



# ANALYSE AV PROM

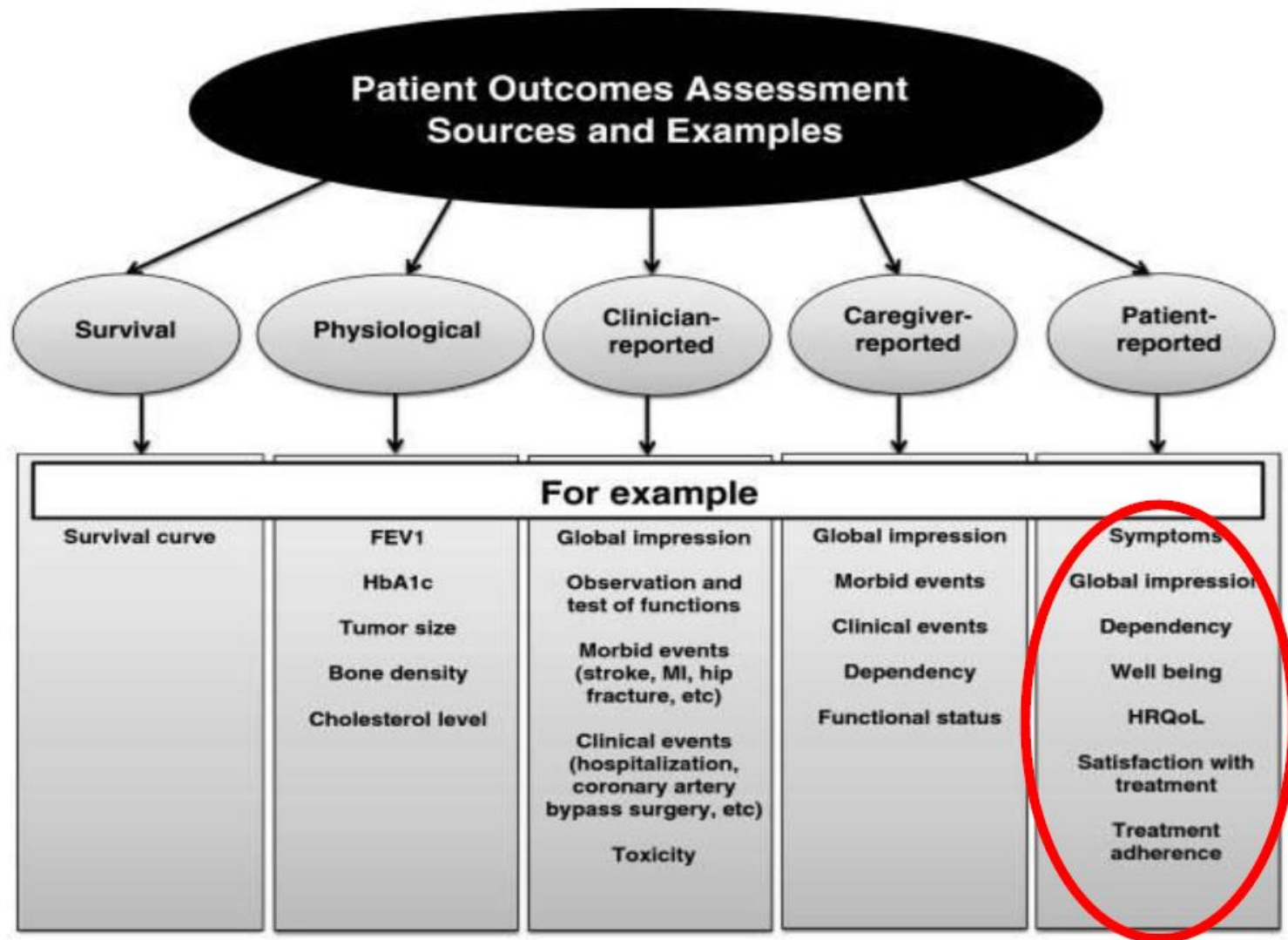
Kyrre Breivik

Fagsenter for pasientrapporterte data

RKBU Vest -Regional Kunnskapssenter  
for Barn og Unge, Norge

# Pasientrapporterte data

- *Pasient rapporterte utfallsmål (Patient Reported Outcome-PRO)*: Rapportering om helsetilstand direkte fra pasienten uten fortolkning fra helsepersonell eller studiemedarbeider (US Food & Drug Administration)
- *Patient Reported Outcome Measures (PROM)*: spørreskjema (instrument) som måler pasientrapporterte utfallsmål
- Patient Reported Experience Measures (PREM): spørreskjema (instrument) som måler pasientenes opplevelse og tilfredshet med helsetjenesten
- En målsetning at alle nasjonale kvalitetsregistre skal inkludere pasientrapporterte data



