



## Entwurf der RiLiBÄK 2006 – welchen Qualitätsansprüchen genügt die Richtlinie selbst?

Eine Richtlinie ist ein Handlungsvorschrift mit bindendem Charakter, aber nicht gesetzlicher Natur. Sie wird jedoch aufgrund einer Verordnung ausgegeben, die auf einem Gesetz beruht. Wegen dieser weitreichenden Bedeutung ist an eine Richtlinie der Anspruch zu stellen, dass die in ihr formulierten Ziele objektiv begründbar und – sofern es sich bei den darin gestellten Anforderungen um berechenbare Größen handelt - mathematisch korrekt wiedergegeben werden. Gerade in einer Zeit, in der auch die Politik erkennt, dass ein Bürokratieabbau vonnöten ist, sollte eine Richtlinie sich auf die unabdingbar notwendigen Inhalte beschränken.

Bereits in der zur Zeit gültigen „Richtlinie der Bundesärztekammer zur Qualitätssicherung quantitativer laboratoriumsmedizinischer Untersuchungen“ (RiLiBÄK) steht die Liste der Verfahren, die einer Nachprüfung durch die Behörden unterworfen sind, in einem Anhang. Eine wissenschaftliche Grundlage für die Auswahl dieser Verfahren gibt es nicht. Entscheidend für deren Auswahl war die Verfügbarkeit eines Referenz- oder Sollwertmethodenkonzeptes und die Möglichkeit der Durchführung externer Ringversuche.

Die Grenzen für die zulässige Impräzision oder die zulässige Unrichtigkeit wurden ursprünglich willkürlich festgelegt, anschließend aufgrund einer Anwenderbefragung korrigiert. Auch die zukünftige RiLiBÄK, deren Entwurf zur Zeit intensiv diskutiert wird, wird einen solchen Anhang besitzen, und wieder wird er sich an den Ergebnissen einer Anwenderbefragung orientieren.

Ob diese Anwenderbefragung repräsentativ ist oder ob sie überhaupt wahrheitsgemäße Angaben wiedergibt, soll hier nicht zur Diskussion stehen. Erstaunlich ist der Umstand, dass sich der Ausschuss der Experten, die von der Bundesärztekammer (BÄK) hierfür berufen wurden, auf ein derartiges Verfahren einlässt, gibt es doch einen mathematischen Zusammenhang zwischen dem klinischen Entscheidungsintervall und dem bei der Messung erforderlichen Variationskoeffizienten (VK). Ob das Entscheidungsintervall dem Referenzbereich entnommen oder aus der intra- und interindividuellen Streuung berechnet wird, ist dabei unerheblich. Entscheidend ist die Berechenbarkeit der Wahrscheinlichkeit einer diagnostische Fehlentscheidung aufgrund eines zu großen VK, wie sie von Westgard<sup>1</sup> und Hylthoft<sup>2</sup> gezeigt wurde. Haeckel und Wosnick<sup>3</sup> und viele weitere Autoren haben dies an zahlreichen Beispielen demonstriert. Die hierfür notwendige Kenntnis der intra- und interindividuellen Streuung ist kein Problem, weil auch diese publiziert sind<sup>4,5</sup>. Solche mathematisch berechenbaren VK sollten nicht durch praktische Erfahrungswerte ersetzt werden.

Das SGB V schreibt vor, die Patientenversorgung auf das Notwendige und Ausreichende zu begrenzen. Da Laboruntersuchungen stets in der Trias von Genauigkeit, Schnelligkeit und Kosten optimiert werden müssen, ist eine übertriebene Forderung an die Präzision gleichbedeutend mit einer Verschlechterung der Verfügbarkeit und/oder der Kosten. Das klinisch Notwendige kann nicht durch eine Anwenderbefragung ermittelt werden, sondern muss aus den wissenschaftlich verfügbaren Daten abgeleitet und im Konsens einer breiten, öffentlichen Diskussion festgelegt werden. Eine übertriebene technische Präzision zu fordern, nur weil sie machbar ist, verstößt gegen das Wirtschaftlichkeitsprinzip.

