität der Patientenversorgung verbessern ließe.

Der aktuelle Entwurf der zukünftigen Richtlinie enthält nun auch den Teil B, welcher der eigentlichen alten RiliBÄK entspricht, mitsamt dem Anhang mit allen Verfahren und der jeweils maximal zulässigen Abweichung für die interne Qualitätskontrolle, bei denen die Pflicht zur externen Qualitätskontrolle besteht. Grundlage dieses Teils B ist nicht mehr das von Shewhart 1930 für die Fertigungsindustrie entwickelte Modell der statistischen Qualitätskontrolle (4) und seine Übertragung auf die klinische Chemie durch Levey und Jennings (5) in der Weiterentwicklung durch Westgard (6 und 7), sondern die Vereinfachungen, die von Macdonald (8) im Sommer dieses Jahres publiziert wurden.

Das neue Modell macht die Ermittlung von Unpräzision ($RE = \text{random error}$) und Unrichtigkeit (Bias) überflüssig und betrachtet nur noch die mittlere quadratische Messabweichung ($TE = \text{total error} = \text{root mean square (rms)}$). Während die Unpräzision die Standardabweichung bezogen auf den Mittelwert darstellt, entspricht die mittlere quadratische Messabweichung nun der Standardabweichung bezogen auf den Sollwert. Macdonald zeigt auch den Zusammenhang zwischen diesen Kenngrößen:

$$\text{rms} = \Delta = \text{TE} = \sqrt{(n-1)^{-1} \sum (x_i - \mu)^2} = \sqrt{((n-1)/n) * \text{RE}^2 + \text{Bias}^2} \quad (\text{a})$$

Die Variablen in dieser Gleichung wurden nicht in der Benennung von Macdonald übernommen, sondern durch die in der Labormedizin üblichen, von Westgard geprägten Bezeichnungen ersetzt. Bisher gab die Richtlinie in Spalte 4 eine maximal zulässige Unpräzision ($\text{RE}_a = \text{allowable random error}$) und in Spalte 5 eine maximal zulässige Unrichtigkeit ($\text{Bias}_a = \text{allowable bias}$) vor, aus denen eine maximal zulässige Abweichung des Einzelwertes (TE_a) in Spalte 7 nach folgender Formel berechnet wurde:

$$\text{TE}_a = z * \text{RE}_a + \text{Bias}_a \quad (\text{b})$$

Mit dem Multiplikator z wird festgelegt, wie hoch der Anteil an Kontrollmessungen ist, die außerhalb des zulässigen Bereiches sind, obwohl stabile Messbedingungen vorliegen. Diese Wahrscheinlichkeit für eine irrtümliche Fehleranzeige ($p_{fr} = \text{probability of false rejection}$) ist bei einer Kontrollmessung 33 %, wenn $z = 1$ (67 % - Vertrauensbereich) ist, beträgt 5 % für $z = 2$ (95 %-Vertrauensbereich) und nur noch 0,1 % für $z = 3$ (99 % Vertrauensbereich). In der zur Zeit noch gültigen RiliBÄK wurde $z = 2$ gewählt. Bei einem stabilen Verfahren, dessen Präzision und Richtigkeit den Vorgaben entsprechen, ist also zu erwarten, dass 5 % aller Kontrollergebnisse - obwohl sie korrekt sind - zu einer irrtümlichen Fehleranzeige führen. Beim Einsatz von 2 Kontrollproben erhöht sich diese Fehleranzeige auf 10 %.

Da Messwerte grundsätzlich normal verteilt sind, beträgt die Wahrscheinlichkeit, dass eine Kontrollmessung ober- oder unterhalb des $3s$ -Bereiches liegt, genau 0,135 % (Abb. 1), wobei s die analytische Unpräzision des Verfahrens aus der Vorperiode darstellt. Bezieht man Unpräzision und Unrichtigkeit auf die vorgegebene oder in der Vorperiode ermittelte maximal erlaubte Unpräzision, so kann man mittels dieser s -Metrik auch die Wahrscheinlichkeit dafür bestimmen, Verfahrensfehler zu erkennen.