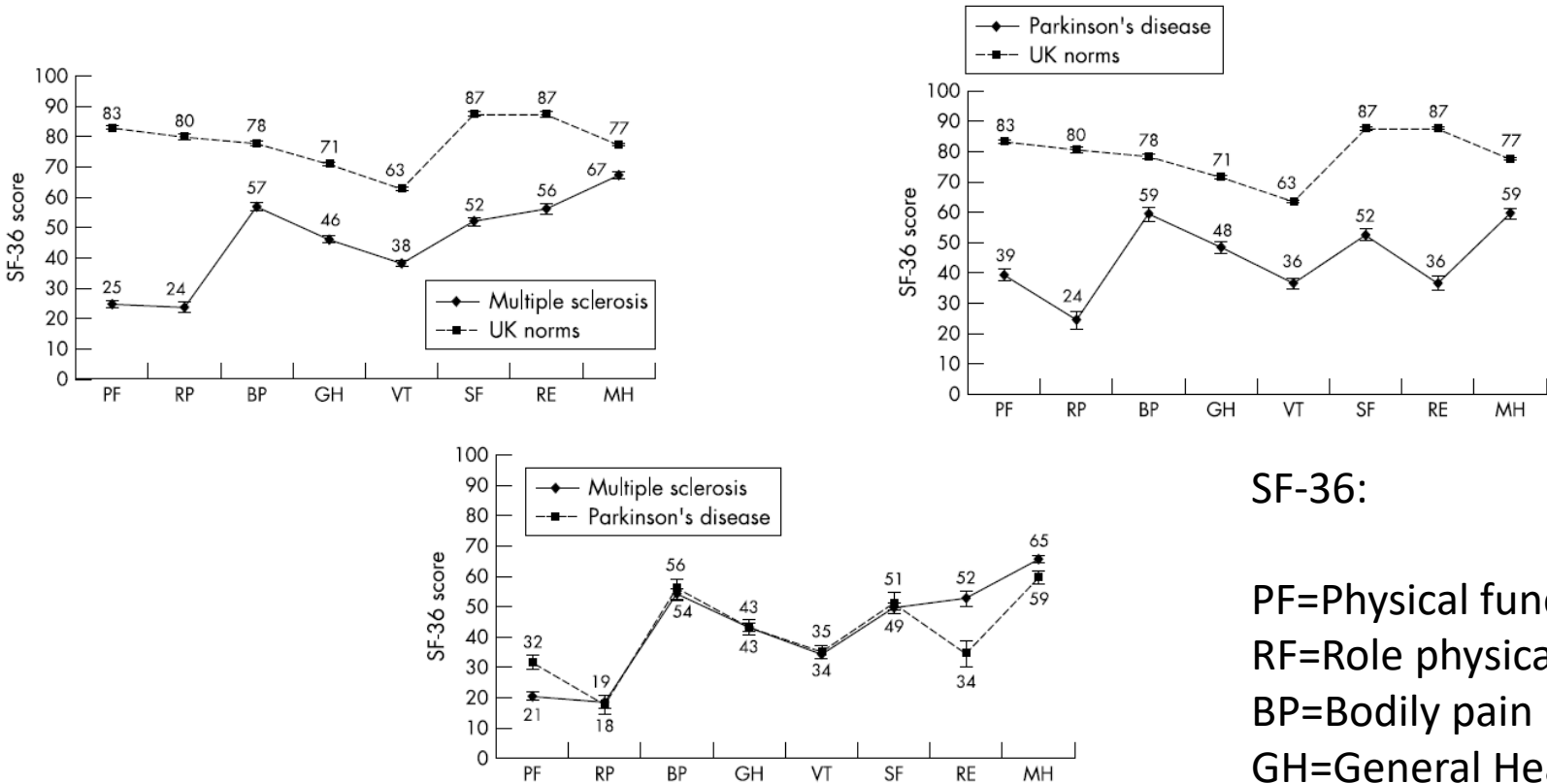
# Hvorfor bruke PROM?

- Blant annet for å få systematisk kunnskap om selvrapportert helse og livskvalitet, bivirkninger i en pasientgruppe
- Dette kan også brukes til å følge opp virksomhetens resultat mht. de problemene pasientene opplever
- Vurdere ulike behandlings innvirkning på pasientenes selvrapporterte helse
- Kan brukes i klinisk praksis
- Brukes i helseøkonomiske analyser (kostnad –nytte)
- Som grunnlag for pasientsikkerhetsarbeid

# Generiske versus sykdomsspesifikke PROM skjema

- Generiske skjemaer måler aspekter på generell helse (RAND-36/SF-36, EQ5D)
  - Styrker:
    - Gjør det mulig å sammenligne helse og effekt av behandling på tvers av pasientgrupper.
    - Kan fange opp uventede behandlingseffekter for sykdomsgrupper
  - Svakheter:
    - Kan være lite sensitiv for forandring på relevante utfallsmål for spesifikke sykdomsgrupper (f.eks. kreft).
    - Fare for at de kan oppleves som for generelle og lite relevant for pasienter.
    - Effekter kan være vanskelig å tolke

# Kan sammenlikne helseprofiler på tvers av grupper



**Figure 4** Graph showing SF-36 scores for multiple sclerosis patients compared with Parkinson's disease patients, controlling for age, sex, marital status, social class, employment, and years since diagnosis of multiple sclerosis or Parkinson's disease. Error bars = SEM. BP, bodily pain; GH, general health perceptions; MH, mental health; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; VT, vitality.

SF-36:

- PF=Physical functioning
- RP=Role physical
- BP=Bodily pain
- GH=General Health
- VT=Vitality
- SF=Social functioning
- RE=Role emotional
- MH=Mental health

Riazi, A., Hobart, J. C., Lamping, D. L., Fitzpatrick, R., Freeman, J. A., Jenkinson, C. A. A., ... & Thompson, A. J. (2003). Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(6), 710-714.

