

Georg Hoffmann*, Ralf Lichtinghagen und Werner Wosniok

Ein einfaches Verfahren zur Schätzung von Referenzintervallen aus routinemäßig erhobenen Labordaten

Simple estimation of reference intervals from routine laboratory data

DOI 10.1515/labmed-2015-0082

Eingang 3.9.2015; Akzeptanz 22.10.2015; vorab online veröffentlicht 17.11.2015

Zusammenfassung: Nach Empfehlungen der IFCC und anderer Organisationen sollen medizinische Laboratorien alle Referenzintervalle selbst erheben oder zumindest individuell anpassen, um Besonderheiten der jeweiligen Methoden und Patientenkollektive abzubilden. In der Praxis ist diese Forderung allerdings nur schwer umzusetzen. Deshalb erarbeiten zwei DGKL-Arbeitsgruppen („AG Richtwerte“ und „AG Bioinformatik“) Methoden zur Schätzung von Referenzintervallen aus routinemäßig erhobenen Labordaten. In der vorliegenden Arbeit beschreiben wir ein visuelles Verfahren, das direkt auf einem Excel-Blatt ohne Programmierkenntnisse durchgeführt werden kann. Die Patientenwerte werden dafür in einem Punktediagramm gegen Quantile der Standardnormalverteilung aufgetragen (sog. QQ-Plot unter Verwendung der NORM.INV-Funktion von Excel). Falls die untersuchte Population vorwiegend nicht-krankte Personen mit annähernd normalverteilten Werten enthält, bilden deren Punkte eine Gerade. Häufig sind die Werte eher lognormalverteilt; dann erkennt man die Gerade nach Logarithmierung der Originalwerte. Werte, die nicht zur angenommenen theoretischen Verteilung passen, sind in dieser Darstellung durch eine Abweichung von der Geradenform gut erkennbar und können eliminiert werden. Aus den bereinigten Werten werden dann Mittelwert und Standardabweichung

berechnet, um das Referenzintervall ($\mu \pm 2\sigma$) zu schätzen. Die Methode liefert mit simulierten Werten und Echtdaten plausible Ergebnisse. Mit steigendem Anteil von Werten, die nicht zum Modell passen, tendiert das Verfahren zur Unterschätzung der Standardabweichung. In allen Fällen, in denen der QQ-Plot keine eindeutige Gerade liefert, ist das vorgeschlagene Verfahren nicht anwendbar.

Schlüsselwörter: indirekte Schätzung; MS Excel; Quantil-Quantil-Plot; Referenzintervall.

Abstract: According to the recommendations of the IFCC and other organizations, medical laboratories should establish or at least adapt their own reference intervals, to make sure that they reflect the peculiar characteristics of the respective methods and patient collectives. In practice, however, this postulate is hard to fulfill. Therefore, two task forces of the DGKL („AG Richtwerte“ and „AG Bioinformatik“) have developed methods for the estimation of reference intervals from routine laboratory data. Here we describe a visual procedure, which can be performed on an Excel sheet without any programming knowledge. Patient values are plotted against the quantiles of the standard normal distribution (so-called QQ plot) using the NORM.INV function of Excel. If the examined population contains mainly non-diseased persons with approximately normally distributed values, the respective dots form a straight line. Quite often the values are rather lognormally distributed; in this case the straight line can be detected after logarithmic transformation of the original values. Values, which do not match with the assumed theoretical distribution, deviate from the linear shape and can easily be identified and eliminated. Using the reduced data set, the mean value and standard deviation are calculated and the reference interval ($\mu \pm 2\sigma$) is estimated. The method yields plausible results with simulated and real data. With

***Korrespondenz:** Prof. Dr. med. Georg Hoffmann, Trillium GmbH, Medizinischer Fachverlag, Jesenwanger Straße 42b, 82284 Grafrath, Deutschland, Tel.: +49-(0)-8144/93905-0,

Fax: +49-(0)-8144/9390-29, E-Mail: georg.hoffmann@trillium.de

Ralf Lichtinghagen: Institut für Klinische Chemie der Medizinischen Hochschule Hannover (MHH), Hannover, Deutschland

Werner Wosniok: Institut für Statistik, Department 3 der Universität Bremen, Bremen, Deutschland

increasing number of results, which do not match with the model, it tends to underestimate the standard deviation. In all cases, where the QQ plot does not yield a substantial linear part, the proposed method is not applicable.

Keywords: indirect estimation; MS Excel; quantile quantile plot; reference interval.

Einleitung

Der größte Teil aller labormedizinischen Entscheidungen basiert auf dem Vergleich von Messwerten mit Referenzintervallen. Diese legen fest, welche Resultate als „normal“ – oder präziser ausgedrückt, welche Personen als „nicht krank“ – einzustufen sind. Aufgrund seiner enormen Bedeutung für Diagnostik, Prognostik, Verlaufs- und Therapiekontrolle sollte man annehmen, dass der Begriff Referenzintervall eindeutig definiert ist und dass die Verfahren für die Ermittlung der Entscheidungsgrenzen exakt festgelegt sind. Theoretisch ist dies nach der derzeit gültigen IFCC-/CLSI-Leitlinie von 2008 [1] auch der Fall:

- Das Referenzintervall ist definiert als der zentrale 95%-Bereich der Werte (*observed values*) einer gesunden (*apparently healthy*) Population.
- Jedes medizinische Laboratorium ist angehalten, Referenzintervalle für alle angebotenen Analyte an mindestens 120 Referenzindividuen (*reference individuals*) durch Bestimmung der 2,5- und 97,5-Perzentile selbst zu ermitteln oder zumindest existierende Angaben anhand von 20 gesunden Kontrollpersonen zu überprüfen.

Nach DIN EN ISO 15189 muss jedes Laboratorium die Grundlagen der entsprechenden Entscheidungswerte dokumentieren und den Nutzern mitteilen [2].

In der Praxis wird von diesen Vorgaben jedoch häufig abgewichen. Zum einen erweist sich der Begriff „gesund“ als unscharf (beispielsweise bei alten Menschen), zum anderen erscheint die Forderung, dass jedes medizinische Laboratorium für Hunderte oder Tausende angebotener Tests geeignete Referenzindividuen untersuchen soll, unrealistisch und teilweise – etwa bei gesunden Neugeborenen – sogar unethisch [3, 4]. Auch wenn die Forderung nach 120 Referenzindividuen auf den ersten Blick maßvoll erscheint, so ist diese Zahl bei geschlechtsabhängigen Werten mit 2 und bei Altersabhängigkeit mit noch höheren Faktoren zu multiplizieren, sodass man im Einzelfall durchaus tausend und mehr Personen benötigen kann. Deshalb übernehmen die meisten Laboratorien Angaben von Diagnostika-Herstellern oder anderen Autoren ohne

die geforderte Prüfung und Dokumentation. Dabei kann es zu erheblichen Diskrepanzen zwischen verschiedenen Quellen kommen [5]. So unterscheiden sich zum Beispiel die Referenzintervalle für Elektrolyte oder Proteine auf viel besuchten Internetseiten um bis zu 30% (Tabelle 1).

Einen Ausweg aus dem Dilemma bieten indirekte Verfahren, die Referenzintervalle aus den in der täglichen Routine erhobenen Laborresultaten abzuschätzen versuchen. Die Grundidee besteht darin, Werte zu eliminieren, die aufgrund von Modellannahmen nicht zur erwarteten Verteilung passen. Aus den verbleibenden Werten werden dann theoretische Perzentilen berechnet. Die Grundlagen für diese Verfahren wurden bereits um 1960 gelegt (Übersicht bei Ref. [6]), doch erst mit der Einführung des Personal Computers in den 1980er-Jahren stand ausreichend Rechenkapazität für eine allgemeine Anwendung zur Verfügung.

Besondere Erwähnung verdient das 2007 publizierte RLE-Verfahren (RLE=Reference Limit Estimator) von Arzideh et al. [7], das von der „Arbeitsgruppe Richtwerte“ der DGKL mithilfe des Statistikpakets R entwickelt wurde. Dieses mathematisch anspruchsvolle Verfahren basiert auf transformierten Rohdaten, berechnet die theoretische Dichtefunktion im Zentrum der Verteilung und eliminiert potenzielle Ausreißer aufgrund lokaler Abweichungen von dieser Funktion an sog. „truncation points“. Es kam erfolgreich zur Schätzung der Referenzintervalle von Enzymen [7], Elektrolyten [8] und Blutbildparametern [9] zum Einsatz und wurde auch in multizentrischen Studien evaluiert [10, 11]. Das RLE-Verfahren benötigt allerdings sehr große Datensätze von mehreren tausend Fällen, die nicht immer zur Verfügung stehen, und zudem erweist sich das Programmpaket R für Anwender ohne Informatik- und Statistikerfahrung oft als Einstiegshürde.

Aus diesem Grunde evaluierten wir im Rahmen einer Kooperation zwischen den DGKL-Arbeitsgruppen Richtwerte und Bioinformatik [12] weniger anspruchsvolle Verfahren, die mit geringem Aufwand und ohne Programmierkenntnisse durchführbar sein sollten. Unser

Tabelle 1: Im Internet publizierte Referenzintervalle für gebräuchliche Messgrößen (Stand 5.10.2015).

Quelle	Natrium, mmol/L	Calcium, mmol/L	Fibrinogen, mg/dL	Gesamteiweiß, g/dL
DocCheck.com	135–148	2,20–2,65	180–350	6,6–8,3
Laborlexikon.de	135–145	2,02–2,60	150–450	6,1–8,1
Thieme.de	135–150	2,3–2,6	200–400	6,0–8,4
Wikipedia.org	135–145	2,2–2,6	180–350	6,0–8,0

besonderes Augenmerk galt den Originalmethoden von J. Pryce 1960 [13] und R. Hoffmann 1963 [14], die auf dem Vergleich von Echtdateen mit theoretischen Normalverteilungen basierten und damals ohne Computer auf „Wahrscheinlichkeitspapier“ durchgeführt wurden [15]. In der vorliegenden Arbeit präsentieren wir ein Verfahren, bei dem das visuelle Vorgehen mit Microsoft Excel nachgebildet wird [16] und vergleichen es mit einem anderen „modifizierten Hoffmann-Verfahren“ [4] aus der jüngeren Literatur.