Da Messfehler grundsätzlich normalverteilt sind, kann die Wahrscheinlichkeit, eine systematische Drift  $S$  (shift oder bias, ausgedrückt als Vielfaches der Methodenvarianz  $\sigma$ ) mit Hilfe von  $N$  Kontrollproben dadurch zu entdecken, dass mindestens eine davon außerhalb der berechneten Fehlergrenze  $G$  gemessen wird, angegeben werden als

$$P = \sum_{k=1}^n \binom{n}{k} * p^k * (1 - p)^{n-k} = \text{BINOMVER}(1; N; p; -) + \dots + \text{BINOMVER}(k; N; p; -)$$

wobei die Einzelwahrscheinlichkeit dafür, dass eine Kontrolle außerhalb des erlaubten Bereiches ist, dann aus der Gaußschen Normalverteilung mit

$$p = 1 - \left(1 / \sqrt{2\pi}\right) * \int_{-\infty}^{G-S} e^{-0,5 * (t)^2} dt + 1 - \left(1 / \sqrt{2\pi}\right) * \int_{-\infty}^{G+S} e^{-0,5 * (t)^2} dt$$

$$= 1 - \text{NORMALVERT}(G - S; 0; 1; +) + 1 - \text{NORMALVERT}(G + S; 0; 1; +)$$

berechnet werden kann. Die jeweils zweite Schreibweise der Formeln entspricht der Schreibweise im Tabellenkalkulationsprogramm „Microsoft Excel“.

Analog erhält man die Einzelwahrscheinlichkeit, einen zufälligen Fehler R (wieder als Vielfaches von  $\sigma$  ausgedrückt) zu entdecken, aus der Gleichung

$$p = 1 - 2 * \left(1 / R * \sqrt{2\pi}\right) * \int_{-\infty}^G e^{-0,5 * (t/R)^2} dt$$

$$= 1 - 2 * \text{NORMALVERT}(G; 0; R; +)$$

Die Wahrscheinlichkeit dafür, eine fehlerfreie Analysenserie irrtümlich wiederholen zu müssen (probability of false rejection), kann ebenfalls berechnet werden, wie Westgard<sup>1</sup> es dargestellt hat. Diese Wahrscheinlichkeit leitet sich aus der ersten Gleichung ab, wenn die Drift gleich Null gesetzt wird.

Die bisherigen und vermutlich auch zukünftigen Vorgaben der RiLiBÄK sehen  $N = 2$  (Kontrollproben) und  $G = 3$  (Fehlergrenze) vor. Mit den oben dargestellten Gleichungen kann ausgerechnet werden, wie effizient dieses Kontrollsystem ist. Einen systematischen Fehler, der doppelt so groß wie die Methodenvarianz ist, wird hiernach lediglich in einem Drittel der Fälle entdeckt; einen zufälligen Fehler der gleichen Größenordnung sogar nur in einem Viertel der Fälle. Selbst einen systematischen Fehler von  $3\sigma$  übersieht man immer noch in einem Viertel und einen zufälligen Fehler dieser Größe sogar in der Hälfte der Fälle. Die laborinternen Grenzen dieses Kontrollkonzeptes sind somit wenig tauglich zur Überwachung der Messungen und sollten dringend abgeschafft werden.

Wesentlich effizienter erscheinen die Vorgaben für die zulässige Impräzision und Unrichtigkeit, wenn die Kontrollmessungen am Ende der Kontrollzyklen bewertet werden. Angesichts der größeren Zahl an Messungen, die in die Berechnung der Impräzision und Unrichtigkeit eingehen, sind die Wahrscheinlichkeiten, Fehler aufzudecken deutlich größer. Überschreitungen der Größenordnung  $2\sigma$  werden mit nahezu 100 % Wahrscheinlichkeit erkannt.

Im bislang letzten Entwurf wurde auf dieses effiziente Kontrollverfahren verzichtet und dafür die Berechnung der mittleren quadratischen Messabweichung aus 7 Messungen neu eingeführt. Hierdurch sinkt die Wahrscheinlichkeit, eine Drift oder eine Zunahme der Streuung zu entdecken deutlich, auch wenn hierfür engere Grenzen festgelegt werden. Gleichzeitig steigt die Zahl der irrtümlich verworfenen Messreihen an. Das neue Verfahren kann noch nicht verlässlich beurteilt werden, weil die Definition der Eingriffsgrenzen fehlt und weil in den Dokumentationsvorschriften Werte enthalten sind, auf deren Berechnung zuvor verzichtet wurde.

In der noch gültigen RiLiBÄK wurde die maximale zulässige Abweichung des Einzelwertes (Spalte 7) berechnet nach der Formel: Zweifache Unpräzision plus Unrichtigkeit. Die Verdoppelung der Unpräzision in dieser Berechnung bedeutet, dass man in der Formel

$$\text{Gesamtfehler} = \text{Unpräzision} + z * \text{Unrichtigkeit}$$

den z-Wert gleich 2 gewählt hat, was einer Irrtumswahrscheinlichkeit von 0,05 entspricht. Es sollten also 95 % aller Messungen innerhalb des Intervalls liegen, wenn sie den Vorgaben hinsichtlich Präzision und Richtigkeit entsprechen. Umgekehrt betrachtet liegen aber 5 % der Messungen der Qualitätskontrolle außerhalb der Grenzen – die Messungen müssen also verworfen werden !