Sinn einer Qualitätskontrolle bei labordiagnostischen Untersuchungen ist die Erkennung von Fehlbestimmungen, bevor die Patientenmesswerte freigegeben werden. Die bisherige Qualitätskontrolle erkannte eine Unrichtigkeit von $3s$ aber nur in 50 % der Fälle (Abb. 2), obwohl dies bereits eine klinisch relevante Drift des Messverfahrens darstellen kann. Die Empfindlichkeit zur Erkennung einer steigenden Unpräzision ist noch geringer. Ein Anwachsen um den Faktor 3 erkennt man nur in einem Drittel der Fälle (Abb. 3).

Selbst dann, wenn stets 2 Kontrollproben gemessen werden, verbessert sich die Empfindlichkeit zur Erkennung von Fehlern nur geringfügig. Eine $3s$ -Drift würde zu 75 % und ein Anwachsen der Unpräzision auf das Dreifache würde zu 53 % entdeckt, während gleichzeitig der Anteil irrtümlicher Fehleranzeigen auf 0,5 % steigen würde.

Dies zeigt : das Vorperiodenkonzept mit der Festlegung laborinterner Fehlergrenzen eignet sich wenig zur Erkennung analytischer Fehler bei Einzelmessungen, wenn nicht zusätzliche Kontrollregeln zu Hilfe genommen werden. Die Anwendung z.B. der Westgard-Regeln wird von den meisten MTAs nicht als strenges Regelwerk, sondern mehr aus der Erfahrung heraus praktiziert. Dies trägt aber deutlich zur besseren Fehlererkennung bei. Eine zwingende Umsetzung solcher Regeln ist aber nicht

empfehlenswert, weil der Einsatz der Regeln abhängig von den analytischen Eigenschaften des Messverfahrens und seinen potentiell möglichen Fehlern ist.

Eine hohe Empfindlichkeit zur Erkennung analytischer Probleme besitzt die Berechnung der Unpräzision und Unrichtigkeit am Ende eines Kontrollzyklusses. Leider fehlt dieses hervorragende Verfahren im neuen Entwurf der Richtlinie, obwohl es von Macdonald keineswegs als überflüssig eingestuft wurde. Einzig die Bewertung der Einzelmessung an den Grenzen der maximal zulässigen Abweichung (TE_a) und die Bewertung des mittleren quadratischen Fehlers (TE) am Ende des Kontrollzyklus ist in der Richtlinie noch enthalten.

Da es bisher möglich war, dass jeder Messwert eines Kontrollzyklusses innerhalb der vorgegebenen Grenzen lag, bei der Berechnung am Monatsende aber die Grenzen für die erlaubte Unpräzision oder Unrichtigkeit überschritten wurden, fordert Macdonald die Beseitigung dieses Problems und schlägt zur Berechnung der zukünftigen Grenzen folgendes vor:

$$TE_a = \sqrt{(n-1)/n * RE_a^2 + Bias_a^2} \quad (c)$$

Für Verfahren, die nicht in der Anlage der RiliBÄK aufgeführt sind, sollen eigene Grenzen in einer Vorperiode wie folgt ermittelt werden:

$$TE_a = \sqrt{(n-1)/n * z^2 * RE^2 + Bias^2} \quad (d)$$

Bei Macdonald und in der RiliBÄK wird der Korrekturfaktor nicht mit dem Buchstaben "z" bezeichnet, sondern mit dem Buchstaben "k". Macdonald schlägt aufgrund statistischer Betrachtungen vor, diesen Faktor gleich drei zu wählen; im Entwurf der RiliBÄK wurde diese Festlegung bisher vergessen, was eine Bewertung des Verfahrens unmöglich macht. Im Folgenden wird deshalb davon ausgegangen, dass man dem Vorschlag von Macdonald auch bei dieser Festlegung folgen wird.

Will man die alte RiliBÄK mit der neuen Fassung vergleichen, so empfiehlt sich die Betrachtung der Extremwerte. Für den Fall einer vernachlässigbaren Unpräzision kann man bei genügend großer Anzahl von Messungen die Gleichungen b bis d wie folgt schreiben

$$TE_a = z * RE_a + Bias_a$$

$$TE_a = \sqrt{(n-1)/n * RE_a^2 + Bias_a^2}$$

$$TE_a = \sqrt{(n-1)/n * 9 * RE^2 + Bias^2}$$

Dies zeigt, dass das alte und das neue Verfahren hinsichtlich der Empfindlichkeit zur Erkennung einer Drift identisch sind.