Material und Methoden

Für die Evaluation des Verfahrens standen Analysenergebnisse von 246 gesunden Probanden aus einer Studie der Medizinischen Hochschule Hannover (MHH) zur Verfügung. Die Grundlagen der Studie sind in Ref. [17] beschrieben. Als Bezugsverfahren für den Methodenvergleich diente das RLE-Verfahren der AG Richtwerte [7]; dafür erhielten wir validierte Patientendatensätze von Dr. F. Arzideh, Universität Bremen.

Die Berechnungen für Abbildung 1 und 9 wurden mit der englischen Version von R durchgeführt, alle anderen mit Microsoft Excel. Zwischen den verschiedenen Excel-Versionen bestanden keine grundsätzlichen Unterschiede; lediglich beim Aufruf der Funktion NORM.INV (s. u.) ist zu beachten, dass der Punkt in älteren Versionen (vor 2010) entfällt (NORMINV statt NORM.INV). Zur einfacheren Erzeugung von Zufallszahlen, Histogrammen

etc. empfiehlt es sich, das Add-in „Analyse-Funktionen“, das im Lieferumfang von Excel enthalten ist, zu installieren (eine Anleitung findet sich zum Beispiel unter <https://support.office.com>).

Wahrscheinlichkeitsnetze (sog. „Wahrscheinlichkeitspapier“) für die visuelle Prüfung von Verteilungen ohne Computer sind im Internet erhältlich (zum Beispiel <http://gpr.physik.hu-berlin.de/Downloads/Papiere.html>) (vgl. Abbildung 3). Es handelt sich dabei um eine Art Millimeterpapier, dessen Skala auf der y-Achse an eine Dichtefunktion angepasst ist. Im Fall der Gauß'schen Glockenkurve sind die Linien folglich in der Mitte bei einer Wahrscheinlichkeit von 50% am dichtesten und an den Rändern jenseits von 1% und 99% weit auseinandergezogen.

Für das Bezugsverfahren [7] wurde das frei zugängliche Statistikpaket R (www.r-project.org) mit den Zusatzpaketen *geoR* und *msm* eingesetzt. Zur vereinfachten Bedienung entwickelte die AG Richtwerte ein Excel-Frontend namens „Reference Limit Estimator“, das ebenfalls im Internet bezogen werden kann (www.dgkl.de, Arbeitsgruppe Richtwerte im Menüpunkt AGs & Sektionen).

Praktische Durchführung

Das nachfolgende Kapitel soll es dem Leser erleichtern, die dargestellten Ergebnisse selbst nachzuvollziehen und eigene Auswertungen durchzuführen. Basis des QQ-Plot-Verfahrens ist die NORM.INV-Funktion von Excel [18]. Der Name (norm=normal, inv=inverse) bezeichnet die

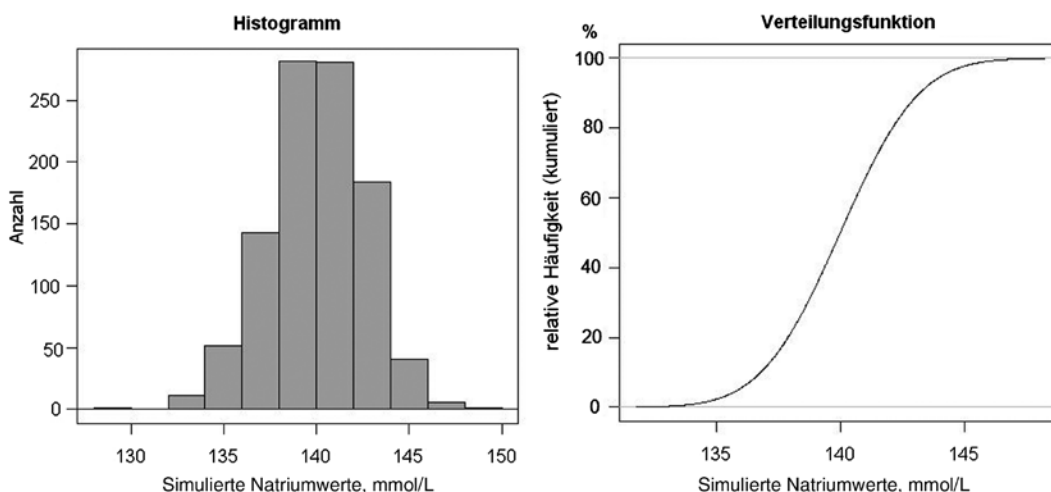


Abbildung 1: Links ist ein Histogramm aus normalverteilten Natriumwerten dargestellt (Mittelwert 140 mmol/L, Standardabweichung 2,5 mmol/L, Gesamtanzahl 1000).

Rechts findet sich die s-förmige Verteilungsfunktion $\Phi(x)$, die angibt, welcher theoretische Anteil auf alle Werte links vom angegebenen Wert x entfällt (bis zum Mittelwert 140 zum Beispiel 50%). Als relative Häufigkeit bezeichnet man die Anzahl in jeder Säule, dividiert durch die Gesamtzahl.

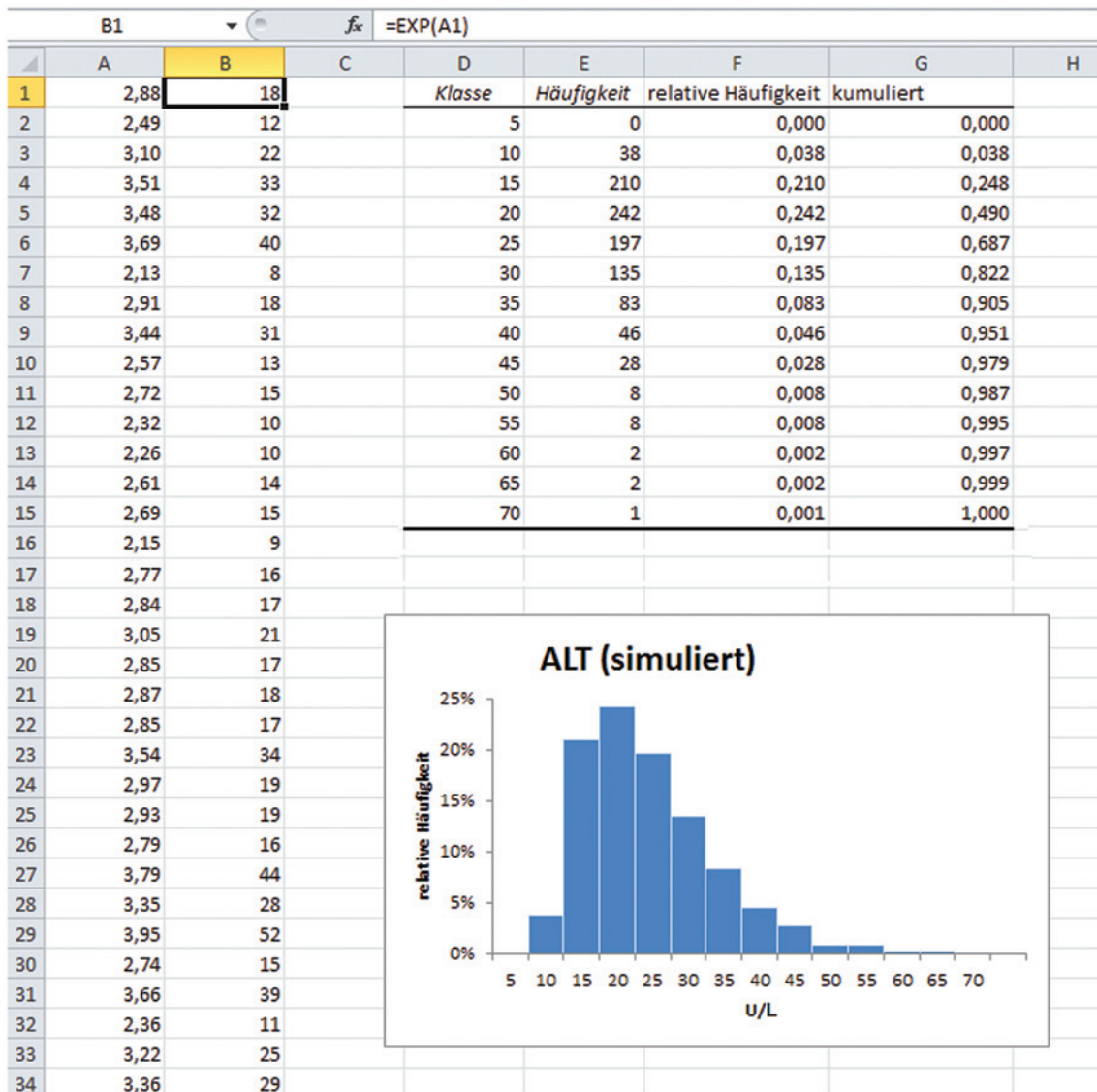


Abbildung 2: Simulation von lognormalverteilten Werten am Beispiel der ALT.

In Spalte A wurden normalverteilte Werte ($3,0 \pm 0,4$) erzeugt, die Exponenten zur Basis e repräsentieren; in Spalte B wurden diese mit der Exponentialfunktion EXP delogarithmiert. Das theoretische Maximum des Histogramms liegt folglich bei $e^{3,0} = 20$ und die 97,5-Perzentile bei $e^{3,8} = 45$.

