Originalarbeiten

Zum Stichprobenfehler im Human-Biomonitoring

Dieter Helm

Umweltbundesamt, Fachgebiet II 1.4, Postfach 33 00 22, D-14191 Berlin

Korrespondenzadresse: Dr. Dieter Helm (dieter.helm@uba.de)

DOI: http://dx.doi.org/10.1065/uwsf2003.12.071

Zusammenfassung

Ziel und Absicht. Human-Biomonitoring ermittelt die Konzentration von Xenobiotika in Populationen mittels Stichproben, wobei notwendigerweise Stichprobenfehler auftreten. Deren Höhe sollte bestimmt werden.

Methoden. Für eine fiktive Population von 200.000 Personen wurden mehrere, unterschiedlich breite Konzentrationsverteilungen simuliert. Davon wurden verschieden große Stichproben gezogen und der Stichprobenfehler bestimmt. Dieser ist definiert als die prozentuale Differenz der geometrischen Mittelwerte von Stichprobe und Population.

Ergebnisse und Schlussfolgerungen. Der Stichprobenfehler ist von der Stichprobengröße und der Breite der Konzentrationsverteilung abhängig; seine Schätzung ist für beliebige Xenobiotika möglich, sofern diese logarithmisch normal verteilt sind und der Stichprobenumfang zwischen 10 und 50.000 liegt. Für die Schätzung wurde eine Formel abgeleitet.

Ausblick. Der Stichprobenfehler sollte, gemeinsam mit dem Fehler der Messung, bei der Darstellung und Interpretation von Ergebnissen berücksichtigt werden.

Schlagwörter: Geometrisches Mittel; Human-Biomonitoring; logarithmische Normalverteilung; Monte-Carlo-Simulation; Stichprobenfehler; Stichprobentheorie; Umweltprobenbank; Umwelt-Survey

Abstract

About the Sampling Error in Human Biomonitoring

Goal and Scope. Human biomonitoring determines the concentration of xenobiotics in populations by means of smaller samples, thus necessarily arising sampling errors. These are determined.

Methods. For a fictitious population of 200,000 persons, differently broad xenobiotic concentration distributions were simulated. Samples of varying size were randomly drawn and the sampling error, defined as the proportional difference between the geometric means of sample and population, was determined.

Results and Conclusions. The sampling error depends on the sample size and the width of the concentration distribution; its estimation is possible for any xenobiotic, given it has lognormal distribution, and the sample size is between 10 and 50,000. For its estimation an equation was derived.

Outlook. When presenting and interpreting results of human biomonitoring, the sampling error must be considered, together with the uncertainty of the measurement.

Keywords: Environmental specimen bank; geometric mean; German environmental survey; human biomonitoring; lognormal distribution; Monte Carlo simulation; sampling error; sampling theory

Einleitung

Human-Biomonitoring ermittelt die Konzentrationen von Xenobiotika in Humanproben. Das Umweltbundesamt betreibt die beiden auf unterschiedliche Teile der Bevölkerung Deutschlands bezogenen Studien Umwelt-Survey und Umweltprobenbank [1–3]. Weitere Beispiele für national ausgerichtete Studien sind NHANES [4], PheeCad [5] und Cadmibel [6]. Eine multinational ausgerichtete Studie ist EXPOLIS [7]. Allen diesen Studien ist gemeinsam, dass ausgehend von den Messwerten einer Stichprobe auf die Belastung einer größeren Population und gegebenfalls auf deren gesundheitliche Implikationen geschlossen werden soll.

Ziel der vorliegenden Arbeit ist die nähere Betrachtung des Stichprobenfehlers im Human-Biomonitoring. Stichprobenfehler ist hier definiert als die prozentuale Abweichung einer statistischen Kenngröße der Stichprobe von der 'wahren', aber unbekannten Kenngröße der Population. Nach dem Gesetz der großen Zahlen nimmt der Unterschied zwischen Population und Zufallsstichprobe mit wachsenden Stichprobenumfang ab. Das gilt auch für Stichproben, die nach bestimmten Auswahlprinzipien als repräsentativ für die Grundgesamtheit gezogen werden, sofern die Ziehungsmerkmale unabhängig sind von der Belastung mit Xenobiotika [8].

