

Int Poster J Dent Oral Med 1999, Vol 1 No 3, Poster 20

## Der Vergleich von Reliabilitäten zweier bildgebender Verfahren bei asymmetrischen Item-spezifischen Bias

**Sprache:** Deutsch

**Autor:** Frank Krummenauer

Department of Medical Statistics and Documentation, University of Mainz, Germany

**Datum/Veranstaltung/Ort:**

45. Jahrestagung der Deutschen Region der Internationalen Biometrischen Gesellschaft

16.-19.03.1999

Dortmund, Deutschland

### Fragestellung

Beim Vergleich bildgebender Verfahren müssen auch deren Reproduzierbarkeiten gegenübergestellt werden, insbesondere wenn keine der beiden Diagnosemethoden eine exakt messende Referenz darstellt. Das separate Bestimmen der jeweiligen instrumentellen Reliabilitäten (siehe etwa Prawat et al., 1995) führt jedoch zu drastischen Verzerrungen und mitunter sogar umgekehrten Effekten, wenn eine der beiden Methoden einem merklichen Item-spezifischen Bias unterliegt, also eine Interaktion zwischen Messverfahren und Studienteilnehmern, welche nur bei wenigen der Studienteilnehmer auftritt, dann aber bei jeder Messwiederholung verzerrend zum Tragen kommt. Es soll daher ein Modell zur simultanen Schätzung der Reliabilitäten motiviert werden, welches die Schätzwerte der Reliabilitäten direkt an einen potentiellen Item-spezifischen Bias adjustiert.

### 1. Die DIGIGRAPH-Studie

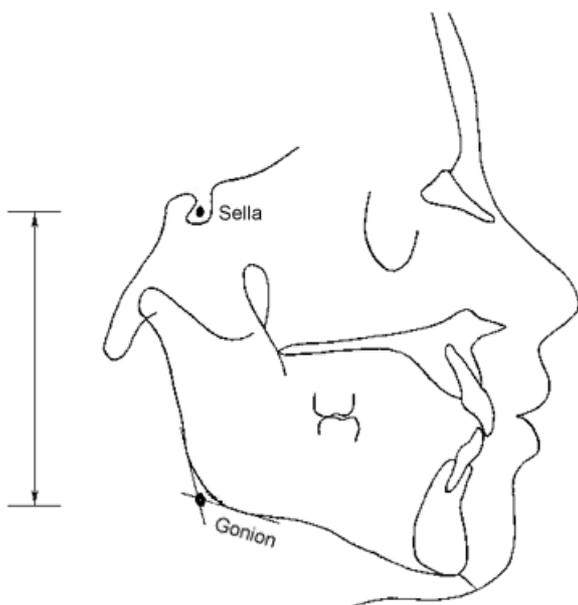
Frage: Kann der kephalometrische Goldstandard (laterales Fern-Röntgenbild) in der Diagnostik kieferorthopädischer Patienten durch das Videosystem DigiGraph ersetzt werden?

Motivation: DigiGraph erzeugt keine Strahlenbelastung und ermöglicht häufige Messungen im Laufe einer Therapie

Studie: 50 13-jährige Patientinnen der Klinik und Poliklinik für Kieferorthopädie der Universität Mainz (Direktor: Prof. Dr. Sergl)

Merkmal: posteriore Gesichtshöhe (Strecke Sella-Gonion nach Jarabak) als einer von 32 untersuchten Parametern

Daten: je zwei Meßauswertungen pro Methode und Patientin



## 2. Das Ergebnis

Separate Schätzung der Reliabilitäten nach Fleiss (1986):

$R_1 = 0.99$  DigiGraph

$R_2 = 0.72$  Fernröntgen (Goldstandard)