Umkehrfunktion der s-förmigen Verteilungsfunktion¹ einer Normalverteilung (Abbildung 1) und hat die folgende Syntax:

$$x = \text{NORM.INV}(p; \mu; \sigma)$$

p = Wahrscheinlichkeit, μ = Mittelwert, σ = Standardabweichung.

Um beispielsweise einen zufälligen Natriumwert aus einer Normalverteilung mit den Parametern $\mu = 140$ und

$\sigma = 2,5$ zu simulieren, gibt man auf einem Excelblatt in Zelle A1 folgende Formel ein:

$$= \text{NORM.INV}(\text{Zufallszahl}(); 140; 2,5)$$

Die Funktion „Zufallszahl()“ in dieser Formel liefert gleichverteilte Kommazahlen zwischen 0 und 1 und deckt somit den gesamten Bereich möglicher Wahrscheinlichkeiten p ab.

In der vorliegenden Arbeit spielt die Simulation solcher Zufallszahlen mit bekannter Verteilung eine große Rolle, denn sie liefert vorhersagbare Ergebnisse, anhand derer die Qualität eines modellbasierten statistischen

¹ Eine Umkehrfunktion berechnet nicht y aus x , sondern x aus y , im vorliegenden Fall also aus einer vorgegebenen kumulativen Häufigkeit (y) das entsprechende Quantil bzw. Testresultat (x).

Verfahrens beurteilt werden kann (siehe zum Beispiel Abbildung 3, Tabellen 3 und 6). Wenn man die obige Formel aus der Zelle A1 in die darunter liegenden 999 Zellen von A2 bis A1000 des Excelblattes kopiert, erhält man insgesamt tausend normalverteilte Zufallszahlen, die realistische Natriumwerte mit einer typischen Häufung um den Mittelwert von 140 repräsentieren: Im Bereich zwischen $\mu-1\sigma$ und $\mu+1\sigma$ werden 68% und zwischen $\mu-2\sigma$ und $\mu+2\sigma$ 95% erwartet. Dieser 2σ - oder 95%-Bereich entspricht dem gesuchten Referenzintervall.

Hat man die Analyse-Funktionen installiert, so wählt man im Menüpunkt „Daten“ den Unterpunkt „Datenanalyse“. Man ruft die Funktion „Zufallszahlengenerierung“ auf und erhält eine Maske für die Eingabe der Parameter (also zum Beispiel für eine Variable mit 1000 Zufallszahlen, Verteilung „Standard“, Mittelwert 140, Standardabweichung 2,5).

Auch schiefe Verteilungen lassen sich mit der NORM.INV-Funktion simulieren. Um zum Beispiel lognormalverteilte Zufallswerte für die Alanin-Aminotransferase (ALT) bei Männern mit einem Modus bei 20 U/L ($e^{3,0}$) und einer 97,5-Perzentile bei 45 U/L ($e^{3,8}$) zu erhalten, erzeugt man in Spalte A eine Normalverteilung $3,0 \pm 0,4$ und trägt in Zelle B1 die Formel

$$=EXP(A1)$$

ein. Diese kopiert man wiederum in die 999 darunter liegenden Zellen A2 bis A1000 auf dem Excelblatt und erhält so in Spalte B lognormalverteilte Werte (Abbildung 2).

Schließlich lassen sich mit diesem Verfahren auch gemischte Populationen erzeugen, die der realen klinischen Situation nahekommen, zum Beispiel 800 unauffällige Natriumwerte ($140 \pm 2,5$) und 200 erniedrigte Werte ($130 \pm 4,5$). Das Ergebnis ist in Tabelle 2 dargestellt; die angegebenen relativen Häufigkeiten erhält man mit der Excel-Funktion „Häufigkeit“ bzw. dem Menüpunkt „Histogramm“ in den Analyse-Funktionen.

Visuelle Prüfung auf Normalverteilung

Grundlage der hier vorgestellten visuellen Verfahren sind zweidimensionale Diagramme, in denen beobachtete und theoretische Wahrscheinlichkeiten gegeneinander aufgetragen werden. Im einfachsten Fall geschieht das ohne Computer in einem Wahrscheinlichkeitsnetz (engl. Probability Plot, deutsche Abkürzung PP-Plot).

Abbildung 3 demonstriert das Vorgehen am Beispiel der simulierten Natriumwerte aus Tabelle 2: Liegen die Punkte auf einer Geraden, so kann eine Normalverteilung

Tabelle 2: Simulation von tausend Natriumwerten mithilfe der NORM.INV-Funktion.

Natrium, mmol/L	nur Nicht-Kranke		20% Kranke	
	Anzahl	kumuliert	Anzahl	kumuliert
117,5	0	0,0%	0	0,0%
120,0	0	0,0%	1	0,1%
122,5	0	0,0%	7	0,8%
125,0	0	0,0%	22	3,0%
127,5	0	0,0%	27	5,7%
130,0	0	0,0%	54	11,1%
132,5	0	0,0%	30	14,1%
135,0	19	1,9%	48	18,9%
137,5	142	16,1%	137	32,6%
140,0	335	49,6%	277	60,3%
142,5	325	82,1%	255	85,8%
145,0	158	97,9%	125	98,3%
147,5	20	99,9%	16	99,9%
150,0	1	100,0%	1	100,0%

Die linke Seite zeigt Zufallswerte einer Normalverteilung ($\mu=140$, $\sigma=2,5$), die die Resultate von Nicht-Kranken repräsentieren. Auf der rechten Seite wurden bei 20% der Werte die Parameter geändert, um Hyponatriämien zu simulieren ($\mu=130$, $\sigma=4,5$).

angenommen werden (schwarze Symbole in Abbildung 3). Das Referenzintervall ($\mu \pm 2\sigma$) liegt dann zwischen den Schnittpunkten der Geraden mit der 2,5- bzw. 97,5-Perzentile. Von dieser Geraden abweichende Werte (rote Symbole in Abbildung 3) repräsentieren „Ausreißer“ (hier die Hyponatriämien).

QQ-Plot-Verfahren

Für die Nachbildung des historischen Verfahrens mit Excel wählten wir den sog. Quantile Quantile Plot (QQ-Plot). Im Unterschied zum oben beschriebenen PP-Plot (Abbildung 3) kommt hier die Umkehrfunktion (NORM.INV) zum Einsatz: x- und y-Achse werden getauscht, sodass das Diagramm um 90° gedreht erscheint, und die (neue) x-Achse kann durch die Umrechnung der Wahrscheinlichkeiten in Quantile linear skaliert werden (Abbildung 4).

Das praktische Vorgehen in Excel ähnelt demjenigen bei den obigen Simulationen von Natrium- oder ALT-Werten. Man berechnet mithilfe der NORM.INV-Funktion theoretische Quantile einer Standardnormalverteilung ($\mu=0$, $\sigma=1$), indem man in die Formel anstelle von Zufallszahlen eine systematisch aufsteigende Reihe von Wahrscheinlichkeiten zwischen 0 und 1 eingibt (Abbildung 4). Dank der Computerunterstützung muss man sich hierbei nicht wie in Abbildung 3 auf einige wenige Punkte beschränken, sondern kann für jeden gemessenen Wert eine eigene

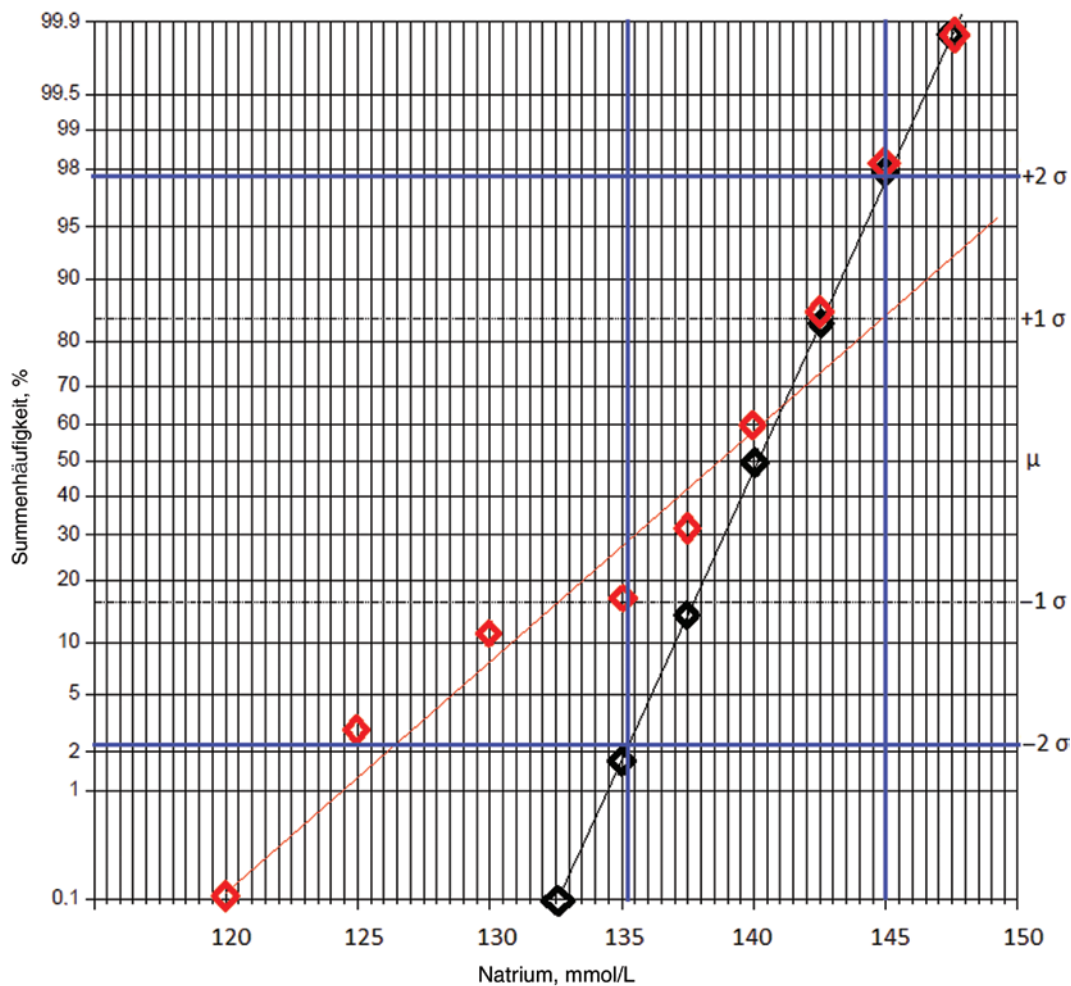


Abbildung 3: Visuelle Prüfung der Natriumwerte aus Tabelle 2 auf Normalverteilung mithilfe eines Wahrscheinlichkeitsnetzes. Schwarz: Nicht-Kranke, rot: gemischte Population mit 20% Hyponatriämien. Die blauen Linien repräsentieren den 2σ - oder 95%-Bereich, der dem Referenzintervall entspricht (hier etwa 135 bis 145 mmol/L).

