

Varianzanalyse für Messwertwiederholungen

– Artikel Nr. 22 der Statistik-Serie in der DMW –

Analysis of variance for repeated measurements

Autoren

R. Bender¹ U. Grouven¹ A. Ziegler²

Institut

¹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

² Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Universität zu Lübeck

Varianzanalyse

Mit Hilfe der Varianzanalyse (analysis of variance, ANOVA) wird untersucht, ob sich drei oder mehr unabhängige Gruppen bezüglich ihrer Mittelwerte einer stetigen Zielvariable statistisch signifikant unterscheiden [1,4]. Die Annahmen, die dieser Methode zugrunde liegen, wurden im Rahmen der DMW-Statistik-Serie schon diskutiert [3,4]. Wichtig ist, dass die Varianzanalyse als Verallgemeinerung des ungepaarten *t*-Tests [3] nur geeignet ist, um die Mittelwerte **unabhängiger** Gruppen zu vergleichen.

Inter- und Intra-Subjekt-Faktoren

Vergleicht man unabhängige Gruppen miteinander, z. B. mehrere Behandlungsarme im Rahmen einer randomisierten Studie, wird hiermit ein **Inter-Subjekt-Faktor** untersucht. Seine Faktorstufen stellen die unterschiedlichen Gruppen dar. Wurden jedoch bei denselben Probanden Messungen dreimal oder öfter durchgeführt, und möchte man diese abhängigen Stichproben miteinander vergleichen, sind die gewöhnlichen ANOVA-Methoden nicht flexibel genug. Sie können nur für unabhängigen Stichproben eingesetzt werden, da sie die Korrelation innerhalb der Daten, die z. B. durch Messwiederholung an ein und derselben Person entsteht, nicht berücksichtigen können. Bei Messwertwiederholungen ist die Zeit ein so genannter **Intra-Subjekt-Faktor**.

Häufig findet man in der Praxis Designs, die sowohl Inter- als auch Intra-Subjekt-Faktoren beinhalten. Liegt die Hauptfragestellung der Studie in der Untersuchung von Gruppenunterschieden, also der Inter-Subjekt-Faktoren, so kann man häufig die Intra-Subjekt-Faktoren durch Verwendung geeigneter Kenngrößen (z. B. Berechnung von Mittelwert oder Fläche unter der Kurve für jeden Probanden) aus dem Design „he-

rausmitteln“. Es bleibt ein Design mit Inter-Subjekt-Faktoren übrig, das mit den Methoden der gewöhnlichen Varianzanalyse oder – im einfachsten Fall – mit Hilfe des *t*-Tests untersucht werden kann. Dieses Vorgehen haben wir im Artikel über Verlaufskurven näher erläutert [2].

Sollen in einer Studie jedoch Effekte eines Intra-Subjekt-Faktors untersucht werden, wird eine flexiblere Methode benötigt, die auch die Betrachtung abhängiger Stichproben erlaubt. Ein solches Verfahren ist die Varianzanalyse für Messwertwiederholungen [7]. Je nach Studiendesign und Datenlage ergeben sich unterschiedliche Auswertungsmodelle. Die grundlegenden Modelle der Varianzanalyse für Messwertwiederholungen werden im Folgenden kurz beschrieben und erklärt.

Modell mit einem Intra-Subjekt-Faktor

Das einfachste Modell der Varianzanalyse für Messwertwiederholungen ist das mit nur einem Intra-Subjekt-Faktor. Besteht dieser Faktor aus lediglich zwei Kategorien (zwei abhängige Stichproben), kommt der gepaarte *t*-Test für die Datenanalyse in Frage. Liegen jedoch drei oder mehr Faktorstufen vor, so wird die Varianzanalyse für Messwertwiederholungen als Auswertungsmethode benötigt.