Konzentrationen von Xenobiotika in Umwelt- und Human-Proben folgen in der Regel einer so genannten logarithmischen Normalverteilung. Dies ist empirisch belegt [1,9], sowie durch theoretische Überlegungen [10,11] und experimentelle Ansätze [12] untermauert. Logarithmische Normalverteilungen, auch Lognormalverteilungen genannt, sind eingipflige Verteilungen, die sich von der bekannten Normalverteilung (Gauss'sche 'Glockenform') vor allem durch ihre ausgeprägte Asymmetrie unterscheiden: sie steigen links steil an und fallen lang ausgezogen nach rechts ab [8]. Zu Stande kommt die logarithmische Verteilung dadurch, dass Zufallsgrößen nicht additiv sondern multiplikativ zusammenwirken [10].

Als statistische Kenngröße zur Beschreibung einer Lognormalverteilung ist das arithmetische Mittel ungeeignet, da es den Mittelwert überschätzt. In der Praxis wird meist das geomet-

© ecomed verlagsgesellschaft AG & Co. KG, D-86899 Landsberg und Ft. Worth/TX • Tokyo • Mumbai • Seoul • Melbourne • Paris

rische Mittel verwendet [1–6,13,14]. Als Maß für die Breite und Schiefe einer logarithmischen Normalverteilung kann die Standardabweichung s der logarithmierten Daten $ln(x_i)$ herangezogen werden. Weitere übliche Angaben sind das Konfidenzintervall des geometrischen Mittels und Perzentile[1–3,9].

1 Methoden

Ergebnisse der Umweltprobenbank [3], des Umwelt-Surveys [9,13] und anderer Studien [4–6,14] wurden im Hinblick auf die Breite von Konzentrationsverteilungen ausgewertet. Wenn Angaben zur Breite fehlten, wurde s aus dem Stichprobenumfang, dem geometrischen Mittelwert und dessen Konfidenzintervall berechnet.

Um den Stichprobenfehler zu bestimmen, wurden logarithmische Normalverteilungen unterschiedlicher Breiten erzeugt. Diese Populationsverteilungen umfassten jeweils 200.000 fiktive Personen; der geometrische Mittelwert war für alle Populationen identisch, die Breite der Verteilungen variierte zwischen 0,25 und 1,50. In einer Monte-Carlo-Simulation wurden von jeder Population mittels Zufallszahlen Stichproben gezogen; die Stichprobengrößen lagen zwischen 10 und 150.000. Jede Stichprobenziehung wurde 1.000-fach wiederholt. Für jede Stichprobenziehung wurde der geometrische Mittelwert der Stichprobe berechnet und der Betrag seiner Differenz zum bekannten geometrischen Mittelwert der Population bestimmt. Die Verteilung dieses Betrags ergab die Fehlerverteilung [8]. Die Kenngrößen der Fehlerverteilung wurden bestimmt. Zur Ableitung des Stichprobenfehlers wurde das 95. Perzentil ausgewählt. Es besagt, dass mit einer Wahrscheinlichkeit von 95 Prozent der Stichprobenfehler bei einer Stichprobe der genannten Größe n und der Breite s unter dem angegeben Wert liegen wird: $E_{0,95;n;s}$. Alle Berechnungen wurden mit SPSS 9.0 durchgeführt.

2 Ergebnisse

2.1 Die Breite empirischer Verteilungen

Die Breite von Konzentrations-Verteilungen von Xenobiotika in Humanproben liegt in der Größenordnung von etwa 1. Sie ist abhängig von der betrachteten Substanz und von der Zusammensetzung des Kollektivs. Einige Beispiele sind in **Tabelle 1** aufgeführt.

Homogene Kollektive zeigen engere Verteilungen als heterogenere. Substanzen, die über wenige, klar abgegrenzte Pfade inkorporiert werden, zeigen eine breitere Verteilung als Substanzen mit diffusen Einträgen. Beispiele: Die hauptsächliche Quelle für Quecksilber im Urin sind Zahnfüllungen mit Amalgamen. Da die größte Gruppe der Kinder von 6 bis 14 Jahren nicht Amalgamträger sind, aber bei einer Minderheit eine größere Anzahl (> 7) von Zähnen behandelt wurden [13], ist bei ihnen die korrespondierende Verteilung recht breit (s = 1,23). Für die Bleikonzentration im Blut gibt es dagegen eine Reihe von Einflussgrößen, deren jede für sich genommen nur schwach wirksam ist [9,13]; dem zufolge sind die Verteilungen enger (s zwischen 0,21 und 0,39 bei 20- bis 29-jährigen Studenten bzw. s = 0,53 bei 18- bis 69-jährigen Erwachsenen der deutschen Allgemeinbevölkerung). Für die deutlich heterogenere Bevölkerung der USA ermittelte das CDC im Jahr 1999 einen Wert von 0,86 bei den 20- bis 39-jährigen Erwachsenen [14].