# Sykdomsspesifikke skjemaer

- Sykdomsspesifikke skjemaer måler aspekter som er spesielt relevante for den aktuelle tilstanden (f.eks. MIDAS, QLQ-C30)
  - Styrker:
    - Er utformet for å respondere på endring i et behandlingsløp for den aktuelle tilstanden. Kan derfor være mer sensitiv for forandring for denne gruppen pasienter.
    - Vil ofte kunne oppleves som mer relevant for disse pasientene
  - Svakheter:
    - Er ofte lite egnet til å sammenlikne resultatene på tvers av sykdomsgrupper.
    - Kan overse uventede positive/negative behandlingseffekter på helseområder som ikke er med i målingen.
    - Bruk av ulike sykdomsspesifikke instrumenter kan gjøre det vanskelig å kombinere resultatene i f.eks. en meta-analyse
- Ja takk begge deler! Ofte bra å ha med begge deler i et register

# Valg av spørreskjema

- Hva ønsker en å få svar på og hvilket spørreskjema gir best mål på dette?
  - Validitet (hvorvidt det måler det en ønsker at det skal måle)
    - Inholdsvaliditet
    - Konstrukt validitet
  - Reliabilitet (pålitelighet av data)
    - Test-retest
    - Indre konsistens
  - Sensitivitet (instrumentet sin evne til å avdekke forskjeller mellom pasienter eller grupper av pasienter)
  - Responsivitet (Instrumentet sin evne til å fange opp endringer i pasientens tilstand)
- Viktig å være kritisk til hvilken metode som er brukt for å teste måleegenskapene til skjemaet



# Analyse av PROM-data

---



En utfordring med PROM er at en ofte måler relativt abstrakte konstrukt basert på subjektive vurderinger (f.eks. Helse-relatert livskvalitet). Når en skal analysere PROM-data er det viktig å bruke en korrekt målemodell for det konstruktet en ønsker å måle

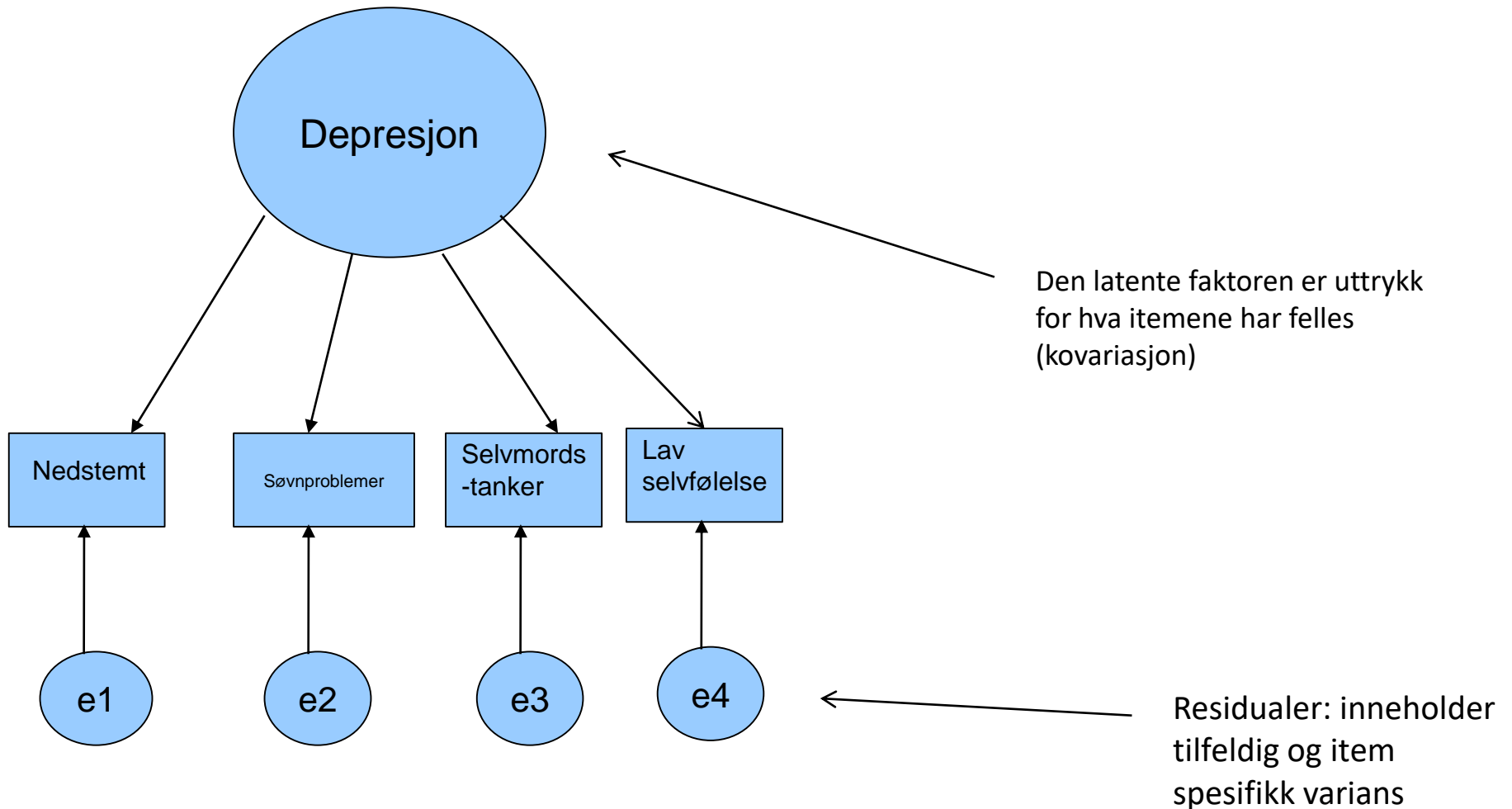


Dette er viktig mht:

Hvilke metoder en skal bruke for å undersøke måleegenskapene til instrumentet

I andre statistiske analyser vil bruk av en korrekt målemodell ofte kunne bidra til og mer valide resultater

# Refleksiv indikatormodell



# Refleksiv indikator modell

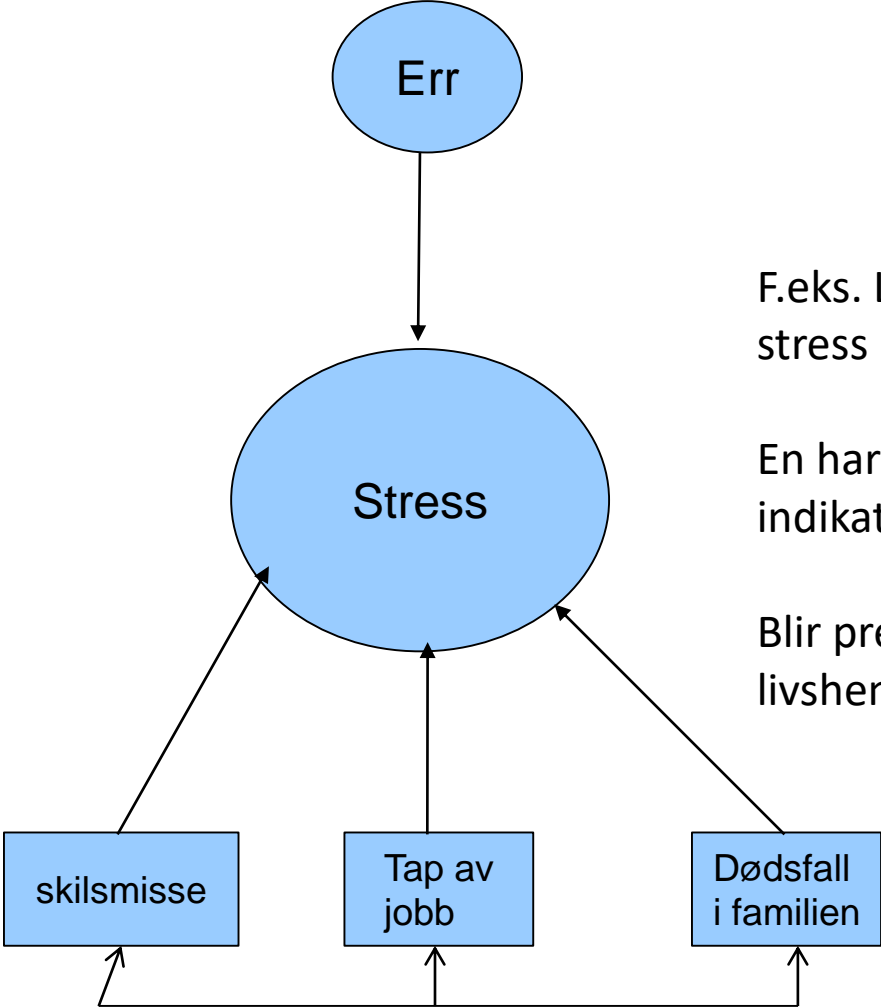
En antar som regel at det underliggende latente konstruktet eksisterer uavhengig av målingen (realisme)

Indikatorer er utbyttbare

Bør ha relativ høy korrelasjon mellom indikatorene

Unidimensjonal (1 faktor) er et ideal

# Kausal (formative) indikatormodell



F.eks. Latent variabel = stress

En har ikke målt reflekseive indikatorer.