**also:** DigiGraph liefert merklich reliablere Meßwerte

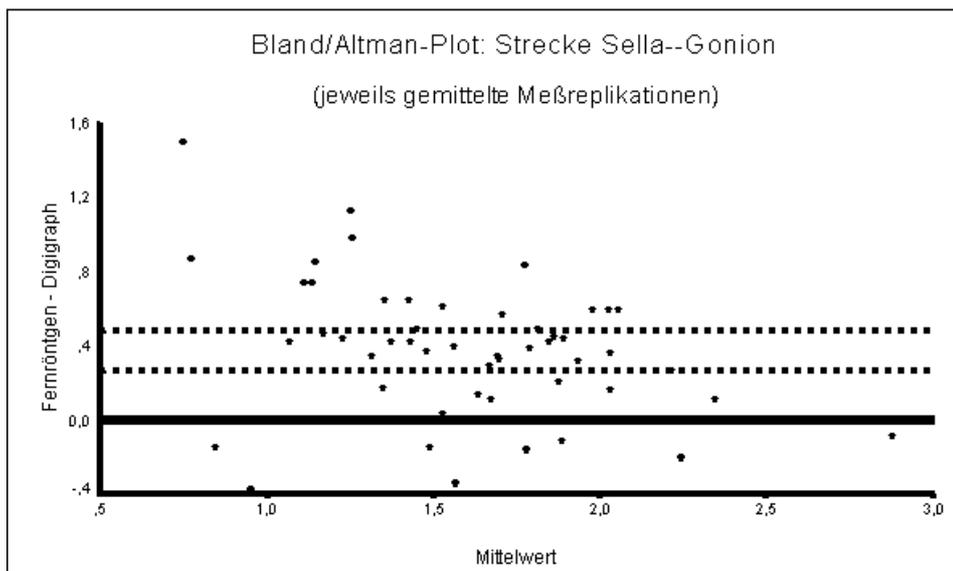
### Univariate Ergebnisse:

siehe Bland/Altman-Plot angewandt auf die jeweils gemittelten Replikationen

- DigiGraph unterschätzt systematisch (und signifikant)
- DigiGraph zeigt merklich (und signifikant) geringere Präzision

???????????

**Wieso zeigt das reliablere Diagnoseverfahren die geringere Präzision?**



## 3. Der Stolperstein

**Problem:** Separate (!) Schätzung der Reliabilitäten

**Modell:**  $X_i = T_i + E_i$  mit "true scores"  $T_i$  und "Messfehlern"  $E_i$   
( $i = 1$ : DigiGraph,  $i = 2$ : Fernröntgen)

**Reliabilität:**  $R_i = \frac{\text{var}(T_i)}{\text{var}(T_i) + \text{var}(E_i)}$

**Klinik:** DigiGraph fixiert bei manchen Kindern die Sella nicht korrekt; bei jeder DigiGraph-Replikation wird die Strecke Sella-Gonion verzerrt erhoben

**aber:** dies ist nicht der Fall beim Fernröntgen

**also:** asymmetrischer Item-spezifischer Bias nicht berücksichtigt

**Ausweg:** Konstitution eines Modells zur

- Schätzung einer zusätzlichen Varianzkomponente für diesen Bias
- Adjustierung der Reliabilitäten an dessen asymmetrisches Auftreten

**also:** gemeinsames Modell mit Korrelation der "true scores" als Parameter

**damit:** simultane Schätzung aller Parameter (LISREL: Long, 1983)

## 4. Ein Modell-Vorschlag

Umparametrisiertes Faktoranalyse-Modell:

$$X_{ih}^{[j]} = w_i^{[j]} + e_{ih}^{[j]} \quad \begin{cases} i = 1, 2 & \text{(Methode)} \\ j = 1, \dots, 50 & \text{(Patientin)} \\ h = 1, 2 & \text{(Replikation)} \end{cases}$$

$X_{ih}^{[j]}$  = Replikation  $h$  mit Methode  $i$  an Kind  $j$

$w_i^{[j]}$  = Interaktion zwischen Kind  $j$  und Methode  $i$

$$\text{var}(w_i^{[j]}) = \sigma_{wi}^2$$

$e_{ih}^{[j]}$  = residueller Fehler (Replikation  $h$ , Methode  $i$ , Kind  $j$ )

$$\text{var}(e_{ih}^{[j]}) = \sigma_{ei}^2$$

- alle  $w_i^{[j]}$  und  $e_{ih}^{[j]}$  normalverteilt
- $e_{ih}^{[j]}$  untereinander und mit  $w_i^{[j]}$  unkorreliert