Der zweite Extremfall geht von einer vernachlässigbaren Unrichtigkeit aus; die Gleichungen b bis d lassen sich dann wie folgt schreiben:

$$TE_a = z * RE_a + Bias_a \quad (e)$$

$$TE_a = \sqrt{(n-1)/n * RE_a^2 + Bias_a^2} \approx RE_a \quad (f)$$

$$TE_a = \sqrt{((n-1)/n) * 9 * RE^2 + Bias_a^2} \approx 3 * s \quad (g)$$

In diesem Fall erkennt man, dass für die Parameter der Anlage der RiliBÄK, für die Gleichung f gilt, der Korrekturfaktor in Gleichung e gleich 1 gewählt wurde: $z = 1$. Ausgehend von der bisher erlaubten Unpräzision muss man also erwarten, dass nur 68 % der Messwerte innerhalb der zukünftig erlaubten Grenzen liegen werden. Erwartet man bei der internen Kontrolle von Parametern, die in der Anlage gelistet sind (für die Gleichung f gilt), den gleichen 99%-Vertrauensbereich, wie er für die nicht gelisteten Analyte durch den Faktor 3 in Gleichung g definiert ist, so müsste man eine Unpräzision erreichen, die geringer als ein Drittel der bisher erlaubten ist. Aber auch wenn man mit dem bisherigen 95%-Vertrauensbereich arbeiten würde, müsste man die Unpräzision aller Verfahren auf die Hälfte reduzieren.

Die Vermeidung der oben genannten Probleme am Ende eines Kontrollzyklusses wird **durch eine extreme Steigerung der Anforderungen an die Präzision erkauft**. Nur wenn der Korrekturfaktor "z" auch auf die Analyte angewendet würde, die in der Anlage gelistet sind, ließe sich dieses Problem vermeiden.

Macdonald weist darauf hin, dass durch die ausschließliche Bewertung des Gesamtfehlers die Flexibilität im Umgang mit Unpräzision und Unrichtigkeit steigt. Innerhalb der festgelegten Grenzen trifft dies zu, beim Übergang von der alten auf die neue Richtlinie ließe sich diese Flexibilität zugunsten geringerer Ansprüche an die Unpräzision aber nur nutzen, wenn die Unrichtigkeit einen bisher deutlich zu großen Spielraum gelassen hätte.

Bei der Festlegung der Zahlenwerte für die erlaubte Messabweichung geht der Entwurf der RiliBÄK bei manchen Analyten auch noch über die hier skizzierten Festlegungen von Macdonald hinaus. Aufgrund einer nicht repräsentativen Laborbefragung, die von INSTAND e.V. und dem Referenzinstitut für Bioanalytik im Frühjahr 2005 bei Ringversuchsteilnehmern durchgeführt wurde, wurden nach Streichungen der jeweils schlechtesten 10 % dieser Labormeldungen Grenzen festgelegt, die noch höhere Anforderungen an die Analytik zur Folge haben würden.

Beispielhaft soll dies für die GOT (= AST = ASAT) dargestellt werden. Waren bisher eine Unpräzision von 6 % und eine Unrichtigkeit von 11 % sowie eine erlaubte Abweichung des Einzelwertes von 23 % festgelegt, so würde nach der Berechnung von Macdonald in Zukunft nur noch eine Abweichung des Einzelwertes von 12,5 % erlaubt sein. Dies würde bei einem 99 %-Vertrauensbereich - wie dargestellt - eine Unpräzision von weniger als 2 % erfordern. Im Entwurf der Richtlinie wird als maximal erlaubte Abweichung des Einzelwerts sogar nur 11,5 % statt 12,5 % gefordert!