Tabelle 1: Beispiele für die Breite der Verteilung verschiedener Xenobiotika-Konzentrationen in einigen Studien des Umweltbundesamtes

Kollektiv	Jahr	Substanz	Breite[d]
Kinder, 6–14 Jahre [13]	1990/92	Hg/Urin	1,23
Kinder, 6–14 Jahre [13]	1990/92	Pb/Blut	0,42
Erwachsene, 18–69 Jahre [13]	1990/92	Pb/Blut	0,53
Studenten, 20–29 Jahre[a] [3]	1990	Pb/Blut	0,21
Studenten, 20–29 Jahre[a,b]	1999	Pb/Blut	0,39
Studenten, 20–29 Jahre[a,b]	2000	Pb/Blut	0,26
Studenten, 20–29 Jahre[a,b]	2001	Pb/Blut	0,31
Erwachsene, 18–69 Jahre [9]	1998	DDE/Blut	0,94
Erwachsene, 18–69 Jahre [9]	1998	HCB/Blut	0,94
Erwachsene, 18–69 Jahre [9]	1998	Pt/Urin	1,35
Studenten, 20–29 Jahre[a,b]	1998	DDE/Plasma	0,53
Studenten, 20–29 Jahre[a,b]	2001	Pd/Urin	0,66
Studenten, 20–29 Jahre[b,c]	2001	Pd/Urin	0,84

Anmerkungen: a: Universität Münster; b: noch nicht veröffentlichte Ergebnisse der Umweltprobenbank; c: Universität Ulm; d: Standardabweichung s der logarithmierten Daten

2.2 Stichprobenfehler des geometrischen Mittels

Entsprechend der empirisch ermittelten Breiten wurden Populationsverteilungen mit einem s von 0,25 bis 1,50 erzeugt. Einige sind in **Abb. 1** dargestellt. Es sind logarithmische Normalverteilungen mit identischem geometrischen Mittelwert von 10, aber unterschiedlicher Breite und Schiefe. Mit fallendem s wandert der Gipfel der Verteilung nach rechts, gleichzeitig wird die Verteilung enger.

Mit wachsendem Stichprobenumfang nimmt der Stichprobenfehler exponentiell ab, mit wachsender Breite steigt er an (**Tabelle 2**). Beispielsweise ist bei einer Stichprobe von 5.000 Probanden und einer Breite s von 1,50 in 95% aller Fälle maximal ein Fehler mit dem Betrag von 4,18% zu erwarten. Nur in 5% aller Fälle wird der Fehler größer sein. Es ist dabei zu beachten, dass der Fehler sowohl nach unten als auch nach oben auftreten kann. Somit beträgt die Schätzgenauigkeit für das geometrische Mittel hier ±4,18%. Eine Halbierung des Stichprobenfehlers erfordert einen etwa 4-fach größeren Stichprobenumfang, da bei 20.000 Probanden der Stichprobenfehler ±1,95% beträgt.

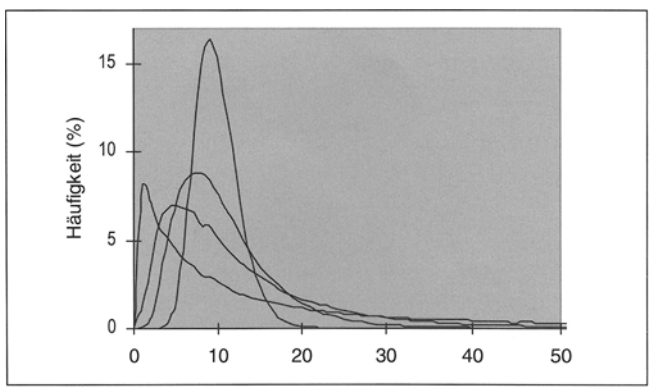

Abb. 1: Einige der verwendeten Populationsverteilungen. (Nur der Wertebereich von 0 bis 50 ist dargestellt.)