**Intrinsische Korrelation der "true scores":**

$$\rho = \text{Corr}(w_1^{[j]}, w_2^{[j]})$$

- Validitätsmaß für DIGIGRAPH gegenüber Goldstandard
- berücksichtigt simultan alle vier Messungen pro Kind
- adjustiert Reliabilitäten an die Kind/Methode-Interaktion:

$$\hat{R}_i = \hat{\rho} \cdot \frac{\hat{\sigma}_{w1} \cdot \hat{\sigma}_{w2}}{\hat{\sigma}_{wi}^2 + \hat{\sigma}_{ei}^2} \quad (i = 1, 2)$$

## 5. Das korrigierte Ergebnis

Varianzkomponenten (zur posterioren Gesichtshöhe):

	$\hat{\sigma}_w^2$	$\hat{\sigma}_e^2$	$\hat{R}$
Röntgen	0.164	0.046	0.72
DIGIGRAPH	0.325	0.002	0.49
$\hat{\rho} = 0.694$			

**Also:** DigiGraph deutlich weniger reliabel als Goldstandard

## Conclusia

Beim derzeitigen Stand nicht bedenkenlos in kieferorthopädischer Diagnostik und Therapie-Kontrolle einsetzbar. Fernziel: Identifikation von Auslösern des obigen "Bias"

Sprich: Bei welchen Kindern scheitert der DigiGraph an der Fixierung der Sella?

## Bibliographie

- Fleiss JL (1986): The Design and Analysis of Clinical Experiments. *John Wiley & Sons. New York*
- Krummenauer F (1999): The Biometrical Comparison of Orthodontic Imaging Methods. *European Journal of Orthodontics. accepted*
- Long JS (1983): Confirmatory Factor Analysis: A Preface to LISREL. *Sage Publications. Beverly Hills and London*
- Prawat JS, Nieberg L, Cisneros GJ, Acs G (1995): A comparison between radiographic and sonically produced cephalometric values. *The Angle Orthodontist 65, 271-276*

Dieses Poster wurde übertragen am 22.07.99 von Dr. Frank Krummenauer.

**Korrespondenzadresse:**

Dr. Frank Krummenauer

Institut für Medizinische Statistik und Dokumentation Mainz

Obere Zahlbacher Strasse 69

D - 55131 Mainz

**Poster Faksimile:**

**Der Vergleich von Reliabilitäten zweier bildgebender Verfahren bei asymmetrischem Item-spezifischen Bias**

Frank Krummenauer  
Institut für Medizinische Statistik und Dokumentation Mainz  
Obere Zahlbacher Straße 69, D-55131 Mainz

**1. Die DIGIGRAPH Studie**

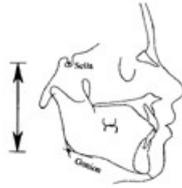
**Frage:** Kann der kraniastrahlentechnische Goldstandard (laterales Fern-Röntgenbild) in der Diagnostik kranioorthopädischer Patienten durch das Videosystem DIGIGRAPH ersetzt werden?

**Motivation:** DIGIGRAPH erzeugt keine Strahlbelastung und ermöglicht leichter Messungen im Laufe einer Therapie

**Studien:** 30 13-jährige Patienten aus der Klinik und Poliklinik für Kieferorthopädie der Universität Mainz

**Merkmale:** posteriore Gesichtsbilder (Strecke-Sella-Gonion nach Jankovic) als eines von 22 unterschiedlichen Parametern (siehe auch erste Kieferorthopädie)

**Daten:** je zwei Messwertreihen pro Methode und Patientin



**2. Das Ergebnis**

Separate Schätzung der Reliabilitäten nach Fleiss (1981)

$R_1 = 0.99$  DIGIGRAPH

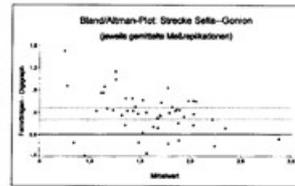
$R_2 = 0.72$  Röntgen (Goldstandard)

**also:** DIGIGRAPH liefert deutlich reliablere Messwerte

**Univariate Ergebnisse:**

- DIGIGRAPH unterschätzt systematisch (und signifikant)
- DIGIGRAPH misst deutlich (und signifikant) geringer Präzision

\*\*\*\*\*  
**Wieso zeigt das reliablere Diagnoseverfahren die geringere Präzision?**



**3. Der Stolperstein**

**Problem:** Separate (?) Schätzung der Reliabilitäten

**Modell:**  $X_{ij} = T_i + E_{ij}$  mit "true scores"  $T_i$  und "Fehlern"  $E_{ij}$  ( $i = 1, \dots, 50$  Patienten,  $j = 1, 2$  Röntgen)