Klasse bilden und deren tatsächliches mit dem theoretischen Quantil vergleichen.

Im Fall des obigen Natriumbeispiels erhält man also tausend Klassen, denen im Diagramm tausend theoretische Quantile zugeordnet werden. Die extremen Wahrscheinlichkeiten 0 und 1 müssen ausgespart bleiben, weil ihnen nicht darstellbare Quantile von minus bzw. plus unendlich entsprechen würden (vgl. auch y-Achse in Abbildung 3).²

Ergebnisse

Abbildung 5 zeigt zur Veranschaulichung die beiden im vorigen Kapitel beschriebenen Beispiele für Natrium

² Dies ist der Grund, warum in der in Abbildung 4 angegebenen Formel nicht durch n , sondern durch $n+1$ dividiert wird: Dadurch erhält man für die höchste Wahrscheinlichkeit nicht $1000/1000=1$, sondern $1000/1001=0,9990$.

(normalverteilt) und ALT (lognormalverteilt). Die x-Achse repräsentiert die theoretischen Quantile der Standardnormalverteilung ($\mu=0$, $\sigma=1$); sie zeigt also ähnlich wie die y-Achse in Abbildung 3 die Abweichungen vom Mittelwert in Vielfachen der Standardabweichung an, nur diesmal in linearer Skalierung. Im Fall der Natriumwerte erhält man im QQ-Plot die erwartete Gerade. Im Fall der ALT resultiert eine gekrümmte Kurve ohne linearen Anteil, die man jedoch in eine Gerade umformen kann, indem man die y-Achse logarithmisch skaliert (Abbildung 5 unten); alternativ kann man die ALT-Werte vor der Analyse logarithmieren und eine lineare y-Achse verwenden.

In Abbildung 6 sind typische QQ-Plots für gemischte Populationen dargestellt. Von den tausend homogen verteilten Natriumwerten ($140 \pm 2,5$ mmol/L) wurden hier 100, 200 bzw. 300 durch erniedrigte Werte ($130 \pm 4,5$) ersetzt, um 10%, 20% und 30% „Ausreißer“ zu simulieren. Wie man bereits visuell erkennt, setzen sich deren Punkte durch

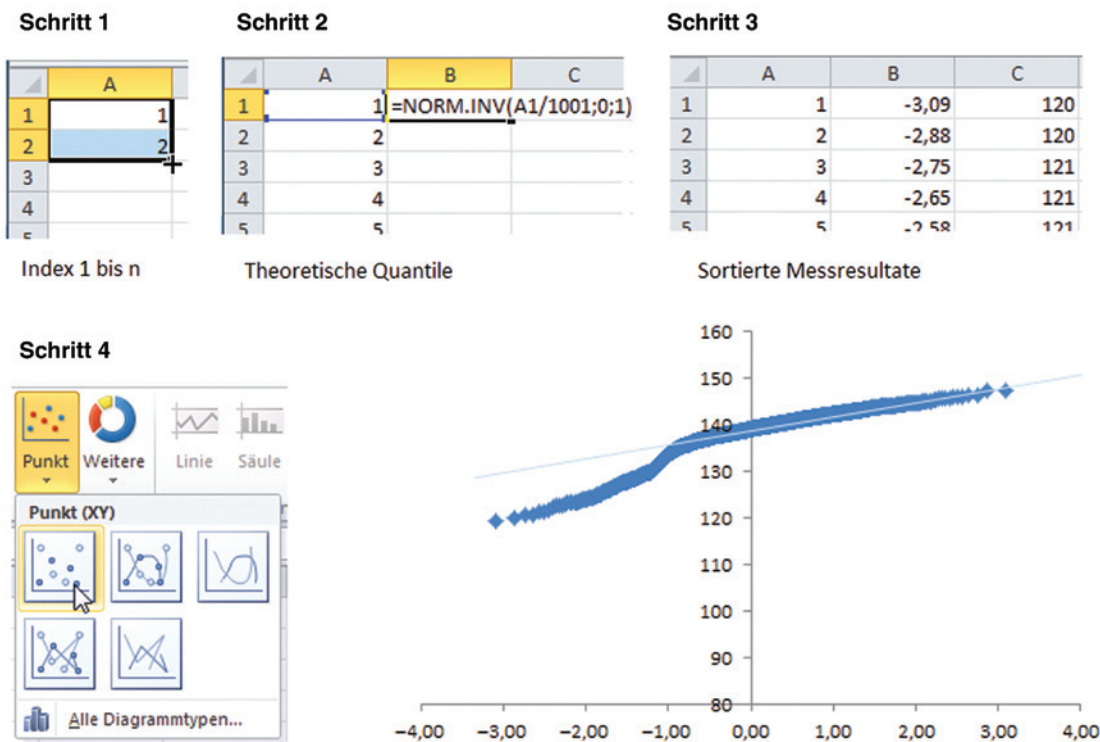


Abbildung 4: Erzeugung eines QQ-Plots mit Excel in vier Schritten.

Schritt 1: In Spalte A erzeugt man eine Zahlenreihe i von 1 bis n . **Schritt 2:** In Spalte B berechnet man daraus mit der Funktion $\text{NORM.INV}(p; \mu; \sigma)$ Quantile einer Standardnormalverteilung. Die Wahrscheinlichkeit p errechnet sich aus dem Index i dividiert durch $n+1$. Für $n=1000$ gibt man in Zelle B1 beispielsweise ein: $\text{=NORM.INV}(A1/1001; 0; 1)$. Diese Formel kopiert man in alle n Zellen der Spalte B. **Schritt 3:** In die Spalte C kopiert man die Messresultate und sortiert sie aufsteigend. **Schritt 4:** Um den gewünschten QQ-Plot zu erhalten, markiert man die Spalten B und C und fügt ein Punktdiagramm ein.

einen Knick in der Geraden ab. Mittelwert und Standardabweichung einer (Teil-) Population wurden dann auf übliche Weise aus denjenigen Werten berechnet, die im QQ-Plot auf einer Geraden lagen. Im vorliegenden Fall wurde der Knick im Bereich zwischen 136 und 137 mmol/L lokalisiert.

Tabelle 3 zeigt, dass durch die Beimengung erniedrigter Werte die Standardabweichungen in den nicht korrigierten Datensätzen immer größer (Zeile 1), die Mittelwerte dagegen immer kleiner werden (Zeile 3), sodass ein zu weites und zu niedriges Referenzintervall resultiert (bei 30% Beimengung zum Beispiel 126 bis 148 mmol/L). Die Elimination der Ausreißer auf Basis des QQ-Plots führt dagegen zu einer bemerkenswert stabilen Schätzung: Der Mittelwert bleibt nun konstant (Zeile 5), und die Standardabweichung nimmt nur geringfügig ab (Zeile 6). Die untere Referenzgrenze ändert sich aufgrund der etwas kleineren Standardabweichung um maximal 1 mmol/L, die obere wird überhaupt nicht beeinflusst.

In einer weiteren Serie von Simulationsexperimenten untersuchten wir, welchen Einfluss die subjektive Ablesung der „Knickstelle“ auf das Endergebnis hat. Wir konnten zeigen, dass die Mittelwerte durch Ableseungenauigkeiten

von bis zu 5% so gut wie überhaupt nicht beeinflusst wurden, während die Standardabweichungen eher unter- als überschätzt wurden. Abbildung 7 zeigt beispielhaft, dass die Standardabweichung – erwartungsgemäß – immer kleiner wird, je mehr Messwerte als Ausreißer definiert und abgeschnitten werden (sprich je weiter der Ablesepunkt in den linearen Teil hinein verlegt wird). Die Referenzgrenzen veränderten sich allerdings auch in diesem Beispiel nur unwesentlich um maximal 1 mmol/L.

Zur praktischen Überprüfung des Verfahrens standen Echtdaten aus einer Studie an 246 klinisch unauffälligen Studentinnen und Studenten zur Verfügung [17]. Da die Teilnehmer nicht nüchtern waren, konnte mit vorwiegend unauffälligen sowie einer begrenzten Zahl pathologischer Ergebnisse bei ernährungsabhängigen Werten gerechnet werden.