Wir betrachten als Beispiel die Giessener Senioren-Langzeitstudie (GISELA), in der die Ernährungsgewohnheiten und der Gesundheitsstatus von ca. 500 Senioren mit einem Mindestalter von 60 Jahren von 1994 bis 2002 untersucht werden [9]. Es wird hier nur der Teildatensatz von *n* = 185 Probanden (121 Frauen und 64 Männer) berücksichtigt, bei denen bezüglich des Body Mass Index (BMI) die Basis- und die Follow-Up-Daten nach 2, 4 und 6 Jahren vollständig vorhanden sind. Außerdem betrachten wir hier nur den BMI als Zielgröße und als mögliche Ein-

Schlüsselwörter

- ▶ Varianzanalyse
- ▶ Messwertwiederholungen
- ▶ Verlaufskurven
- ▶ Wechselwirkung

Key words

- ▶ Analysis of variance (ANOVA)
- ▶ Repeated measurements
- ▶ Serial data curves
- ▶ Interaction

Bibliografie

DOI 10.1055/s-2007-959045
Dtsch Med Wochenschr 2007;
132: e61–e64 · © Georg Thieme
Verlag KG Stuttgart · New York ·
ISSN 0012-0472

Korrespondenz

Privatdozent Dr. rer. biol. hum.

Ralf Bender

Institut für Qualität und
Wirtschaftlichkeit im Gesund-
heitswesen (IQWiG)
Dillenburger Straße 27
51105 Köln
eMail Ralf.Bender@iqwig.de

Tab. 1 Mittelwerte und Standardabweichungen (SD) bezüglich des BMI (Angaben in kg/m²) im Zeitverlauf von 185 Probanden der Giessener Senioren Langzeitstudie (GISELA).

BMI (kg/m ²)	Männer (n = 64)		Frauen (n = 121)		Gesamt	
	Mittelwert	SD	Mittelwert	SD	Mittelwert	SD
Basisjahr	26,15	3,29	26,04	3,90	26,08	3,69
Nach 2 Jahren	26,37	3,36	26,30	3,87	26,33	3,69
Nach 4 Jahren	26,46	3,36	26,25	4,03	26,32	3,80
Nach 6 Jahren	26,58	3,24	26,55	4,06	26,56	3,78

Tab. 2 Varianzanalysetabelle für das Modell mit nur einem Intra-Subjekt-Faktor zur Untersuchung von Zeiteffekten bezüglich des BMI bei 185 Probanden der GISELA-Studie.

	Quadratsumme	Freiheitsgrade	Mittel der Quadrate	F	p-Wert	Korrigierter p-Wert (H-F)
Zeit	21,945	3	7,315	9,56	< 0,0001	< 0,0001
Fehler	422,510	552	0,765			

flussgrößen die Zeit und das Geschlecht. In **Tab. 1** sind die Mittelwerte und Standardabweichungen dieses Teildatensatzes im Zeitverlauf angegeben. Anhand der Mittelwerte lässt sich bei beiden Geschlechtern ein leichter Anstieg des BMI im Zeitverlauf feststellen.

Um zu untersuchen, ob die Zeit einen statistisch signifikanten Effekt auf den BMI hat, kann man die Varianzanalyse für Messwertwiederholungen mit einem Intra-Subjekt-Faktor (Zeit) verwenden. Da Messungen im Basisjahr und nach 2, 4 und 6 Jahren vorliegen, handelt es sich um ein Design mit vier abhängigen Stichproben. Eine Mittelung der Daten über die Zeit kommt hier nicht in Frage, da ja gerade die Veränderung des BMI im Zeitverlauf untersucht werden soll.

Bei der Varianzanalyse für Messwertwiederholungen unterscheidet man grundsätzlich zwei verschiedene Sichtweisen, aus denen zwei verschiedene statistische Auswertungsstrategien hervorgehen. Zum einen kann man die abhängigen Messungen eines Probanden als Realisationen von vier verschiedenen Zielvariablen ansehen (multivariater Ansatz). Zum anderen kann man eine univariate Sichtweise einnehmen, bei der nur eine Zielvariable vorliegt, muss aber dann die Abhängigkeitsstruktur der Daten berücksichtigen (univariater Ansatz). Der multivariate Ansatz erfordert weniger Annahmen über die Abhängigkeitsstruktur, hat aber auch weniger Power als der univariate Ansatz [7]. Zudem ist der multivariate Ansatz nur anwendbar, wenn die Zahl der Faktorstufen des Intra-Subjekt-Faktors relativ klein ist. Im Folgenden beschränken wir uns daher auf die Darstellung der Methoden für den univariaten Ansatz.