Dies wirft zwangsläufig die Frage nach dem **medizinisch Notwendigen** auf. Antworten auf die Frage nach der notwendigen Präzision und Richtigkeit kann man auf sehr unterschiedliche Art und Weise gewinnen. Ein gut validiertes Verfahren ist die Ableitung aus der intra- und interindividuellen Streuung des Analyten. So kann die maximal erlaubte Unpräzision nach Cotlove/Harris (9) und die maximal erlaubte Unrichtigkeit nach Gowans et al. (10) wie folgt berechnet werden:

$$RE_a = 0,5 * intra$$

$$\text{Bias}_a = 0,25 * \sqrt{\text{intra}^2 + \text{inter}^2}$$

wobei die intra- und interindividuelle Streuung der meisten Analyte der Übersicht von Sebastian-Gambaro et. al. (11) oder einer Internetdatenbank (12) entnommen werden können. Für die oben genannte GOT ergibt sich aus der intraindividuellen Streuung von 11,6 % und der interindividuellen Streuung von 13,6 % eine maximal erlaubte Unpräzision von 5,8 %, Unrichtigkeit von 4,46 % und somit bei 95 %-Vertrauensbereich eine maximal erlaubte Abweichung des Einzelwertes von 16 %.

Die **Orientierung am medizinisch Erforderlichen** scheint zwingend geboten, da der Maßstab des technisch Machbaren und seine Ermittlung durch Laborbefragung nicht die Grundlage einer Richtlinie sein darf (SGB V § 12). Eine Überprüfung des Erforderlichen darauf hin, ob es überhaupt technisch realisierbar ist, kann durch eine solche Befragung sicherlich durchgeführt werden.

Für eine Reihe von Laborparametern wurde in Tabelle 1 die am Beispiel der GOT illustrierte Berechnung durchgeführt. Man erkennt für einige Analyte eine gute Übereinstimmung zwischen der Forderung der RiliBÄK und der klinischen Notwendigkeit (Gruppe A). Man erkennt aber in der Gruppe B auch, dass für viele Analyte erheblich überzogene Anforderungen bestehen. Schließlich wird bei der Gruppe C deutlich, dass für manche Analyte die medizinisch notwendige Präzision technisch noch nicht erreichbar ist. Dies trifft auch für die point of care-Diagnostik (POCT) zu, bei der die im Labor erreichbare Präzision und Richtigkeit vielfach nicht erreicht werden kann.

Der dargestellte Ansatz zur Herleitung der analytischen Anforderungen aufgrund der medizinischen Erfordernisse kann nur beispielhaft sein. Ebenso ist eine Begründung aufgrund des Referenzbereiches oder der Entscheidungsgrenzen möglich (13). Eine rationale Begründung der Vorgaben wird um so eher gelingen, je breiter die Diskussion darüber geführt wird.

1. Clin. Lab. 2006;52: 103-104 (Heft 1+2)
2. Clin. Lab. 2006;52: 319-322 (Heft 5+6)
3. Clin. Lab. 2006;52: 439-440 (Heft 7+8)
4. Shewhart W.A. Economic control of quality of the manufactured product. New York NY Van Nostrand 1931
5. Levey S., Jennings E.R. The use of control charts in the clinical laboratory. Am J. Clin. Pathol. 1950; 1059-1066
6. Westgard J.O., Groth T., Aronsson T., Falk H., de Verdier C-H.: Performance characteristics of rules for internal quality control probabilities for false rejection and error detection. Clin Chem 1977, 1857-1867
7. Westgard J.O., Groth T., Aronsson T., de Verdier C-H.: Combined Shewhard-Cusum control chart for improved quality control in clinical chemistry. Clin Chem 1977 1881-1887
8. Macdonald R.: Quality assessment of quantitative analytical results in laboratory medicine by root mean square of measurement deviation. J. Lab. Med. 2006, 111-117
9. Cotlove E., Harris E.K., Williams G.Z.: Biological and analytical components of variation in long-term studies of serum constituents in normal subjects III Physiological and medical implications. Clin. Chem. 1970, 1028-32

10. Gowans EMS, Hylthoft Petersen P., Blaabjerg O, Horder M.: Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J. Clin. Lab. Invest.* 1988, 757-764
11. Sebastian-Gâmbaro Á., Lirón-Hernández F.J., Fuentes-Arderiu X.: Intra- and Inter-Individual biological variability data bank. *Eur. J. Clin. Chem. Biochem.* 1997, 845-852
12. Ricos C. et. al: <http://www.westgard.com/biodatabase1.htm>
13. Fraser C.G., Hylthoft Petersen P., Ricos C., Haeckel R.: Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. 1992, 311-317