Abbildung 8 zeigt beispielhaft die Häufigkeitsverteilungen und QQ-Plots für Gesamteiweiß, Glukose und Triglyzeride. Gesamteiweiß (oben) steht für nicht gestörte Messgrößen, die sich im QQ-Plot als annähernd normalverteilt ohne nennenswerte Abweichungen erwiesen. Ähnlich wie Gesamteiweiß verhielten sich Albumin, Natrium,

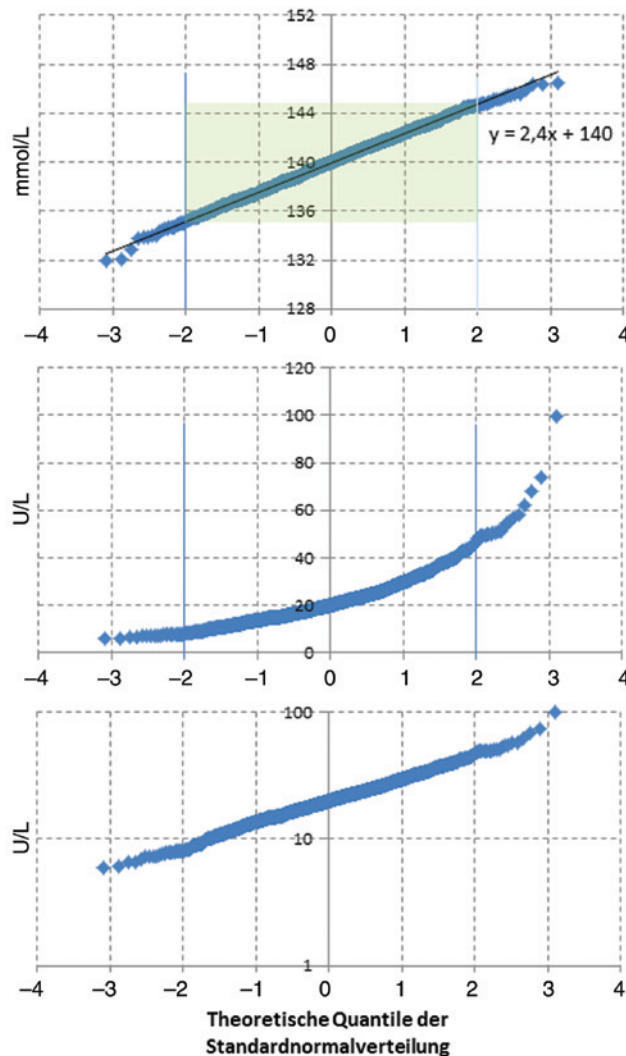


Abbildung 5: QQ-Plots für je tausend simulierte Messwerte. Normalverteilte Natriumwerte bilden eine Gerade (oben), aus deren Regressionsgleichung man den Mittelwert μ und die Standardabweichung σ abschätzen kann. Das Referenzintervall (135 mmol/L bis 145 mmol/L, $\mu \pm 2\sigma$) lässt sich zwischen den Quantilen -2 und $+2$ der Standardnormalverteilung ablesen (grüne Fläche). Lognormalverteilte ALT-Werte führen zu einer gekrümmten Kurve (Mitte), die sich nach logarithmischer Skalierung der y-Achse linearisiert (unten).

Kalium, Calcium und Kreatinin. Für die Glukosewerte (mittleres Diagramm) ergab sich im zentralen Teil ebenfalls eine Gerade, doch fanden sich einige Hypoglykämien (niedrigster Wert 2,2 mmol/L bzw. 40 mg/dL) und Hyperglykämien (höchster Wert 17,5 mmol/L bzw. 315 mg/dL). Auffällig erhöhte Werte fanden sich auch bei Cholesterin, GGT, ALT und AST. Bei den drei Enzymen wurde vor der QQ-Plot-Analyse eine Logarithmierung der Originalwerte durchgeführt, um den gekrümmten Plot in eine Gerade zu überführen (vgl. Abbildung 5 unten).

Die Triglyzeridwerte (unteres Diagramm) wiesen eine komplexe, aus Subpopulationen zusammengesetzte Verteilung auf, die die Grenzen des hier vorgestellten Verfahrens demonstriert. Im QQ-Plot zeigten sich zwei etwa gleich lange gerade Anteile, von denen der eine einen hypoglykämischen Wertebereich zwischen 0,34 und 0,80 mmol/L (30 bis 70 mg/dL), der andere den Bereich von 0,80 bis 2,0 mmol/L (70 bis 175 mg/dL) umfasste. Zusätzlich fanden sich Hypertriglyzeridämien bis zu 3,74 mmol/L (327 mg/dL). Dies ist nicht überraschend, da die Versuchspersonen nicht nüchtern waren. Die Trennung in zwei lineare Anteile mit unterschiedlichen Steigungen blieb auch nach Logarithmieren der Werte erhalten (nicht dargestellt). Wie in der Diskussion (Abbildung 10) näher ausgeführt wird, sind solche Daten mit dem vorliegenden Verfahren nicht auswertbar.

Von diesem Beispiel abgesehen waren aber alle QQ-Plots über einen großen Abschnitt linear (gegebenenfalls nach vorheriger Logarithmierung der Werte). Etwaige Ausreißer setzten sich durch einen Knick klar ab und wurden von der weiteren Berechnung ausgeschlossen. Die so erhaltenen Referenzintervalle stimmten trotz der relativ kleinen Fallzahl von weniger als 250 mit den an der MHH etablierten Angaben gut überein (Beispiele in Tabelle 4).

Für einen Methodenvergleich standen uns zwei umfangreiche Datensätze aus der klinischen Routine zur Verfügung: 17.506 Kreatininwerte zwischen 0,26 und 22,61 mg/dL sowie 56.137 AST-Werte zwischen 1 und 200 U/L. Die Auswertung erfolgte jeweils getrennt nach Männern und Frauen. Abbildung 9 stellt beispielhaft die mit den beiden Verfahren erhaltenen Grafiken für Kreatinin bei Frauen ($n=9393$) dar. Man erkennt im Histogramm der Vergleichsmethode (links) die symmetrisch verteilte Population der Gesunden, die im QQ-Plot-Verfahren (rechts) durch eine Gerade repräsentiert werden. Bei der Analyse der AST-Daten überprüften wir zusätzlich die Annahme einer Lognormal-Verteilung der Werte [19]. Zu diesem Zweck wurden die Originaldaten vor der Analyse logarithmiert und die daraus berechneten Referenzgrenzen anschließend wieder delogarithmiert.

Tabelle 5 zeigt die gute Übereinstimmung der beiden Verfahren. Die Differenz betrug beim Kreatinin maximal 0,1 mg/dL, bei der AST 3 U/L. Die logarithmische Transformation führte im Fall der AST zu einer leichten Verbreiterung und Rechtsverschiebung des Referenzintervalls.

Diskussion

Das hier vorgestellte Verfahren ist nicht grundsätzlich neu, sondern repräsentiert nur das (vorläufig) letzte Glied in einer

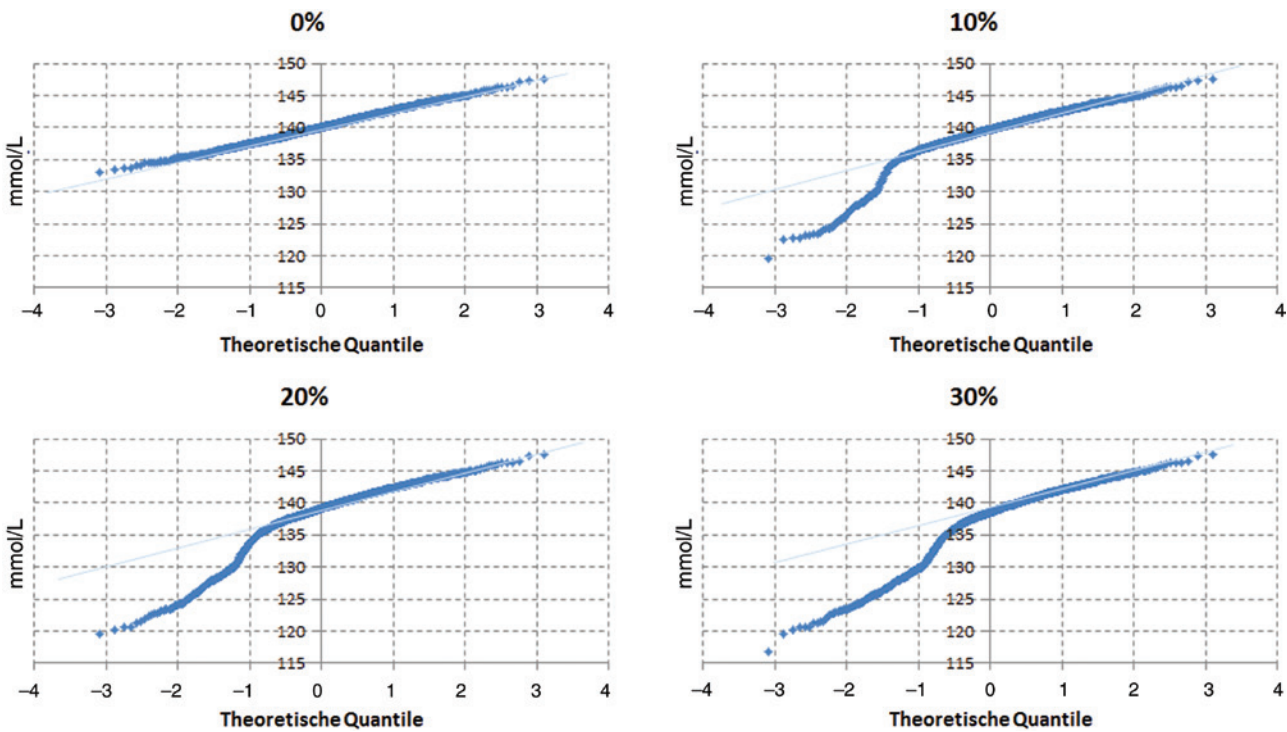


Abbildung 6: Simulation von je tausend Natriumwerten mit unterschiedlich großem prozentualen Anteil an Hyponatriämien (farbig markierter Bereich). Man erkennt jeweils ein Geradenstück, von dem sich die erniedrigten Werte durch einen Knick deutlich absetzen.