Überraschenderweise dürfen bei geeigneter Modellierung – auch wenn die Stichproben abhängig sind – die gewöhnlichen *F*-Tests der Varianzanalyse verwendet werden, wenn die Korrelation der Messungen ein gewisses symmetrisches Muster aufweist, die so genannte Sphärizität [7]. Ist diese Annahme verletzt, so sind die gewöhnlichen *F*-Tests jedoch nicht gültig, da sie zu kleine *p*-Werte liefern. Die Sphärizitätsbedingung ist in der folgenden für die Praxis wichtigen Situation gerade nicht erfüllt. Bei zeitlich aufeinander folgenden Messungen sind dicht beieinander liegende Messungen stärker korreliert als weiter auseinander liegende, so dass sich der Grad der Korrelation systematisch ändert. In unserem Beispiel der GISELA-Studie betra-

gen die Korrelationskoeffizienten des BMI im Basisjahr mit den Messungen nach 2, 4 und 6 Jahren 0,96, 0,93 und 0,91. Das bedeutet, dass die Korrelation der Daten wie erwartet mit der Zeit abnimmt. Für solche Situationen wurden verschiedene Korrekturfaktoren für den *F*-Test vorgeschlagen, die zu korrigierten *p*-Werten führen. Sie sind auch bei Verletzung der Sphärizitätsbedingung gültig. Einer dieser Korrekturfaktoren geht auf Huynh und Feldt [8] zurück und wird als „Huynh-Feldt-Epsilon-Korrektur“ bezeichnet [7].

Dieses Verfahren wird auf unsere Beispieldaten angewendet. In **Tab. 2** wird die Varianzanalysetabelle nach dem univariaten Ansatz zum Test auf Zeiteffekte bezüglich des BMI dargestellt. Es wird sowohl der *p*-Wert für den gewöhnlichen *F*-Test, als auch der korrigierte *p*-Wert mit Huynh-Feldt-Epsilon-Korrektur (H-F) angegeben. Aufgrund des hier hoch signifikanten Zeiteffekts (*p* < 0,0001), gibt es in diesem Beispiel keinen relevanten Unterschied zwischen dem nicht korrigierten und dem korrigierten *p*-Wert.

Modell mit einem Intra- und einem Inter-Subjekt-Faktor

Um neben dem Zeiteffekt auch zu untersuchen, ob sich der BMI zwischen Männern und Frauen in der GISELA-Studie signifikant unterscheidet, muss das Modell erweitert werden. Da es sich bei den beiden Geschlechtergruppen um unabhängige Stichproben handelt, ist das Geschlecht ein Inter-Subjekt-Faktor. Für Inter-Subjekt-Faktoren können im Rahmen des Varianzanalysemodells für Messwertwiederholungen die gewöhnlichen *F*-Tests ohne Korrektur verwendet werden. Bei den Modellen der Varianzanalyse für Messwertwiederholungen, die sowohl Inter- als auch Intra-Subjekt-Faktoren enthalten, werden zwei Varianzanalysetabellen erstellt, eine für die Inter- und eine für die Intra-Subjekt-Faktoren. Zu beachten ist, dass es nicht nur Haupteffekte geben kann, sondern auch Wechselwirkungen [4] zwischen den Inter- und Intra-Subjekt-Faktoren, die in der Tafel der Varianzanalyse für die Intra-Subjekt-Faktoren mit aufgenommen werden. Die Bedeutung von Wechselwirkungen wird ausführlich im nächsten Abschnitt diskutiert. In unserem Beispiel gibt es also insgesamt Ergebnisse von drei zu unterscheidenden Effekten: Zeit, Geschlecht und die Wechselwirkung zwischen Zeit und Geschlecht. Die Betrachtung der Wechselwirkung ist deshalb wichtig, da es ja durchaus sein kann, dass es unterschiedli-

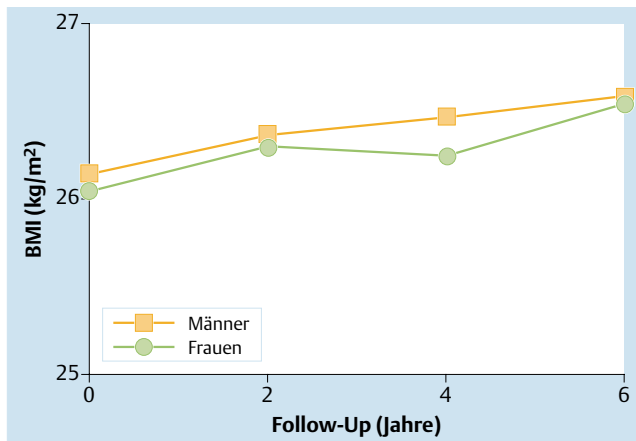


Abb. 1 Mittelwerte des BMI (Angaben in kg/m²) von 185 Frauen und Männern der Giessener Senioren Langzeitstudie (GISELA) im Zeitverlauf.

che Verlaufskurven des BMI bei Männern und Frauen geben kann, die sich nicht allein durch vorhandene Haupteffekte darstellen lassen. Tab. 3a und 3b sind die beiden Varianzanalysetabellen, in **Abb. 1** die mittleren Verlaufskurven für Frauen und Männer grafisch dargestellt.