Stichprobengröße	Breite s				
	0,25	0,50	0,75	1,00	1,50
	Stichprobenfehler (%)				
10	15,08	32,34	50,03	62,96	105,07
20	10,81	21,78	32,91	44,10	68,05
50	7,03	13,75	20,63	27,21	40,87
100	5,11	9,45	14,20	20,59	30,90
200	3,79	6,81	10,09	15,02	22,23
500	2,27	4,37	6,54	9,15	13,55
1.000	1,64	2,90	4,35	6,61	9,90
2.000	1,17	2,03	3,04	4,64	6,99
5.000	0,71	1,33	1,99	2,79	4,18
10.000	0,52	0,97	1,45	1,88	2,83
20.000	0,38	0,66	0,98	1,30	1,95
50.000	0,26	0,40	0,60	0,76	1,14
100.000	0,19	0,22	0,33	0,43	0,64
150.000	0,15	0,13	0,20	0,24	0,36

2.3 Ableitung der Fehlerfunktion

Aus Tabelle 2 ist ersichtlich, dass zwischen Stichprobengröße und Stichprobenfehler eine nicht-lineare, doch zunächst unbekannte Beziehung besteht. Mittels doppelt logarithmischem Auftragen kann geprüft werden, ob eine Funktion der Klasse $y = b_0x^{b1}$ vorliegt. Funktionen dieser Klasse sind in der Natur sehr häufig. Das doppelt logarithmische Auftragen entspricht formal einem Logarithmieren und Umformen der allgemeinen Funktionsgleichung $y = b_0x^{b1}$ zu $Y = b_1X + b_0$, was einer Geradengleichung entspricht. Wenn also die Werte sich bei der Abtragung zu einer Gerade ordnen, ist erwiesen, dass eine Funktion dieser Klasse vorliegt.

In **Abb. 2** wurden Stichprobengröße und Stichprobenfehler (95. Perzentil der Fehlerverteilung des geometrischen Mittels) für alle Verteilungsbreiten auf diese Weise dargestellt (Symbole in Abb. 2). Es ist erkennbar, dass sich die Punkte linear anordnen. In Folge dessen wurde mit SPSS (Prozedur CURVEFIT) eine Kurvenanpassung an ein Modell der Form (1) vorgenommen. Mit sehr hoher Modellgüte wurden die Koeffizienten b_0 und b_1 für alle Breiten *s* bestimmt (**Tabelle 3**).

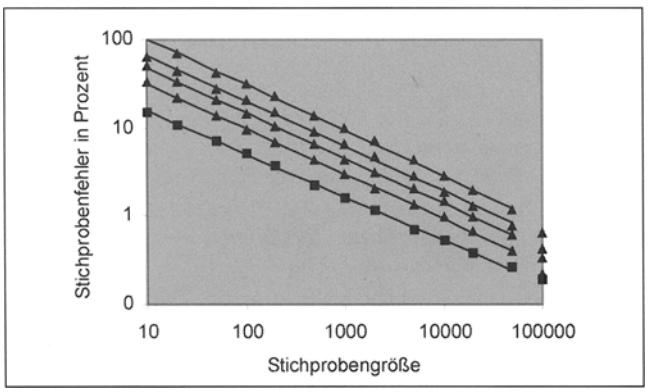

Abb. 2: Stichprobenfehler (95. Perzentil der Fehlerverteilung des geometrischen Mittels) als Funktion der Stichprobengröße für verschiedene Breiten einer logarithmischen Normalverteilung

Tabelle 3: Modellparameter für die Anpassung der Fehlerfunktion an ein exponentielles Modell

s	b_0	b_1	p	R^2
0,25	46,75	–0,4850	<0,0001	0,999
0,50	101,28	–0,5097	<0,0001	1,000
0,75	154,32	–0,5120	<0,0001	1,000
1,00	214,12	–0,5128	<0,0001	0,999
1,50	337,61	–0,5190	<0,0001	0,999

Die allgemeine Form des Modells lautet:

$$E_{0,95;n} = b_0 \cdot n^{b1} \quad (1)$$

Es resultierten also fünf Funktionen, deren Verlauf in Abb. 2 eingetragen wurden (durchgezogene Linien). Um eine gemeinsame Formel für alle Breiten *s* zu finden, wurde versucht b_0 und b_1 als Funktion von *s* darzustellen. Auch diese Anpassungen gelangen mit hoher Modellgüte.