Eine Reduzierung irrtümlich verworfener Analysenserien auf 3 % ( $z = 3$ ), 1 % ( $z = 3,5$ ) oder 0,1 % ( $z = 4$ ) hätte einen sehr großen zulässigen Bereich zur Folge, sofern nicht gleichzeitig die Anforderungen an Präzision und Richtigkeit überdacht oder die gesamte Kontrollstrategie geändert werden. Mit einer Vergrößerung des zulässigen Bereiches sinkt gleichzeitig die Empfindlichkeit der Qualitätskontrolle zur Erkennung auftretender Fehler.

Im Entwurf der neuen Richtlinie wird die Spalte 7 nicht mehr nach obiger Formel berechnet, sondern als Wurzel aus der Summe der quadratischen Unpräzision und der quadratischen Unrichtigkeit und stellt damit die Vorgabe für die mittlere quadratische Abweichung dar:

$$\text{Spalte 7} = \sqrt{\text{Unpräzision}^2 + \text{Unrichtigkeit}^2}$$

Eigentlich berechnet sich die mittlere quadratische Abweichung, wie im Glossar des Teiles A der neuen Richtlinie dargestellt, nach der Formel:

$$\Delta = \sqrt{\text{Unrichtigkeit}^2 + \left\{ n / (n-1) \right\} * \text{Unpräzision}^2}$$

Die mittlere quadratische Abweichung soll aus 7 Messwerten berechnet werden. Die Spalte 7, die den Faktor  $n/(n-1) = 7/6$  unberücksichtigt lässt, stellt damit hinsichtlich der Unpräzision um 16 % höhere Ansprüche, als mathematisch begründet wären. Da die Vorgaben für Unpräzision und Unrichtigkeit aus einer oben bereits erwähnten Anwenderbefragung ermittelt wurden, ist mit diesem höheren Anspruch auch eine entsprechende Quote an Laboren verknüpft, die den Ansprüchen nicht mehr genügt .

Im derzeitigen Entwurf der Richtlinie ist das Schicksal der alten Spalte 7 unklar. Wird sie durch die neue Spalte ersetzt, dann hätte dies zur Folge, dass auch der Einzelwert gemäß dieser Grenze beurteilt wird. Diese deutliche Reduzierung des zulässigen Fehlers des Einzelwertes ist aber weder mathematisch zu begründen noch aus der Anwenderbefragung abzuleiten. Sie wäre einfach die Folge dessen , dass die Spalte 7 neu für einen anderen Zweck berechnet wurde.

Zwei Beispiele sollen die dargestellten Zusammenhänge verdeutlichen:

Die noch gültige RiLiBÄK schreibt für den Analyten „Eisen“ eine maximal zulässige Unpräzision von 4 % vor. Angesichts eines Referenzbereiches von ca.  $\pm 65$  % und einer circadianen Rhythmik von über 50 % und zudem einer Monatsrhythmik bei Frauen, ist dies ein groteskes Beispiel überzogener Vorgaben. Hier übersteigt die Festlegung der Grenzen für die Messabweichung die medizinischen Erfordernisse bei weitem.

Des weiteren schreibt die derzeit gültige RiLiBÄK für Glucose im Plasma/Serum eine maximal zulässige Unpräzision von 5 %, eine maximal zulässige Unrichtigkeit von 6 % vor und erlaubt eine maximale Abweichung des Einzelwertes von 16 %. Im aktuellen Entwurf sind diese Werte unverändert. Bevor die RiLiBÄK 2003 überarbeitet und korrigiert wurde, waren für die Unpräzision 4 %, für die Unrichtigkeit 7 % und für den Einzelwert 15 % angegeben. Angesichts eines Referenzintervalls von  $\pm 26$  % ist der Anteil in Wirklichkeit hypoglykämischer Patienten, deren Messung jedoch eine Hyperglykämie vortäuscht, durchaus nicht vernachlässigbar. Würde man die Vorgaben aus der intraindividuellen Streuung von 5,7 % und der interindividuellen Streuung von 6,9 % ableiten<sup>5</sup>, so dürften die Unpräzision höchstens 2,9 %, die Unrichtigkeit höchstens 2,2 % und damit der Gesamtfehler höchsten 6,9 % betragen.