**Reliabilität:**  $R_i = \frac{\text{Var}(T_i)}{\text{Var}(X_{ij})}$

**Klinik:** DIGIGRAPH liefert bei manchen Kindern die Sella nicht korrekt; bei jeder DIGIGRAPH-Replication wird die Strecke-Sella-Gonion vermessen

**also:** asymmetrischer Item-spezifischer Bias nicht berücksichtigt

**Ausweg:** Konstruktion eines Modells zur Schätzung einer zusätzlichen Varianzkomponente für diesen "Bias"

**Annahme:** separate Schätzung "true scores"  $T_i$  als unabhängig angenommen

**also:** gemeinsames Modell mit Korrelation der "true scores" als Parameter

**idea:** "true scores" als gemeinsamen Faktor in Faktorenanalyse-Modell betrachten

**damit:** simultane Schätzung aller Parameter (LISREL; Long, 1983)

**4. Ein Modell-Vorschlag**

1-parametrisches Faktorenanalyse-Modell:

$$X_{ij} = \mu_j + \omega_{ij} \quad \begin{cases} i = 1, 2 & \text{(Methode)} \\ j = 1, \dots, 50 & \text{(Patienten)} \\ A = 1, 2 & \text{(Replication)} \end{cases}$$

$X_{ij}^{(k)}$  = Replication k mit Methode i an Kind j

$\omega_{ij}^{(k)}$  = Interaktion zwischen Kind j und Methode i

$\text{Var}(\omega_{ij}^{(k)}) = \sigma_{\omega_{ij}^{(k)}}^2$

$\sigma_{\omega_{ij}^{(k)}}^2$  = methodischer Fehler (Replication k, Methode i, Kind j)

$\text{Var}(\sigma_{\omega_{ij}^{(k)}}^2) = \sigma_{\omega_{ij}^{(k)}}^4$

- alle  $\omega_{ij}^{(k)}$  und  $\sigma_{\omega_{ij}^{(k)}}^2$  normalverteilt
- $\sigma_{\omega_{ij}^{(k)}}^2$  untereinander und mit  $\omega_{ij}^{(k)}$  unkorreliert

**Intrinsische Korrelation der "true scores":**

$$\rho = \text{Corr}(\omega_{i1}^{(1)}, \omega_{i2}^{(1)})$$

- Validitätsmaß für DIGIGRAPH gegenüber Goldstandard
- berücksichtigt sowohl alle vier Messungen pro Kind
- adjustiert Reliabilitäten an die Kind/Methode-Interaktion:

$$R_i = \rho \frac{\sigma_{\omega_{i1}^{(1)}}^2 + \sigma_{\omega_{i2}^{(1)}}^2}{\sigma_{\omega_{i1}^{(1)}}^2 + \sigma_{\omega_{i2}^{(1)}}^2} \quad (i = 1, 2)$$

**5. Das korrigierte Ergebnis**

**Varianzkomponenten:** (nur posteriore Gesichtsbilder)

	$\mu_j^2$	$\sigma_{\omega_{ij}^{(k)}}^2$	$R$
Röntgen	0.164	0.046	0.72
DIGIGRAPH	0.325	0.002	0.49

$$\rho = 0.694$$

**also:** DIGIGRAPH deutlich weniger reliabel als Goldstandard

**Conclusion:** beim derzeitigen Stand nicht bedenkenlos in kieferorthopädischer Diagnostik und Therapie-Kontrolle einsetzbar

**Fernziel:** Identifikation von Ansätzen des obigen "Bias"

**Spezif:** bei welchem Kindern schneidet die DIGIGRAPH an der Fixierung der Sella?

**Literatur**

1. Fleiss JL (1981) *The Design and Analysis of Clinical Experiments*. John Wiley & Sons, New York
2. Krummenauer F (1999) The Biometrical Comparison of Orthodontic Imaging Methods. *Journal of Dentistry* accepted
3. Long JS (1983) *Confirmatory Factor Analysis: A Preface to LISREL*. Sage Publications, Beverly Hills and London
4. Prewitt JS, Naeberg L, Cameron GJ, Ans G (1993) A comparison between radiographic and visually produced cephalometric values. *The Angle Orthodontist* 65, 271-276