Tabelle 3: Schätzung von Referenzintervallen für Natrium (Zielwert 135 bis 145 mmol/L) aus gemischten Populationen mit 0% bis 30% erniedrigten Werten.

	0%	10%	20%	30%
(A)				
Mittelwert	140	139	138	137
Standardabweichung	2,6	4,1	5,1	5,7
Untergrenze	135	131	128	126
Obergrenze	145	147	148	148
(B)				
Mittelwert	140	140	140	140
Standardabweichung	2,6	2,4	2,3	2,3
Untergrenze	135	135	136	136
Obergrenze	145	145	145	145

(A) Berechnung aus allen tausend Werten. (B) Nach visueller Elimination der Werte, die im QQ-Plot nicht auf der Geraden liegen.

Kette von Vorläuferversionen, die alle auf den amerikanischen Biostatistiker Robert G. Hoffmann [14] zurückgehen. Dieser propagierte 1963 erstmals die „indirekte Schätzung“ von Referenzintervallen aus Messungen an stationären und ambulanten Patienten anstelle von Referenzpersonen. Zwei Voraussetzungen mussten dafür erfüllt sein: Die Messwerte sollten in etwa einer Gauß’schen Normalverteilung

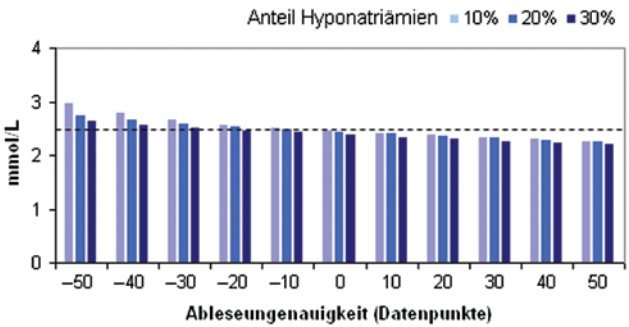


Abbildung 7: Identisches Simulationsexperiment wie in Abbildung 6. Die Lokalisation des Knicks wurde jedoch im Rahmen einer Fehleranalyse um 10 bis 50 Messpunkte nach links (negative Werte) bzw. rechts (positive Werte) verschoben. Die dadurch bedingte Fehlschätzung der Standardabweichungen vom Sollwert (gestrichelte Linie) betrug maximal 15%.

folgen, und die Mehrzahl der Patienten sollte in Bezug auf den untersuchten Analyten klinisch unauffällig sein. Als „proof of concept“ demonstrierte er sein Vorgehen anhand von Glukosemessungen bei Nichtdiabetikern, indem er die Messwerte aufsteigend sortierte und die kumulative Häufigkeit jedes Werts auf Wahrscheinlichkeitspapier auftrug (vgl. Abbildung 3). Das Referenzintervall ermittelte

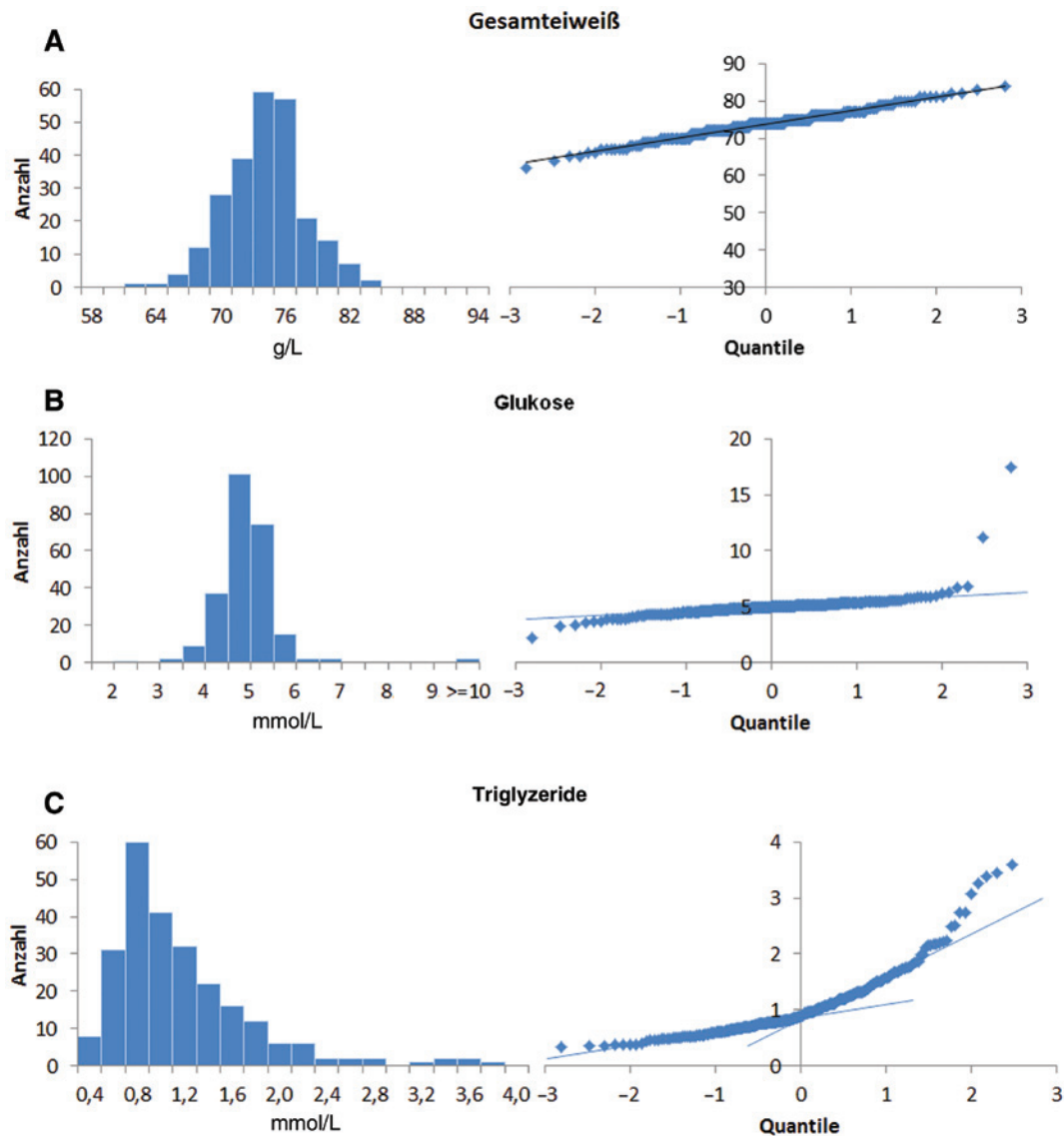


Abbildung 8: Ausgewählte Anwendungsbeispiele des QQ-Plot-Verfahrens.

(A) Linearer QQ-Plot ohne auffällige Werte. (B) Linearer QQ-Plot mit wenigen pathologischen Werten auf beiden Seiten. (C) Komplexer QQ-Plot mit zwei etwa gleich langen geraden Abschnitten sowie pathologischen Werten im oberen Bereich.

Tabelle 4: Schätzung von Referenzintervallen mit dem QQ-Plot-Verfahren bei einer Population von klinisch unauffälligen Studenten, denen Blut unter nicht standardisierten Bedingungen abgenommen wurde [17].

	MW	SD	Geschätztes Referenzintervall	Angabe der MHH
Gesamteiweiß, g/L	73,8	3,7	66–81	65–80
Albumin, g/L	43,8	4,0	36–52	35–52
Natrium, mmol/L	140	1,8	136–143	135–145
Kalium, mmol/L	4,29	0,29	3,7–4,9	3,6–5,4
Calcium, mmol/L	2,38	0,09	2,2–2,6	2,15–2,6
Glukose, mmol/L	4,95	0,44	4,1–5,8	3,9–5,5

er visuell, indem er durch den zentralen Part der Punkte eine Gerade legte und deren Schnittpunkte mit der 2,5- und 97,5-Perzentile bestimmte. Dieser Ansatz fand teils in der ursprünglichen, teils in modifizierter Form Eingang in die Medizin [20–22].

Das in der vorliegenden Arbeit beschriebene Verfahren kombiniert diverse Modifikationen der Originalmethode, um mit möglichst geringem Aufwand möglichst robuste Schätzungen zu erzielen. Grundlage ist ein 2014 publizierter „modified Hoffmann approach“ [4], der die Quantile eines QQ-Plots ebenso mit Excel berechnet, wie in Abbildung 4 beschrieben. Die Autoren schließen daran