Blir predikert av stressende livshendelser

# Kausal indikator modell

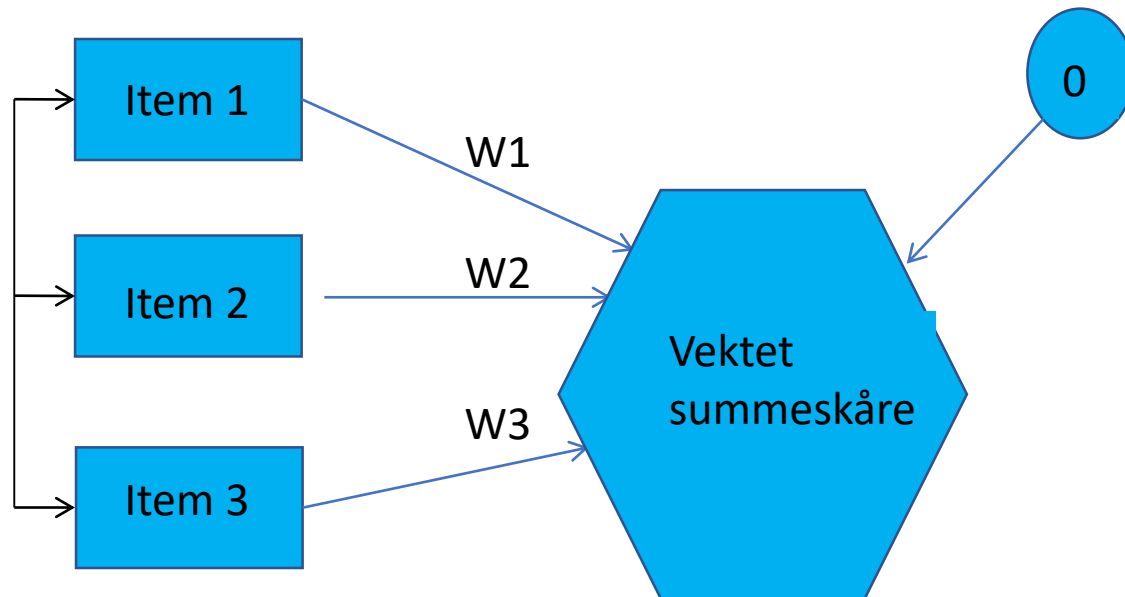
En antar som regel at det underliggende latente konstrukt eksisterer uavhengig av målingen (realisme)

Indikatorer er ikke utbyttbare

Indikatorene trenger ikke å være korrelert men skal en «conceptual unity» ved at de er relatert til definisjonen av konstruktet

Indikatorene kommer som regel fra flere dimensjoner

# Composite (formative) modell



Composite indekser blir definert (formet) av indikatorene og blir som oftest sett som pragmatiske verktøy heller enn et uttrykk for noe som finnes utenfor målingen. Slår ofte sammen mange dimensjoner. F.eks. APGA skåre, SES, Dow Jones indeks

# Composite indikator modell

Indikatorene er selve konstruktet og eksisterer ikke utenfor målingen

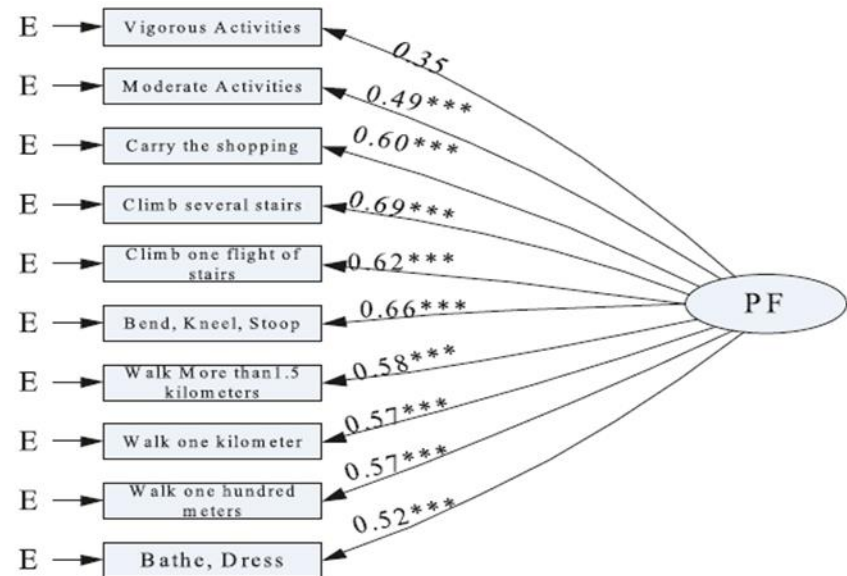
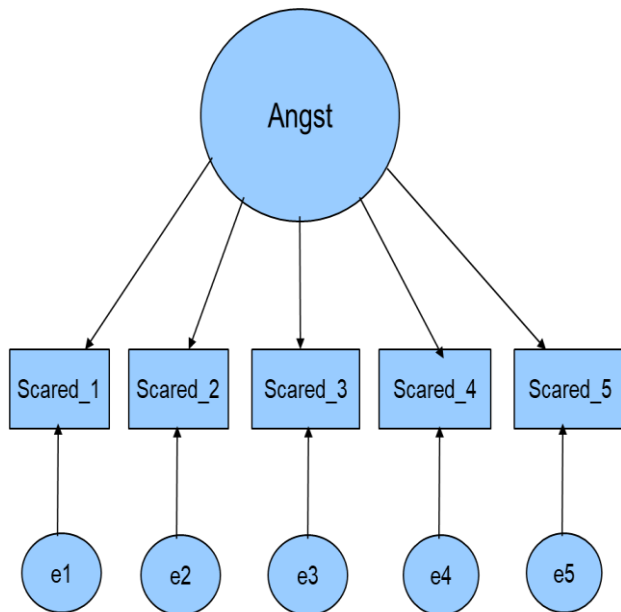
Indikatorer er ikke utbyttbare