Abb. 1: Bei einem 3s-Vertrauensbereich liegen 0,135 % aller Werte oberhalb und 0,135 % aller Werte unterhalb der Grenzen: Wahrscheinlichkeit für eine irrtümliche Fehleranzeige

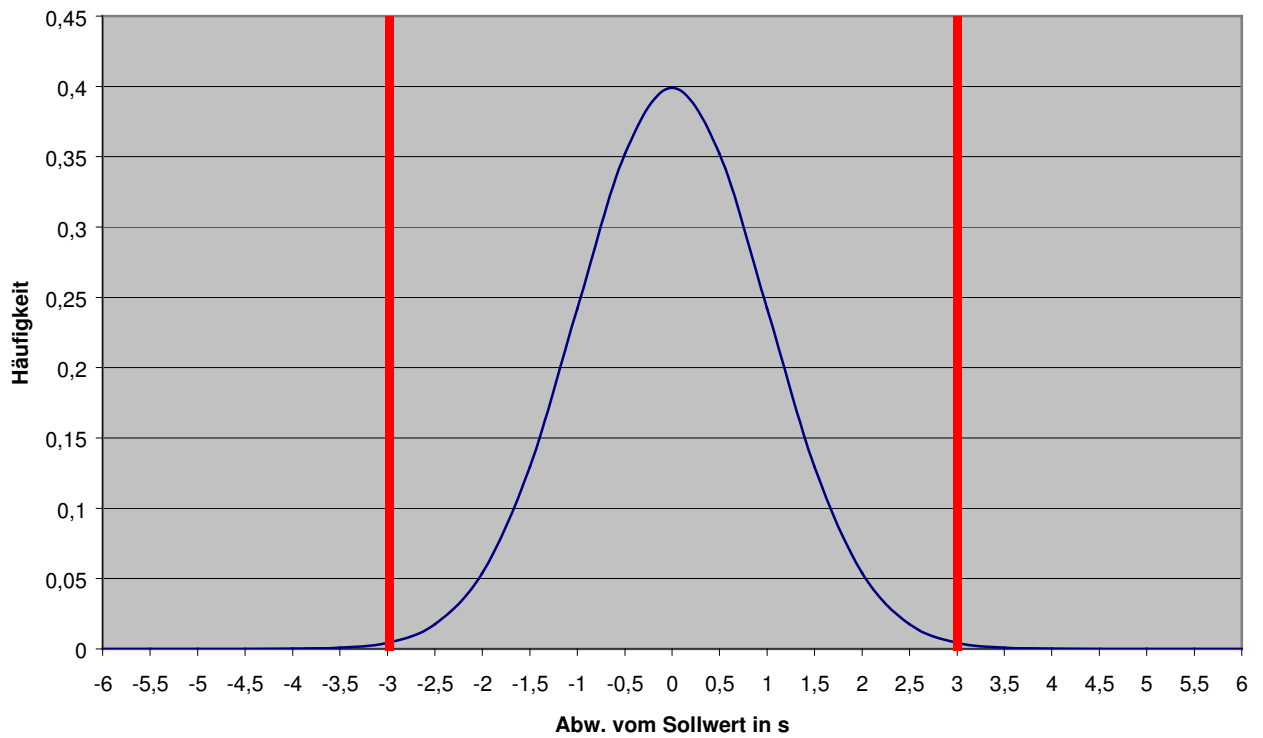


Abb. 2.: Eine Drift in der Größenordnung von 3s wird nur in der Hälfte der Fälle erkannt.