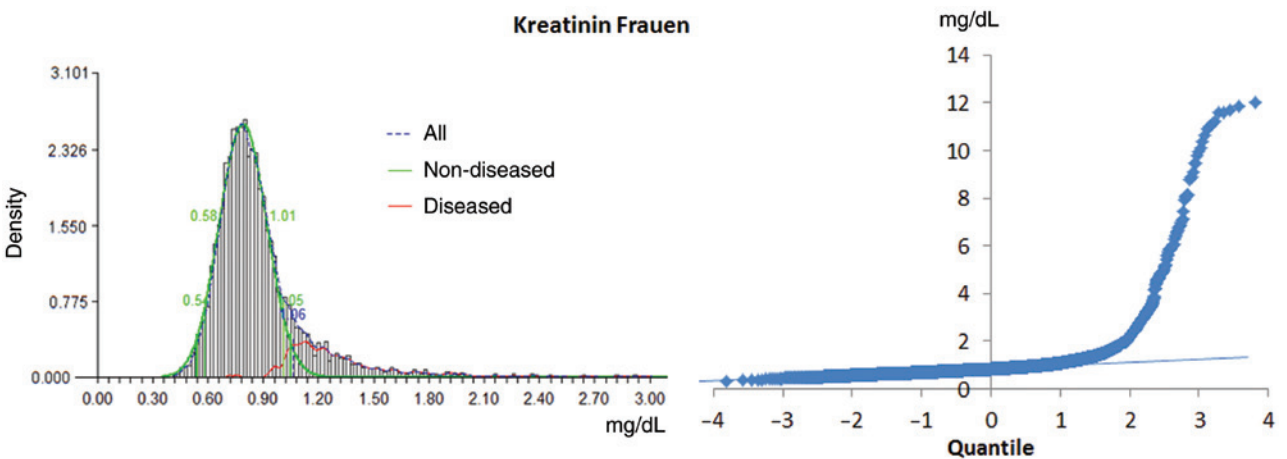


Abbildung 9: Methodenvergleich der Referenzintervallschätzung aus klinischen Routinedaten am Beispiel für Kreatinin bei 9393 Frauen. Links ist das RLE-Verfahren der AG Richtwerte dargestellt (Berechnung mit R, Originaldarstellung in Englisch mit Dezimalpunkten und konventionellen Einheiten); die grüne Linie gibt die Dichte der angenommenen Normalverteilung wieder, die rote Linie die Dichte der nicht zum Modell passenden, erhöhten Werte (die x-Achse ist rechts abgeschnitten). Rechts ist der entsprechende QQ-Plot zu sehen (Berechnung mit Excel); die Daten der nicht-pathologischen Hauptpopulation stellen sich als Gerade dar, die Ausreißer sind durch einen Knick klar abgrenzbar.

Tabelle 5: Methodenvergleich der Referenzintervallschätzung aus klinischen Routinedaten.

Messgröße und Verteilungsmodell	Männer		Frauen	
	Verfahren 1	Verfahren 2	Verfahren 1	Verfahren 2
Kreatinin, mg/dL (N)	0,6–1,4	0,7–1,3	0,5–1,2	0,5–1,1
AST, U/L (N)	12–33	9–33	12–29	10–29
AST, U/L (LN)	13–36	13–38	13–32	13–33

Trotz eines hohen Anteils an Ausreißern liefern beide Verfahren recht ähnliche Schätzwerte. Erläuterungen zu den Modellannahmen finden sich im Text. Verfahren 1=QQ-Plot-Verfahren, Verfahren 2=Vergleichsmethode der AG Richtwerte. N, Normalverteilungs-Modell; LN, Lognormalverteilungs-Modell.

jedoch eine segmentale Regressionsanalyse mithilfe des Statistikpakets R an, die aufgrund ihrer Komplexität und Fehleranfälligkeit hier nicht übernommen wurde. Stattdessen ermitteln wir den linearen Teil des QQ-Plots visuell und schätzen das Referenzintervall $\mu \pm 2\sigma$ [23] nach Elimination potenzieller Ausreißer direkt über die Berechnung von Mittelwert und Standardabweichung.

Dieses Vorgehen ist nicht nur deutlich einfacher, sondern führt offenbar auch zu robusteren Ergebnissen. Unsere Experimente mit simulierten Daten zeigen, dass sich die Werte kranker Personen und subjektive Ableseungenauigkeiten auf die Schätzung der Referenzintervalle nur wenig auswirken. Mit dem aufwendigeren regressionsbasierten Verfahren erhielten Shaw et al. [4] dagegen zum Teil extrem fehlerhafte Ergebnisse, die die Autoren auf einen zu hohen Prozentsatz kranker Personen zurückführten.

Eine Erklärung für die Fehleranfälligkeit des regressionsbasierten Ansatzes bei gemischten Populationen

liefert die Abbildung 10: Hier wurden drei unterschiedlich große Populationen (300 nicht Kranke, 200 mäßig Kranke und 100 schwer Kranke) mit jeweils normalverteilten arbiträren Messwerten simuliert (100 ± 20 , 150 ± 40 und 250 ± 50). In Summe resultiert daraus eine typisch links-gipflige Verteilung, die an eine Lognormalverteilung erinnert. Im QQ-Plot erkennt man zwar drei in etwa lineare Abschnitte, aber deren Schnittpunkte mit der y-Achse entsprechen nicht den erwarteten Mittelwerten: Statt 100, 150 und 250 liest man 120, 100 und 180 ab. Auch die Steigungen der Teilgeraden in Abbildung 10 entsprechen nicht den erwarteten Standardabweichungen.

Für eine korrekte Berechnung müsste man die Daten der drei Geradenstücke jeweils isoliert betrachten und gegen die Quantile dazu passender Standardnormalverteilungen mit $n=300$, 200 bzw. 100 auftragen. Bei summarischer Betrachtung aller 600 Datenpunkte verschieben sich die jeweiligen Quantile um die Daten aus den anderen

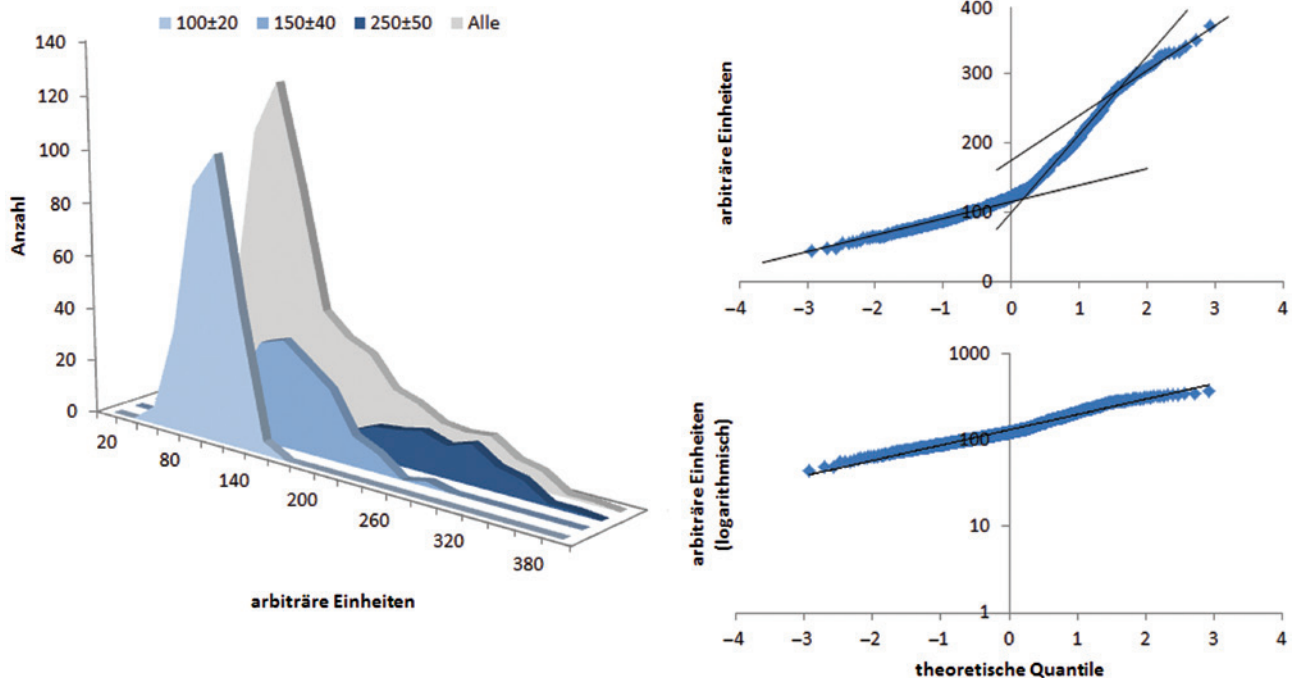


Abbildung 10: Histogramme (links) und QQ-Plots für ein gemischtes Patientenkollektiv.

Mittels Simulation wurden drei normalverteilte Wertemengen erzeugt (blaue Flächen), aus denen in Summe eine linksgeipflige Verteilung (graue Fläche) resultiert. Im QQ-Plot (rechts oben) sind die drei Subpopulationen optisch gut unterscheidbar, aber die Lage der geraden Abschnitte (Steigung und Intercept) erlaubt keine Rückschlüsse auf die zugrundeliegenden Mittelwerte und Standardabweichungen. Auch bei logarithmischer Transformation der y-Achse (rechts unten) sind die drei linearen Teilstrecken erkennbar, werden aber durch die automatische Skalierung von Excel stark abgeflacht.

Tabelle 6: Schätzung von Referenzintervallen einer gemischten Population (Abbildung 10) mit dem hier vorgestellten Verfahren sowie durch segmentale Regressionsrechnung nach Shaw et al. [4].

	Soll (Simulation)	Aktuelles Verfahren	Shaw et al.
1 (n=300)	60–140	62–125	66–177
2 (n=200)	70–230	83–257	–124 to 324
3 (n=100)	150–350	262–350	45–305

Der Grad der Rotfärbung symbolisiert das Ausmaß der jeweiligen Fehlschätzung; im Extremfall erhält man mit dem Regressionsverfahren einen negativen unteren Grenzwert.

Geraden. Dadurch werden die Mittelwerte teils über- und teils unterschätzt und die Standardabweichungen durchwegs überschätzt. Im Extremfall resultieren aus diesen Fehlern sogar negative untere Grenzwerte (Tabelle 6, rechte Spalte).

Aus dem bisher Gesagten lassen sich folgende fünf Empfehlungen ableiten:

- Das hier beschriebene QQ-Plotverfahren eignet sich vor allem für die rasche visuelle Beurteilung, ob sich an Patienten erhobene Messwerte grundsätzlich für die Berechnung von Referenzintervallen eignen.