Wie im einfaktoriellen Modell des letzten Abschnitts gibt es auch im mehrfaktoriellen Modell einen signifikanten Zeiteffekt (korrigierter p -Wert mit Huynh-Feldt-Epsilon-Korrektur: $p=0.0002$). Obwohl die Verlaufskurve der Männer zu allen Zeitpunkten oberhalb der Kurve der Frauen verläuft, ist der Effekt des Geschlechts nicht statistisch signifikant ($p=0.854$). Ebenso wenig gibt es eine statistisch signifikante Wechselwirkung zwischen Zeit und Geschlecht (korrigierter p -Wert mit Huynh-Feldt-Epsilon-Korrektur: $p=0.7263$). Mit den Daten dieses Beispiels lässt sich also nur eine Änderung des BMI im Zeitverlauf, nicht aber ein Unterschied zwischen Männern und Frauen, weder im Mittel über die Zeit noch im Zeitverlauf belegen.

Bedeutung der Wechselwirkung

Um die Bedeutung der Wechselwirkung zwischen einem Inter- und einem Intra-Subjekt-Faktor im Rahmen eines Varianzanalysemodells für Messwertwiederholungen zu illustrieren, vertau-

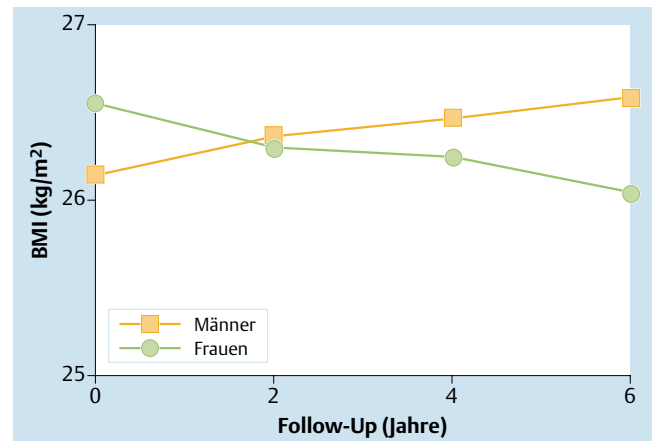


Abb. 2 Künstliche Daten der Giessener Senioren-Langzeitstudie (GISELA) nach Vertauschen der Zeitpunkte 0 und 6 bei den Frauen zur Illustration einer Wechselwirkung.

schen wir einmal die Daten für die Zeitpunkte 0 und 6 Jahre nur bei den Frauen. Die resultierenden Verlaufskurven sehen dann aus wie in **Abb. 2**.

Die Zeitverläufe in diesen künstlichen (!) Daten sind nun völlig unterschiedlich. Bei Männern gibt es wie vorher einen kontinuierlichen Anstieg, während die Verlaufskurve der Frauen jetzt kontinuierlich abfällt. Auf den F -Test für den Inter-Subjekt-Effekt Geschlecht hat diese Datenmanipulation keinen Einfluss. Da man quasi den über alle Zeitpunkte gemittelten Unterschied betrachtet, spielt die Reihenfolge der Zeitpunkte gar keine Rolle. Somit bleibt es auch bei diesen Daten dabei, dass es keinen signifikanten Haupteffekt des Geschlechts gibt ($p=0.854$). Die Datenmanipulation hat aber eine drastische Änderung in den statistischen Ergebnissen der Varianzanalyse bezüglich des Zeitverlaufs zur Folge. Es gibt nun keinen signifikanten Haupteffekt der Zeit mehr ($p=0.935$), aber dafür eine signifikante Wechselwirkung zwischen Geschlecht und Zeit ($p < 0.0001$).