Indikatorene trenger ikke å være korrelert og kan være en ren samlev variabel

Indikatorene kommer som regel fra flere dimensjoner

# Målemodeller for pasientrapporterte data

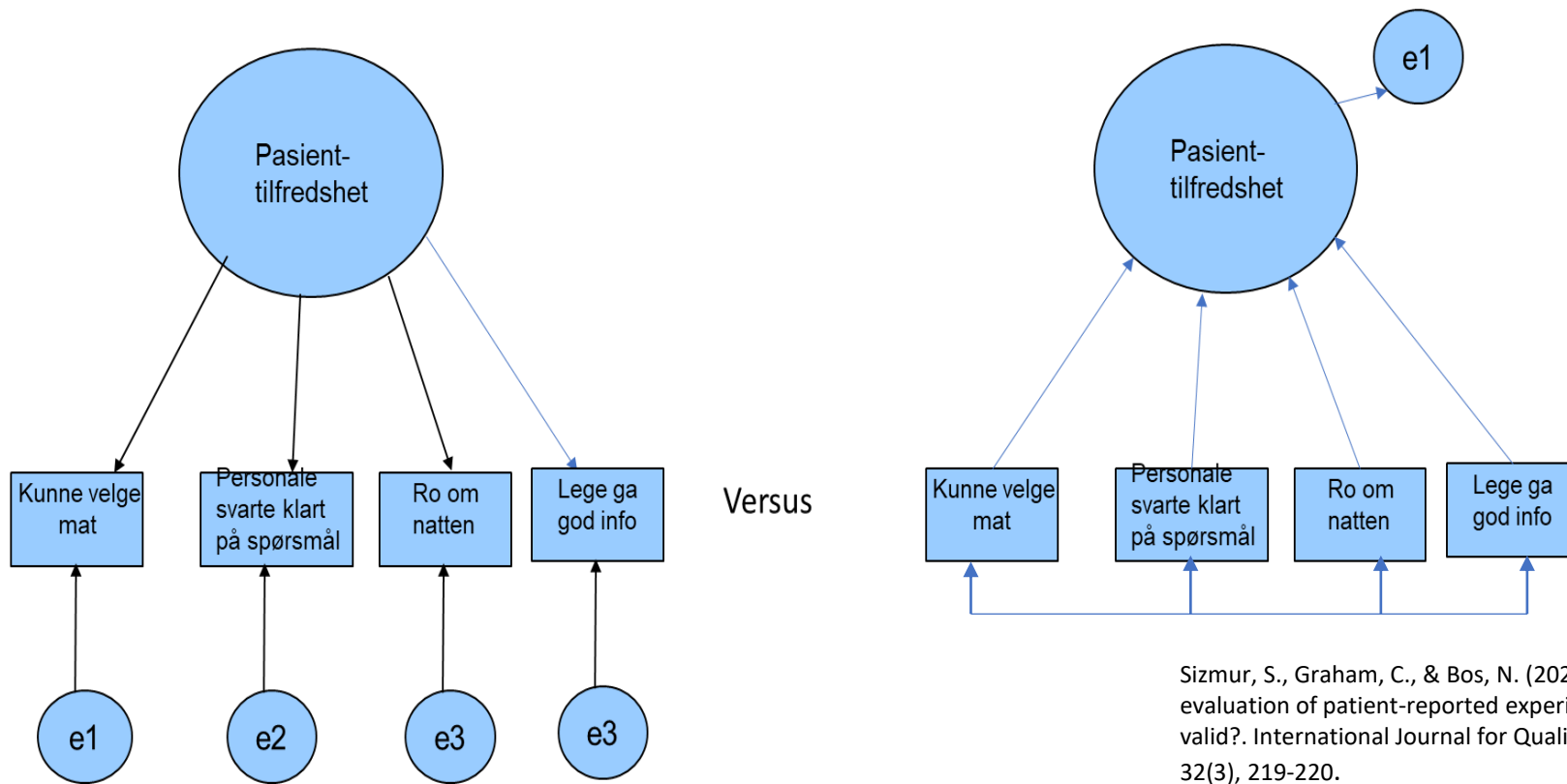
- Mange forskere benytter en refleksiv indikator modell som en ryggmargsrefleks ovenfor alle skalaer
- Når en måler relativt spesifikke domener som f.eks. angst, fatigue, smerte er dette ofte ok men hva med fysisk funksjon (PF)?





# Målemodeller for pasientrapporterte data

- PREM (Patient Reported Experience Measures) mål på pasienterfaringar
  - Pasientens opplevelse av og tilfredshet med helsetjenesten.
- En refleksiv indikatormodell blir regel brukt også her, selv om en kausal indikatormodell ofte kan gi mer mening (Sizmur et. al., 2020)



Sizmur, S., Graham, C., & Bos, N. (2020). Psychometric evaluation of patient-reported experience measures: is it valid?. *International Journal for Quality in Health Care*, 32(3), 219-220.