- Erhält man in der Grafik eine durchgängige Gerade (Abbildung 5 oben), so kann man von einer Normalverteilung ausgehen und das Referenzintervall aus $\mu \pm 2\sigma$ abschätzen.
- Etwaige Ausreißer sind in der Regel durch Abweichungen von der Geradenform gut erkennbar (Abbildung 8 Mitte, Abbildung 9 rechts) und können bis zu einem Anteil von 20 bis 30% (Tabelle 3) vor der Berechnung von μ und σ eliminiert werden.
- Weicht die gesamte Kurve von der Geradenform ab (Abbildung 5 Mitte), so kann man die Messwerte logarithmieren und den QQ-Plot erneut erstellen. Erhält man auf diese Weise eine Gerade, so wiederholt man die Punkte 2 und 3 mit logarithmierten Werten und delogarithmiert die Ergebnisse.
- Wenn auch durch Logarithmieren keine Gerade erzeugt werden kann (Abbildung 10 rechts unten), so muss man mit erheblichen Fehlschätzungen rechnen (Tabelle 6). In diesem Fall sollte ein geeigneteres Patientenkollektiv oder besser noch ein Kollektiv aus nicht-kranken Referenzpersonen zusammengestellt werden.

Neben der einfachen Logarithmierung wurden zahlreiche weitere Transformationsverfahren empfohlen, um

schiefe Verteilungen einer Normalverteilung anzunähern [23]. So kommt zum Beispiel beim RLE-Verfahren der Arbeitsgruppe Richtwerte standardmäßig die Box-Cox-Transformation aus der Familie der Powertransformationen zum Einsatz [7]. Einige Autoren schlagen vor, die Originaldaten bei unbekannter Verteilungsfunktion vor der Anwendung statistischer Modelle grundsätzlich zu transformieren [19, 24]. Dabei ist allerdings zu bedenken, dass auffällige Datenstrukturen durch Transformationen verschleiert werden können (vergleiche Abbildung 10), und dass man bei Vorliegen schiefer Verteilungen stets nach physiologischen oder klinischen Gründen fahnden sollte, ehe man sie auf rein statistischer Basis „normalisiert“.

In einer 2015 erschienenen Arbeit empfehlen Tate et al. [25] den Einsatz indirekter Verfahren vor allem zur Anpassung publizierter Referenzintervalle an lokale Gegebenheiten (Analytik, Präanalytik, spezielle Patientenkollektive usw.). Die Autoren fordern, dass die Mehrzahl der Messwerte aus einer klinisch unauffälligen Population stammt und dass die auffälligen Werte statistisch klar abgrenzbar sind. Dieser Forderung trägt die vorliegende Arbeit in besonderer Weise Rechnung: Wie die Abbildung 9 zeigt, lassen sich Ausreißer mit bloßem Auge anhand des „Knicks“ im QQ-Plot leicht identifizieren, während beispielsweise Abweichungen von der Glockenkurvenform der Normalverteilung weniger deutlich sind.

Das hier beschriebene Verfahren steht nicht in Konkurrenz zu anderen Methoden, sondern ist eine nützliche Ergänzung. Im Vergleich zur IFCC-Referenzmethode [1] kommt es ohne Rekrutierung gesunder Probanden aus, im Vergleich zum RLE-Verfahren der Arbeitsgruppe Richtwerte [7] benötigt es keine Installation des Statistikprogramms R und liefert auch mit rund 200 Messwerten bereits plausible Resultate. Als nachteilig kann man die visuelle Beurteilung des QQ-Plots ansehen, weil sie subjektiv und nicht automatisierbar ist. An einer Automatisierung wird deshalb derzeit gearbeitet.

Dennoch bietet die Visualisierung auch Vorteile gegenüber rein rechnerischen Verfahren, denn sie liefert auf einen Blick viele Informationen zur Datenqualität und Verteilungsform, zu etwaigen Ausreißern und Subpopulationen usw. So werden Messwerte, die die Schätzung von Referenzintervallen beeinträchtigen könnten, relativ sicher erkannt. Nicht einmal bei ausgewählten Referenzindividuen kann man letztlich ausschließen, dass sie an klinisch „stummen“ Erkrankungen (Fettleber, Atherosklerose) leiden oder beeinflussende Substanzen (Alkohol, Medikamente) einnehmen. Dies gilt umso mehr für routinemäßig an klinischen Kollektiven erhobene Labordaten. Deshalb sollte das QQ-Plot-Verfahren unabhängig von

der letztendlich verwendeten direkten oder indirekten Methode grundsätzlich als eine Art Vorfilter eingesetzt werden.

Danksagung: Die Autoren bedanken sich bei der Deutschen Vereinten Gesellschaft für Klinische Chemie und Laboratoriumsmedizin (DGKL e. V.) für die organisatorische Unterstützung der Arbeit im Rahmen der Arbeitsgruppen Bioinformatik und Richtwerte. Ein besonderer Dank gilt Dr. Farhad Arzideh, Universität Bremen, Dr. Norman Bitterlich, Medizin & Service Chemnitz, und Prof. Dr. Frank Klawonn, Helmholtz-Zentrum Braunschweig, für wertvolle Diskussionsbeiträge.

Autorenbeteiligung: Alle Autoren tragen Verantwortung für den gesamten Inhalt dieses Artikels und haben der Einreichung des Manuskripts zugestimmt.

Forschungsförderung: Keine.

Interessenkonflikt: Die Autoren erklären, dass keine wirtschaftlichen oder persönlichen Interessenkonflikte bestehen.

Literatur

1. Clinical and Laboratory Institute Standards document C28-A3: Defining, establishing and verifying reference intervals in the clinical laboratory: approved guideline, 3rd ed., 2008.
2. DAkkS Checkliste zur DIN EN ISO 15189:2014 für medizinische Laboratorien, Abschnitt 5.5.2., verfügbar unter www.dakks.de, zuletzt besucht am 05.10.2015.
3. Haeckel R, Wosniok W, Arzideh F. A plea for intra-laboratory reference limits. Part 1. General considerations and concepts for determination. Clin Chem Lab Med 2007;45:1033–42.
4. Shaw J, Cohen A, Konforte D, Bineh-Marvasti T, Colantonio D, Adeli K. Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clin Biochem 2014;47:166–72.
5. Sonntag O. Ist das normal? – Das ist normal! Über die Bedeutung und Interpretation des sogenannten Normalwertes. J Lab Med 2003;27:302–10.
6. Glick J. Statistics of patient test values: application of indirect normal range and quality control. Clin Chem 1972;18:1504–13.
7. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, et al. A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-laboratory databases demonstrated by catalytic activity concentrations of enzymes. Clin Chem Lab Med 2007;45:1043–57.
8. Arzideh F, Brandhorst G, Gurr E, Hinsch W, Hoff T, Roggenbuck L, et al. An improved indirect approach for determining reference limits from intra-laboratory databases exemplified by concentrations of electrolytes. J Lab Med 2009;33:52–66.
9. Zierk J, Arzideh F, Haeckel R, Rascher W, Rauh M, Metzler M. Indirect determination of pediatric blood count reference intervals. Clin Chem Lab Med 2013;51:863–72.

10. Arzideh F, Wosniok W, Haeckel R. Reference limits of plasma and serum creatinine concentrations from intra-laboratory data bases of several German and Italian medical centres. *Clin Chim Acta* 2010;411:215–21.
11. Arzideh F, Wosniok W, Haeckel R. Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory data bases from several German and Italian medical centres. *Clin Chem Lab Med* 2011;49:659–64.
12. Hoffmann G. IT-Werkzeuge zur Auswertung großer labordiagnostischer Datensätze. *Klin Chem Mitteilungen* 2011;42:124–30.
13. Pryce J. Level of haemoglobin in whole blood and red blood cells, and proposed convention for defining normality. *Lancet* 1960;2:333–6.
14. Hoffmann R. Statistics in the practice of medicine. *J Am Med Assoc* 1963;185:864–73.
15. Cook M, Levell M, Payne R. A method for deriving normal ranges from laboratory specimens applied to uric acid in males. *J Clin Path* 1970;23:778–80.
16. Hoffmann G. Auflösung eines Dilemmas: Referenzintervalle zum Selbermachen. *Trillium Diagnostik* 2014;12:159–61.
17. Lichtinghagen R, Senkpiel-Jörns D, Brand K, Janzen N. Beurteilung des Einflusses verlängerter Stauzeiten auf nicht-normalisierte versus normalisierte klinisch-chemische Messgrößen. *J Lab Med* 2013;37:131–7.
18. <http://office.microsoft.com/de-ch/excel-help/norm-inv-funktion-HP010335690.aspx>, abgerufen am 26.5.2015.
19. Haeckel R, Wosniok W. Observed, unknown distributions of clinical chemical quantities should be considered to be log-normal: a proposal. *Clin Chem Lab Med* 2010;48:1393–6.
20. Neumann G. Determination of normal ranges from routine laboratory data. *Clin Chem* 1968;14:979–88.
21. Reed A, Cannon D, Winkelmann J, Bhasin Y, Henry R, Pileggi V. Estimation of normal ranges from a controlled sample survey. *Clin Chem* 1972;18:57–66.
22. Soldin O, Hoffmann E, Waring M, Soldin S. Serum iron, ferritin, transferrin, total iron binding capacity, hs-CRP, LDL, cholesterol and magnesium in children; new reference intervals using the Dade Dimension Clinical Chemistry System. *Clin Chim Acta* 2004;342:211–7.
23. Reed A, Wu G. Evaluation of a transformation method for estimation of normal range. *Clin Chem* 1974;20:576–81.
24. Harris E, DeMets D. Estimation of normal ranges and cumulative proportions by transforming observed distributions to Gaussian form. *Clin Chem* 1972;18:605–12.
25. Tate J, Yen T, Jones G. Transference and validation of reference intervals. *Clin Chem* 2015;61:1012–5.