An diesem künstlichen Beispiel wird deutlich, dass man bei der Interpretation nicht signifikanter Haupteffekte, aber signifikanter Wechselwirkungen zwischen Inter- und Intra-Subjekt-Faktoren sehr vorsichtig sein muss. Eine Aussage der Art, dass es – wegen der nicht signifikanten Haupteffekte – keinen Unterschied im BMI zwischen Männern und Frauen und keinen Zeiteffekt gibt, wäre nicht korrekt, sondern irreführend. Es gibt zwar keinen

Tab. 3a Varianzanalysetabelle für den Inter-Subjekt-Effekt aus dem Modell mit einem Inter- und einem Intra-Subjekt-Faktor zur Untersuchung von Geschlechtsunterschieden und Zeiteffekten bezüglich des BMI bei 185 Probanden der GISELA-Studie.

	Quadratsumme	Freiheitsgrade	Mittel der Quadrate	F	p -Wert
Geschlecht	1,835	1	1,835	0,03	0,854
Fehler	9890,1	183	54,044		

Tab. 3b Varianzanalysetabelle für den Intra-Subjekt-Effekt und die Wechselwirkung aus dem Modell mit einem Inter- und einem Intra-Subjekt-Faktor zur Untersuchung von Geschlechtsunterschieden und Zeiteffekten bezüglich des BMI bei 185 Probanden der GISELA-Studie.

	Quadratsumme	Freiheitsgrade	Mittel der Quadrate	F	p -Wert	Korrigierter p -Wert (H-F)
Zeit	18,942	3	6,314	8,22	< 0,0001	0,0002
Zeit \times Geschlecht	0,799	3	0,266	0,35	0,7915	0,7263
Fehler	421,711	549	0,768			

Zeiteffekt, der für Männer und Frauen gleich ist, aber sehr wohl Effekte im Zeitverlauf, nämlich unterschiedliche Trends zwischen den Geschlechtern. Und damit gibt es auch einen Unterschied zwischen Männern und Frauen, nämlich einen abfallenden Trend bei Frauen und einen ansteigenden Trend bei Männern. Ein solches Phänomen lässt sich mit Hilfe eines Wechselwirkungsterms im Rahmen von Varianzanalysemodellen für Messwertwiederholungen beschreiben und auch statistisch absichern.

Häufig werden in der Praxis einfach getrennte Analysen für Frauen und Männer durchgeführt. Dieses Vorgehen ist statistisch nicht so effektiv wie eine gemeinsame Analyse aller Daten, und es lässt sich nicht mehr untersuchen, ob gefundene Unterschiede im Zeitverlauf denn auch statistisch signifikant sind.

Wir möchten noch einmal darauf hinweisen, dass die Auswertungen in diesem Abschnitt nur didaktischen Zweck haben und nicht den tatsächlichen Verlauf des BMI bei den Frauen der GISELA-Studie widerspiegeln. Natürlich belegen die echten Daten einen Anstieg des BMI im Verlauf der Zeit auch bei den Frauen.

Bemerkungen

In diesem Artikel haben wir kurz die wesentlichen Eigenschaften der klassischen Varianzanalyse für Messwertwiederholungen vorgestellt. Zur Analyse von Daten mit einer gewissen Abhängigkeitsstruktur gibt es eine Reihe anderer möglicher Verfahren. Die Wahl eines geeigneten Verfahrens hängt stark von der jeweiligen Datensituation sowie der genauen Fragestellung ab. Liegen z. B. sehr viele Zeitpunkte vor, so kommen eher regressionsanalytische Verfahren für abhängige Daten in Frage [7], auf die wir hier nicht weiter eingehen können. Die hier vorgestellten Methoden der Varianzanalyse für Messwertwiederholungen sind geeignet, wenn die Zahl der abhängigen Stichproben nicht zu groß ist.