# Målemodeller for pasientrapporterte data: Helserelatert livskvalitet

- Hva er livskvalitet?
- Et relativt diffust multidimensjonalt begrep hvor det finnes en rekke definisjoner.
  - An evaluation of all aspects of our lives, including, for example, where we live, how we live, and how we play. It encompasses such life factors as family circumstances, finances, housing and job satisfaction (Johnston, 2013)
- Hva er helserelatert livskvalitet?
- Også her finnes det mange definisjoner. Begrep som helserelatert livskvalitet, helsestatus, velvære, livskvalitet og helse brukes ofte om hverandre.
  - Begrepet brukes ofte for å måle hvordan en sykdom eller behandling påvirker en persons livskvalitet.
  - Brukes også i andre sammenhenger som kost/nytte av behandlingseffekt (Bruvik & Mjørud, 2012)

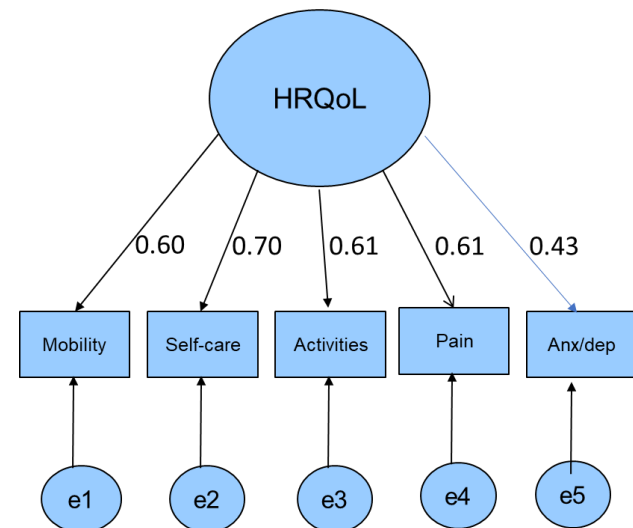
# Målemodeller for pasientrapporterte data: helse relatert livskvalitet

- I forskningslitteraturen blir helse relatert livskvalitet ofte brukt men ikke definert
- Helse relatert livskvalitet blir brukt på:
  - beskrivelsen av et batteri med selvstendige må ulike domener (f.eks. angst, depresjon, fysisk funksjon, fatigue etc)
  - en summeskåre av alle er noen sentrale domener i et generisk eller sykdomsspesifikt skjema
  - Et globalt spørsmål om helse relatert livskvalitet
  - En vektet utility skåre for økonomiske analyser (f.eks. EQ5D)

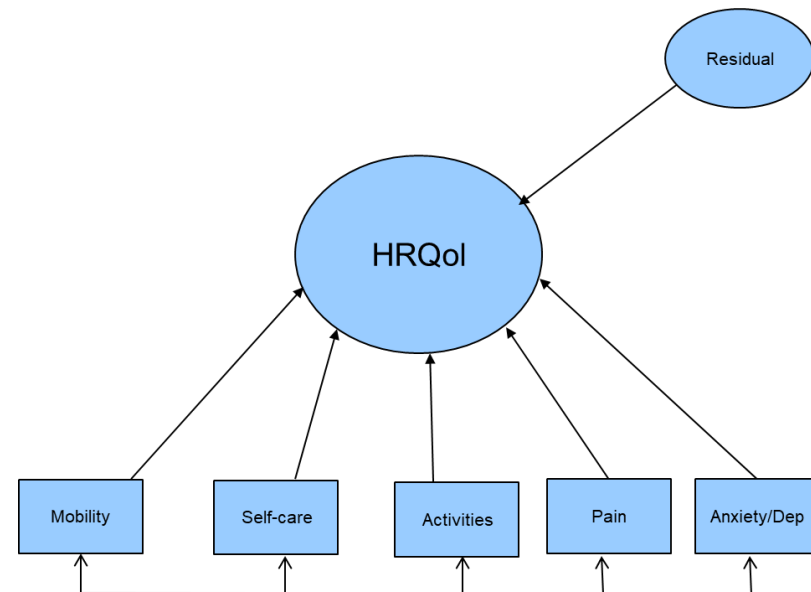
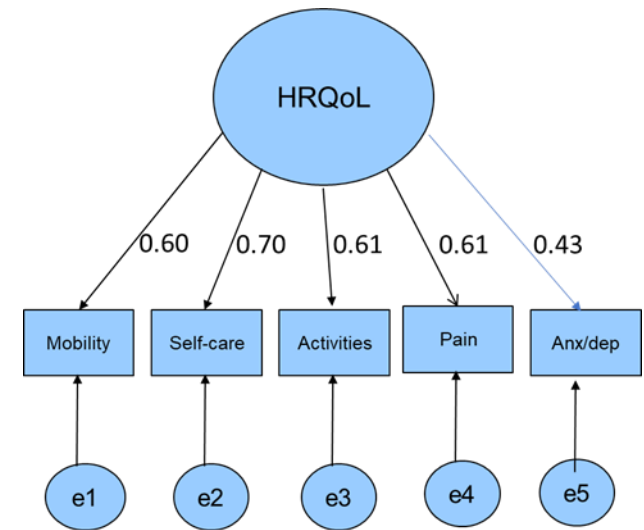
# Modeller for helselerelatert livskvalitet

- Mye av forskningen på helselerelatert livskvalitet har brukt en reflektiv modell

EQ-5D



- Potensielle svakheter ved en refleksiv for HRQOL
- I dette tilfellet er HRQOL representert ved fellesvariansen til gange (mobilitet), personlig stell, vanlige gjøremål, smerter/ubehag og angst/depresjon. Hva betyr dette?
  - Er de refleksive indikatorene utbyttbare?
  - Hva betyr det at ladningen til angst/dep er lavest?
  - Er det grunn til å tro at HRQol er en unidimensjonal?
- Enkelte forskere har argumentert for at en kausal indikatormodeller ofte er mest fornuftig når det gjelder analysen av indikatorene brukt i helserelatert livskvalitetsmål (Bollen et. al., 2009; Boehmer & Luszczynska., 2006; Fayers et. al., 1997)
- Andre igjen har argumentert for at HRQol skjema inneholder både refleksive og kausale indikatorer