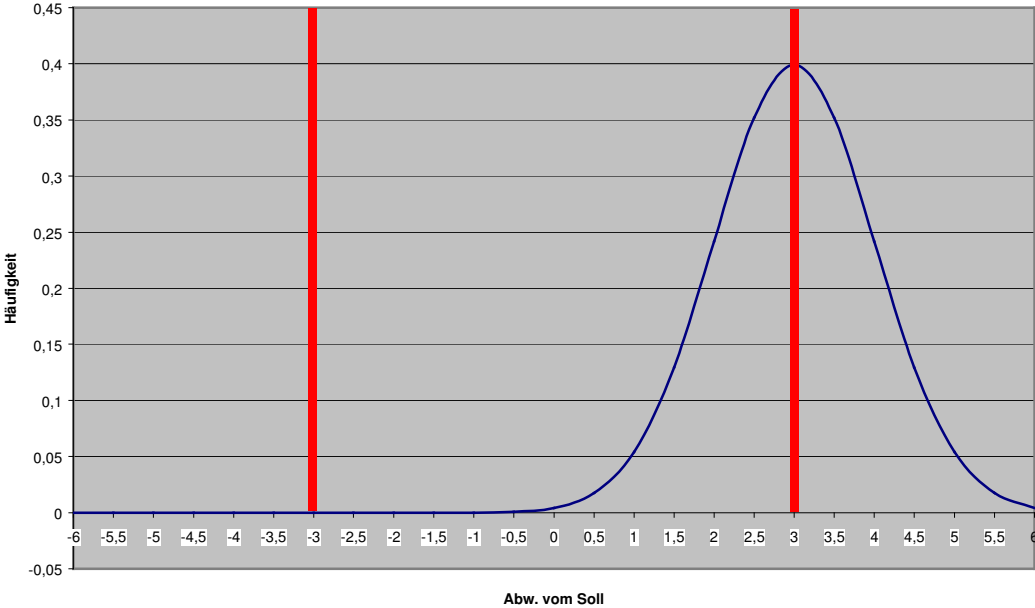


Abb. 3: Erhöht sich die Unpräzision um das Dreifache, so wird dieser Fehler in einem Drittel der Fälle erkannt.

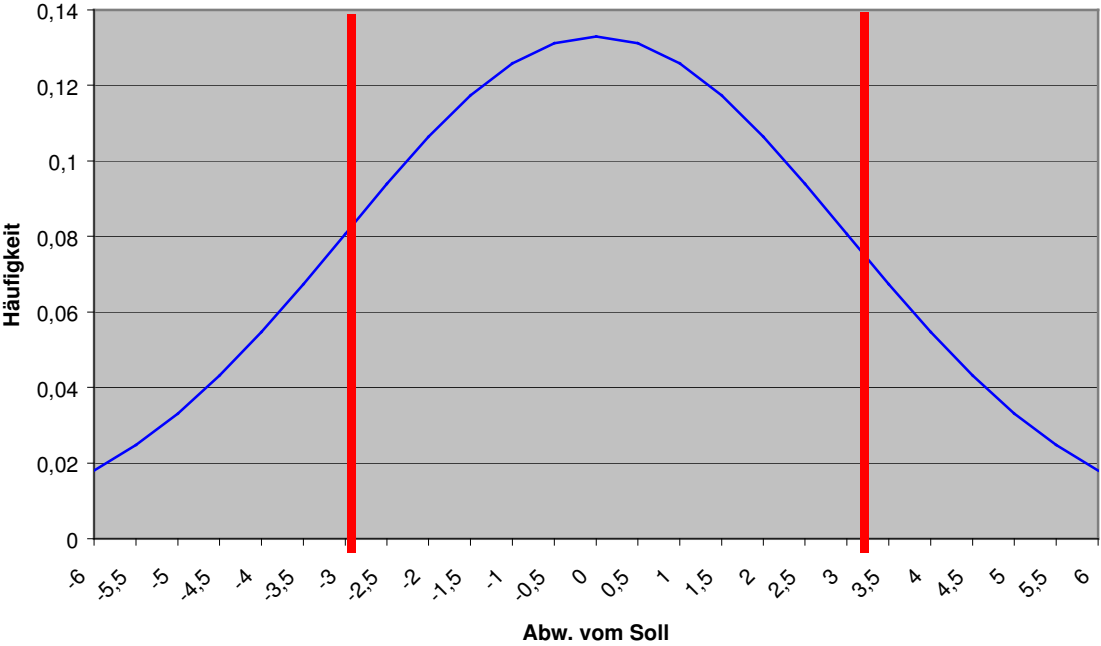


Tabelle 1: Macd = berechnet nach Macdonald aus erlaubter Unpräzision und Richtigkeit der RiliBÄK 2003; med. erf. = berechnet aufgrund der Intra- und Intersubjektiven Variabilität; Entwurf = Vorläufiger Vorschlag im Entwurf der RiliBÄK (Stand August 2006)