Eine Voraussetzung für den sinnvollen Einsatz der Varianzanalysemodelle für Messwertwiederholungen ist jedoch, dass der Datensatz bezüglich der Intra-Subjekt-Faktoren vollständig ist, d. h. dass es bei allen Probanden zu allen Zeitpunkten Messungen geben muss. Fehlt z. B. bei fünf Zeitpunkten bei einem Probanden auch nur ein Wert, so können entweder alle anderen Daten dieses Probanden im Rahmen einer Varianzanalyse für Messwertwiederholungen nicht für die Auswertung verwendet werden (Reduzierung des Stichprobenumfangs) oder man lässt diesen Zeitpunkt auch bei allen anderen Probanden weg (Reduzierung der Zahl der abhängigen Stichproben). Als dritte Möglichkeit kommt noch in Frage, die fehlenden Werte geeignet zu ersetzen. Dies führt aber auch nur unter gewissen Annahmen zu gültigen Schlussfolgerungen, und es dürfen nicht zu viele Werte fehlen. In Beobachtungsstudien ist es jedoch häufig die Regel, dass zu vielen verschiedenen Zeitpunkten viele Werte nicht erhoben werden konnten oder von vorne herein Messungen an Probanden zu unterschiedlichen Zeitpunkten geplant waren. Für solche unbalancierten Designs stellt die Varianzanalyse für Messwertwiederholungen kein geeignetes Verfahren dar. Im Beispiel der GISELA-Studie reduzierte sich der Stichprobenumfang von $n = 532$ auf $n = 182$, obwohl von 6 möglichen Zeitpunkten nur 4 betrachtet wurden. Daher ist das Verfahren der Varianzanalyse für Messwertwiederholungen hier sehr ineffizient und wurde für die Datenanalyse hier nur für didaktische Zwe-

Tab. 4 Übersetzung (deutsch – englisch)

Varianzanalyse	analysis of variance (ANOVA)
Inter-Subjekt-Faktor	between-subject factor
Intra-Subjekt-Faktor	within-subject factor
gepaarter t-Test	paired t test
Messwertwiederholungen	repeated measurements
Sphärizität	sphericity
Varianzanalysetabelle	ANOVA table
Wechselwirkung	interaction
unbalanciert	unbalanced
lineare gemischte Modelle	linear mixed models
verallgemeinerte Schätzgleichungen	generalized estimating equations (GEE)

cke gewählt, nicht jedoch bei der Publikation der Studienergebnisse [9]. Eine weitaus bessere Möglichkeit zur Analyse dieser Daten stellen neuere statistische Methoden dar, die es erlauben, dass die Zahl der Messungen je Proband unterschiedlich ist. In Frage kommen hier lineare gemischte Modelle [6] oder Verfahren, die auf so genannten verallgemeinerten Schätzgleichungen (generalized estimating equations, GEE) beruhen [5]. Auf diese Methoden werden wir in zukünftigen Artikeln der Statistik-Serie eingehen. Bei der Datenanalyse der GISELA-Studie wurden lineare gemischte Modelle eingesetzt [9]. Die englischen Bezeichnungen der hier diskutierten Begriffe zeigt Tab. 4.

kurzgefasst

Mit Hilfe der Varianzanalyse für Messwertwiederholungen lassen sich Unterschiede zwischen mehr als zwei abhängigen Stichproben bezüglich einer stetigen Zielvariablen statistisch untersuchen (Effekte von Intra-Subjekt-Faktoren). Eine gleichzeitige Untersuchung von Unterschieden zwischen unabhängigen Gruppen ist möglich durch Modelle, die sowohl Intra- als auch Inter-Subjekt-Faktoren beinhalten. Neben Haupteffekten können auch Wechselwirkungen zwischen Intra- und Inter-Subjekt-Faktoren eine Rolle spielen, mit denen z. B. unterschiedliche Zeitverläufe zwischen verschiedenen Gruppen beschrieben werden können.

Literatur

- Altman DG, Bland JM. Comparing several groups using analysis of variance. *BMJ* 1996; 312: 1472–1473
- Bender R, Lange S. Verlaufskurven. *Dtsch Med Wochenschr* 2007; 132: e22–e23
- Bender R, Lange S, Ziegler A. Wichtige Signifikanztests. *Dtsch Med Wochenschr* 2007; 132: e24–e25
- Bender R, Ziegler A, Lange S. Varianzanalyse. *Dtsch Med Wochenschr* 2007; 132: e57–e60
- Burton P, Gurrin L, Sly P. Tutorial in Biostatistics: Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med* 1998; 17: 1261–1291
- Cnaan A, Laird NM, Slasor P. Tutorial in Biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med* 1997; 16: 2349–2380
- Crowder MJ, Hand DJ. *Analysis of Repeated Measures*. Chapman & Hall: London, 1990
- Huynh H, Feldt LS. Estimation of the Box correction for degrees of freedom for sample data in randomised block and split-plot designs. *J Educ Stat* 1976; 1: 69–82
- Jungjohann SM, Lührmann P, Bender R, Blettner M, Neuhäuser-Berthold M. Eight-year trends in food, energy and macronutrient intake in a sample of elderly German subjects. *Br J Nutr* 2005; 93: 361–368