Diese Genauigkeitsanforderung ist für diagnostische Zwecke nach Ansicht der Autoren begründet. Für die Verlaufskontrolle des Diabetikers ist das Entscheidungsintervall nicht aus dem Referenzbereich oder der inter- und intraindividuellen Streuung ableitbar. POC-Geräte, die nur diesem Zweck dienen, müssen geringeren Anforderungen unterworfen sein. Der aktuelle Vorschlag seitens der DGKL, für Vollblutmethoden die Serumvorgaben zu verwenden, weil jedes Gerät erst nach Abtrennung der Erythrozyten misst, entbehrt der Orientierung am klinisch Notwendigen.

Es wäre zu begrüßen, wenn das Konzept der Richtlinie der Bundesärztekammer auf die wissenschaftlich begründeten Fakten und die durch sie gestellten Anforderungen auf das Ausreichende und klinisch Notwendige reduziert würden. Hierzu wäre erforderlich:

1. Streichung des Teils A <sup>6</sup>
2. Streichung der Vorperiode.
3. Ableitung der zulässigen VKs für Unpräzision und Unrichtigkeit aus den Referenzbereichen oder den intra-/interindividuellen Streuungen.

Die Befragung der Anwender kann bestenfalls dazu dienen, festzustellen, bei welchen Parametern die klinisch erforderlichen Vorgaben von vielen Laboren nicht erreicht werden und Übergangsvorschriften sinnvoll erscheinen. Das Notwendige nicht zu fordern, weil es vielfach nicht erreicht wird, ist ebenso unbegründet wie die Forderung unnötig hoher Genauigkeit, wenn sie nicht geboten ist.

Vielleicht würde die Bundesärztekammer das hier dargestellte Ziel – wissenschaftlich begründete Vorgaben im gesetzlichen Rahmen der Wirtschaftlichkeit - schneller erreichen, wenn die Entwürfe zur Neufassung der Richtlinie in einer breiteren Öffentlichkeit diskutiert würden. Leider aber tagen die Experten stets hinter verschlossenen Türen. Wenn das Beratungsergebnis bei einer öffentlichen Anhörung den Anwenderkreisen vorgestellt wird, dann ist dies vollkommen unverbindlich, da als Autor der RiLiBÄK der Wissenschaftliche Beirat in Erscheinung tritt, der an keine Meinung gebunden ist. So konnte bei der letzten Fassung der RiLiBÄK zum Thema „Liquor“ ein Beschluss gefasst und nachträglich gerechtfertigt werden, der einerseits mit fehlerhaften Argumenten begründet war und andererseits den fundierten Widerspruch mehrerer Fachgesellschaften erfuhr. Ebenso wurden die hämatologischen Untersuchungen in die Verpflichtung zur Erstellung laboratoriumsinterner Fehlergrenzen einbezogen obwohl die Kontrollproben in der Regel nur 6 Wochen haltbar sind. Erst nach zahlreichen Einsprüchen wurde dies später offiziell im Kommentar zur RiLiBÄK korrigiert. Eine Richtlinie zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen sollte in ihrer Erstellung höheren Qualitätsansprüchen unterliegen.

Es bleibt zu hoffen, dass die BÄK einer öffentlichen, transparenten Entwicklung der RiLiBÄK den Vorrang geben wird, da nur dadurch sicher gestellt werden kann, dass nicht Vorgaben veröffentlicht und über das Medizinproduktegesetz bindend werden, die aufgrund des höherwertigen SGB V ohne praktische Bedeutung bleiben werden.

Die Bedeutung der RiLiBÄK hat in den letzten Jahren nach Ansicht vieler Anwender deutlich an Wert verloren. Eine Orientierung an den klinischen Erfordernissen würde bei der Umsetzung der Richtlinie von Vorteil sein und ihre Akzeptanz allgemein erhöhen.

W. Bauersfeld  
W. Dick  
N. Gässler  
C. Kaiser

1. [www.westgard.com](http://www.westgard.com)
2. Hylthoft, P. et. al., Clin. Chem. Clin. Biochem. 1996;91; 401-7
3. Haeckel, R. und Wosnick, W., Clin Chem. Lab. Med. 2004; 42; 198-203
4. Angeles, M et. al., Eur. J. Clin. Chem. Lab. Med. 1997; 35, 845-852
5. Ricos, C. et al., Scand. J. Clin. Lab. Invest. 1999; 59: 491 - 500
6. BNLD-Informationen, Clin.Lab. 2006;52: 103 - 104