	Analyt	Matrix	max. Abweichung des Einzelwertes		
			Macd	med. erf.	Entwurf
	pH	Blut	0,0	4,5	0,0
	Natrium	Urin	6,4	40,6	6,5
	Kalium	Urin	8,6	51,4	8,5
	CRP (C-reaktives Protein)	Blut	13,6	76,0	13,5
	Phosphat (anorganisch)	Urin	11,7	58,9	11,5
	Calcium	Urin	8,6	39,5	8,5
	Protein, Gesamt-	Urin	11,3	50,2	11,5
	Kreatinin	Urin	12,2	48,5	12,0
Gruppe B	Triglyceride (Gesamt-Glycerin)	Blut	9,4	34,8	9,0
	Creatinkinase (CK) EC 2.7.3.2	Blut	11,2	42,3	11,0
	Lipase	Blut		44,2	12,0
	Albumin	Urin	14,4	48,1	14,5
	Lactat	Blut	10,8	35,1	11,0
	Harnstoff	Urin	13,9	41,4	13,5
	Glutamat-Pyruvat-Transaminase (GPT, ALT)	Blut	12,5	34,7	11,5
	Harnsäure	Urin	13,9	34,7	13,5
	Thyreotropes Hormon (TSH)	Blut	13,5	28,8	13,5
	Cortisol	Blut	17,8	33,4	16,0
	Estradiol, 17-beta	Blut	22,2	44,5	22,0
	Carcinoembryonales Antigen	Blut		28,2	14,0
	Gamma-Glutamyl-Transferase (Gamma-GT)	Blut	11,7	22,8	11,5
	Thrombozyten	Blut	12,7	15,2	8,5
	Harnsäure	Blut	7,2	12,3	7,0
	Leukozyten	Blut	6,4	11,2	6,5
	Harnstoff	Blut	10,6	16,8	10,5
	Bilirubin (gesamt)	Blut	22,4	33,9	22,0
	Hämoglobin	Blut	2,8	5,1	3,5
	Glutamat-Oxalacetat-Transaminase (GOT bzw. AST)	Blut	12,5	16,0	11,5
	Kalium	Blut	4,7	6,2	4,5
	Immunglobulin M	Blut	14,4	17,9	13,0
	Cholesterin (gesamt)	Blut	7,2	9,3	7,0
	Ferritin	Blut	13,5	17,4	13,5
	Lactat-Dehydrogenase (LDH) EC 1.11.1.27	Blut	9,4	11,3	9,0
	Immunglobulin A	Blut	12,2	14,6	12,0
	Phosphat (anorganisch)	Blut	9,4	10,9	9,0
	Glucose	Blut	7,8	8,5	8,0
	pCO2	Blut	6,5	6,5	6,5
	Erythrozyten	Blut	3,9	3,9	4,0
	Alkalische Phosphatase (AP) EC 3.1.3.1	Blut	13,0	11,6	13,0
	Immunglobulin G	Blut	10,0	8,5	10,0

	pCO2	Blut	6,5	5,5	6,5
	Trijodthyronin, gesamt	Blut	14,9	12,5	15,0
	Testosteron	Blut	20,6	15,4	20,5
	Thyroxin, gesamt	Blut	12,8	9,3	12,5
Gruppe C	Protein, Gesamt-	Blut	5,8	3,9	6,0
	Magnesium	Blut	7,8	4,8	7,5
	Kreatinin	Blut	11,7	7,1	11,5
	Calcium (gesamt)	Blut	5,8	2,4	6,0
	Transferrin	Blut		3,6	9,5
	Chlorid	Blut	4,7	1,7	4,5
	Albumin	Blut	12,5	4,1	12,5
	Thromboplastinzeit (%)	Blut	11,4	3,4	11,5
	Natrium	Blut	3,1	0,8	3,0
	aPTT (Aktivierte partielle Thromboplastinzeit)	Blut	10,8	2,2	10,5