

## **Evaluering av måleegenskaper, 15. og 16. november på NIH**

For å vurdere om målemetoder gir reproduserbare og tolkbare resultater, trenger man hensiktsmessige design og statistiske metoder. Kurset omfatter design og analyse for vurdering av reliabilitet, diskriminerende evne, prediksjon, sensitivitet for endring, intern konsistens og begrepsvaliditet (construct validity). Metoder for analyse av ratiodata, intervalldata, ordinaldata og kategoriske data vil bli gjennomgått.

### **Målgruppe**

Doktorgradsstipendiater. Kurset er godkjent som spesialpensum under den valgfrie delen av opplæringsdelen i doktorgradsutdanningen og gir 2,5 studiepoeng.

### **Målsetting**

Etter gjennomgått kurs skal deltakerne kunne velge egnede metoder for evaluering av reliabilitet, diskriminerende evne, prediksjon og endring over tid. De skal kunne gjennomføre en metodestudie, bruke aktuelle statistiske metoder og vurdere ulike kilder til målefeil.

### **Forkunnskaper**

Gjennomgått forskeropplæringskurs i statistikk eller tilsvarende. Kjennskap til terminologi angående enveis og toveis ANOVA.

### **Undervisningsform**

Forelesninger. Deltakerne oppfordres til å beskrive egne målemetoder ved innledningen til kurset, samt å delta aktivt ved å presentere aktuelle problemstillinger underveis.

### **Kursleder**

Professor, dr. philos. Rolf Moe-Nilssen, Inst. for samfunnsmedisinske fag, Universitetet i Bergen.

### **Tid og sted**

15. og 16. november på NIH

Påmelding til [gro.styrmo@nih.no](mailto:gro.styrmo@nih.no) innen 1. november.

<b>DAG 1</b>	<b>Tema</b>	<b>Innhold</b>
0900-0930	Innledning	Presentasjon av deltakere. Om kurset.
0945-1030	Måleteori	Klassifisering. Måling. Målenivå.
1045-1130	Kilder til målefeil	Presisjon og nøyaktighet. Intra- og intertesterrelabilitet, testretestrelabilitet.
1145-1230	Design av metode-evalueringstudier	Reliabilitetsevaluering av et instrument. Sammenligning mellom instrumenter. Krav til testpopulasjon.
1230-1315	Lunsj	
1315-1400	Vurdering av 'Multi item' instrumenter	Bruk av del- og sumscore. Profiler, indekser og indikatorer.
1415-1500	Mål for reliabilitet	Absolutt og relativ reliabilitet.
1515-1600	Statistisk teori	Samvariasjon og enighet. Tilfeldige og systematiske feil. Trend. Fallgruber. Pearson's r. t-test for pardata.
<b>DAG 2</b>	<b>Tema</b>	<b>Innhold</b>
0800-0845	Statistiske modeller for intervalldata, relativ enighet	Homo- og heteroscedastisitet. Reliabilitet og trend. Intraclass correlation (ICC). Coefficient of variation (CV).
0900-0945	Statistiske modeller for intervalldata, absolutt enighet	Bland-Altman's plot. Intrasubjekt standardavvik ( $s_w$ eller SEM). CI for en måling. CI for en differanse.
1000-1045	Intern konsistens av testbatterier	Cronbach's alfa. Kuder-Richardson 20.
1100-1145	Statistiske modeller for enighet, ordinale og kategoriske data	Kappa for to og flere enn to testere. Veid Kappa.
1145-1230	Lunsj	
1230-1315	Sensitivitet for endring	Mål for endring. Minste målbare endring. Effect size og Standardized Response Mean (SRM) Klinisk viktig endring. Responsiveness Regresjon mot middelverdien.
1330-1415	Statistiske modeller for klassifisering	Sensitivitet og spesifisitet. Konstruksjon og tolkning av ROC-kurve.
1430-1515	Risikovurdering og prediksjon	Retro- og prospektive design. Bruk av overlevelsesanalyse.
1530-1600	Men hva med validiteten?	Reliabilitet, klassifisering, sensitivitet for endring, begrepsvaliditet. Spearmans korreksjon for reliabilitet.
	Avslutning	Oppsummering og evaluering

## **Aktuell litteratur**

Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991, 396-419.

Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997;314:572.

Bland JM, Altman DG. Measurement error [corrected and republished article originally printed in *BMJ* 1996 Jun 29;312(7047):1654]. *BMJ* 1996;313:744.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.

Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *J Rehabil Med* 2009; 41(5):343-346  
Domholdt E. Rehabilitation research. Principles and applications. St. Louis, Miss: Elsevier Saunders, 2005, 245-264.

Loge JH, Kaasa S, Hjermland MJ, Kvien TK. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. *J.Clin.Epidemiol.* 1998;51:1069-76.

Moe-Nilssen R. Test-retest reliability of trunk accelerometry during standing and walking. *Arch.Phys.Med.Rehabil.* 1998;79:1377-85.

Moe-Nilssen R, Nordin E, Lundin-Olsson L. Criteria for evaluation of measurement properties of clinical balance measures for use in fall prevention studies. *Journal of Evaluation in Clinical Practice* 2008;14:236-40.

Moe-Nilssen R, Aaslund MK, Hodt-Billington C, Helbostad JL. Gait variability measures may represent different constructs. *Gait Posture* 2010; In press.

Ottenbacher KJ, Tomchek SD. Reliability analysis in therapeutic research: practice and procedures. *Am.J.Occup.Ther.* 1993;47:10-16.

Shrout PE, Fleiss FL. Intraclass correlations: Uses in assessing rater reliability. *Psychol.Bull.* 1979;86:420-8.