$$b_0 = 214{,}73\ s^{1{,}0994} \quad (2)$$

$$b_1 = -0{,}5252 + 0{,}0097/s \quad (3)$$

Werden (2) und (3) in (1) eingesetzt, erhält man

$$E_{0,95;n;s} = 214{,}73 \cdot s^{1{,}0994} \cdot n^{-0{,}5252 + 0{,}0097/s} \quad (4)$$

oder, vereinfacht, für eine gröbere Schätzung:

$$E_{0,95;n;s} \approx 215 \cdot s \cdot n^{-0{,}5} \quad (5)$$

Umstellung von (4) nach *n* ergibt:

$$n = \left(\frac{E_{0,95;n;s}}{214{,}73 \times s^{1{,}0994}} \right)^{\frac{1}{-0{,}5252+0{,}0097/s}} \quad (6)$$

Mit dieser Gleichung kann der Mindeststichprobenumfang für einen vorgegebenen Stichprobenfehler bei bekannter Breite der Populationsverteilung bestimmt werden.

Beispiel: Für eine Studie im Human-Biomonitoring soll ein Stichprobenfehler von ±5% in Kauf genommen werden. Aufgrund von Vorstudien oder Literaturdaten wird eine Breite der Populationsverteilung von ca. 0,25 erwartet. Nach (6) sind also mindestens 100 Probanden erforderlich. Dieser Wert stimmt recht gut mit den Ergebnissen der Tabelle 2 überein; hier wurde mit einer Stichprobengröße von 100 ein Stichprobenfehler von ±5.11% ermittelt.

Gleichung (6) kann für eine gröbere Schätzung von *n* vereinfacht werden:

$$n \approx \left(\frac{E_{0,95;n;s}}{215 \times s} \right)^{-2} \quad (7)$$

Dieser Schätzer würde im obigen Beispiel einen Mindeststichprobenumfang von 116 ergeben.

3 Diskussion

Der Stichprobenfehler im Human-Biomonitoring hängt nur vom Stichprobenumfang und der Breite der Konzentrationsverteilung ab. Sein Betrag muss bei der Auswertung von Studien im Human-Biomonitoring berücksichtigt werden, vor allem wenn Trendaussagen getroffen werden sollen. Ist beispielsweise der Stichprobenfehler zweier zeitlich aufeinanderfolgender Erhebungen ±5%, dann kann innerhalb dieses Bereichs notwendigerweise keine gesicherte Trendaussage getroffen werden.

Mit den vorgestellten Gleichungen können in der Praxis der erforderliche Mindeststichprobenumfang (bei vorgegebenem Stichprobenfehler) oder der zu erwartende Stichprobenfehler (bei vorgegebenem Stichprobenumfang) geschätzt werden. Dies gilt für Stichproben im Bereich von 10 bis 50.000 Probanden, womit allerdings der Bereich der meisten Studien im Human-Biomonitoring abgedeckt sein dürfte.

Der Stichprobenfehler des geometrischen Mittelwerts ist unabhängig von dessen Höhe, wenn, wie in dieser Arbeit geschehen, ausschließlich der relative Stichprobenfehler betrachtet wird. Daher sind die Ergebnisse dieser Arbeit übertragbar auf Konzentrationsverteilungen im Human-Biomonitoring mit beliebigen geometrischen Mittelwerten, sofern diese einer logarithmischen Normalverteilung folgen. Liegt ein anderer Verteilungstyp vor, dann können die vorgestellten Formeln nicht verwendet werden. Sie gelten ebenfalls nicht, wenn die Stichprobe nach Ziehungsmerkmalen gewonnen wird, die nicht unabhängig sind vom betrachteten Xenobiotikum.

Weiterhin ist der Stichprobenfehler unabhängig von der relativen Stichprobengröße; er wird ausschließlich von der absoluten Stichprobengröße bestimmt. D.h. dass eine Stichprobe eines bestimmten Umfangs den selben Fehler liefern wird, gleichgültig, ob die dahinter liegende Population 1.000 oder 100.000 Personen umfasst. Diese vermeintlich paradoxe Aussage kann veranschaulicht werden, wenn man sich eine logarithmische Normalverteilung nicht als Kurvenzug (wie in Abb. 1) vorstellt, sondern als Säulendiagramm, wobei jede Säule die relative Häufigkeit eines bestimmten Konzentrationsbereichs wiedergibt. Da die Probanden nach dem Zufallsprinzip ausgewählt werden, ist für die Wahrscheinlichkeit ihrer Wahl ausschließlich die relative Häufigkeit, nicht aber die absolute entscheidend. Steigt bei gleichbleibendem Verlauf der Konzentrationsverteilung die Größe der betrachteten Population, so ändern sich lediglich die absoluten, aber nicht die relativen Häufigkeiten. Aus diesem Grund ist der Stichprobenfehler von der Populationsgröße unabhängig.

Nicht betrachtet wurde der analytische Fehler von Messungen im Human-Biomonitoring. Um den Gesamtfehler zu erhalten, müssen Stichprobenfehler, systematischer Fehler der Messung und der Zufallsfehler der Messung zusammen betrachtet werden.

Dieser Beitrag ist keine offizielle Stellungnahme des Umweltbundesamtes und gibt nicht notwendigerweise die Amtsmeinung wieder.

Literatur

[1] Seifert B, Becker K, Helm D, Krause C, Schulz C, Seiwert M (2000): The German Environmental Survey 1990/92 (GerES II): Reference concentrations of selected environmental pollutants in blood, urine, hair, house dust, drinking water, and indoor air. J Expos Anal Environ Epidem 10, 552–565

[2] Hoffmann K, Becker K, Friedrich C, Helm D, Krause C, Seifert B (2000): The German Environmental Survey 1990/92 (GerES II): Cadmium in blood, urine, and hair of adults and children. J Expos Anal Environ Epidem 10, 126–135

[3] UBA (1999): Umweltprobenbank des Bundes-Ausgabe 1999. Ergebnisse aus den Jahren 1996 und 1997. Texte 61/99. Werbung & Vertrieb, Berlin, 593 S.

[4] Pirkle JL, Kaufmann RB, Brody DJ, Hickman T, Gunter EW, Paschal DC (1998): Exposure of the U.S. population to lead, 1991–1994. Environ Health Perspect 106 (11) 745–750

[5] Staessen JA, Roels H, Fagard R (1996): Lead exposure and conventional and ambulatory blood pressure. A prospective population study. JAMA 275 (2) 1563–1570

[6] Sartor FA, Rondia DJ, Claes FD, Staessen JA, Lauwerys RR, Bernard AM, Roels JP, Bruaux PJ, Ducoffre GM, Lijnen PJ, Thijs LB, Amery AK (1992): Impact of environmental cadmium pollution on cadmium exposure and body burden. Arch Environ Health 47 (5) 347–353

[7] Jantunen MJ, Hänninen O, Katsouyanni K (1998): Air pollution exposure in european cities: the EXPOLIS study. J Expos Anal Environ Epidem 8 (4) 495–518

[8] Sachs L (1982): Applied Statistics, Springer, New York

[9] Becker K, Kaus S, Krause C, Lepom P, Schulz C, Seiwert M, Seifert B (2000): Umwelt-Survey 1998 Band III: Human-Biomonitoring. Stoffgehalte in Blut und Urin der Bevölkerung in Deutschland. WaBoLu-Hefte 1/02

[10] Aitchison J, Brown JAC (1957): The Lognormal Distribution. University Press, Cambridge

[11] Gaddum J (1945): Lognormal distribution. Nature156 (4) 463–466

[12] Trapp S, Matthies M (1997): Chemodynamics and Environmental Modeling. Heidelberg: Springer

[13] Krause C, Babisch W, Becker K, Bernigau W, Helm D, Hoffmann K, Nöllke P, Schulz C, Schwabe R, Seifert M, Thefeld W (1996): Umwelt-Survey 1990/92. Band Ia: Studienbeschreibung und Human-Biomonitoring: Deskription der Spurenelementgehalte in Blut und Urin der Bevölkerung in der Bundesrepublik Deutschland. WaBoLu-Hefte 1/96

[14] CDC (2001): National Report on Human Exposure to Environmental Chemicals. Centers for Diseases Control and Prevention. Atlanta, Georgia. March 2001

Eingegangen: 15. Mai 2003
Akzeptiert: 18. September 2003
OnlineFirst: 15. Dezember 2003