# Helserelatert livskvalitet

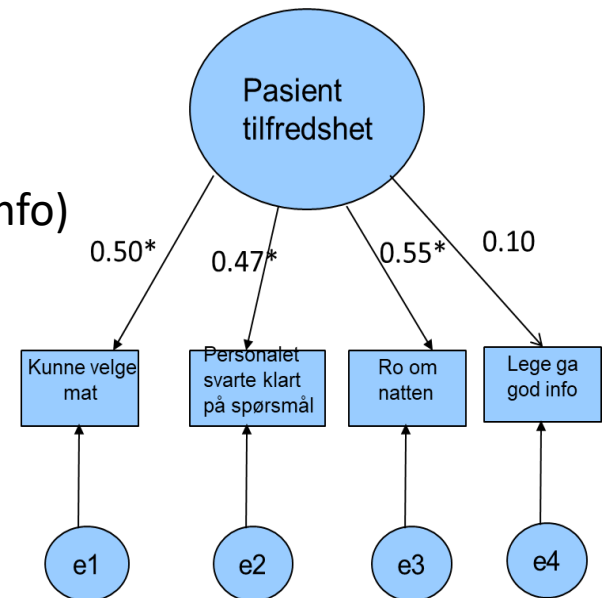
- Costa (2015) argumenterer for at ulike definisjoner av helserelatert livskvalitet fordrer bruk av ulike målemodeller og at forskere bør bruke målemodeller som ligger implisitt i definisjonen de bruker

# Konsekvenser for testing av målegenskaper (for reflekseive og formative (kausal og composite) indikatorer

- Reliabilitet
  - Indre konsistens (f.eks. Chronbach alpha) er kun relevant for modeller som inneholder reflekseive indikatorer. I en kausal indikatormodell er det ønskelig at indikatorene ikke skal være for høyt korrelerte
  - Testing av test-retest er ok for begge typer modeller
- Validitet
  - Testing av unidimensjonalitet er bare relevant for reflekseive modeller
  - I en formativ modell bør indikatorene være korrelerte med et mer generelt mål på konstruktet (f.eks. generell livskvalitet)
  - Innholdsvaliditet: At testen måler hele bredden av begrepet er viktigst for formative modeller
- Sensitivitet
  - Viktig for begge typer modeller
- Responsivitet
  - Viktig for begge typer modeller

# Noen uheldige konsekvenser ved å teste måleegenskapene til en formativ indikator modell basert på kriterier beregnet på en refleksiv modell

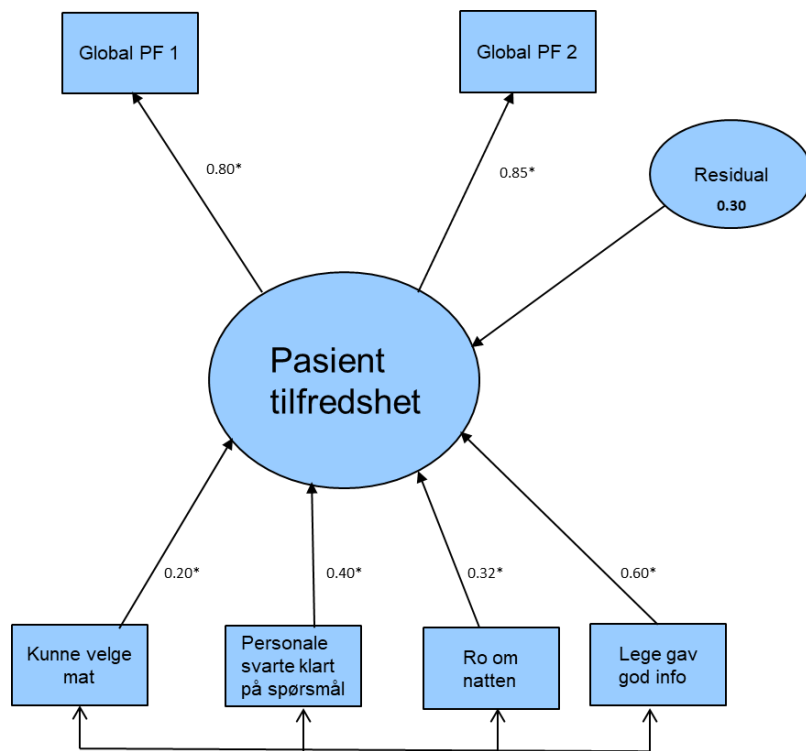
- Klassisk test-teori, eksplorerende faktoranalyse og Item Response Theory blir ofte brukt for å teste måleegenskapene til et skjema og alle bygger på en refleksiv indikator modell
- En skala kan bli forkastet/ikke publisert pga lav intern konsistens reliabilitet (f.eks. Chronbach alpha)
- Potensielt dårlig tilpasning til en 1 faktorløsning
- Kan fjerne potensielt viktige ledd (f.eks. Lege gav god info) pga lav faktorladning





# Å teste måleegenskapene til en kausalindikator modell

- Man må bruke andre kriterier for å undersøke måleegenskapene til en skala basert på en kausal indikatormodell (Diamantopoulos & Winklhofer, 2001)

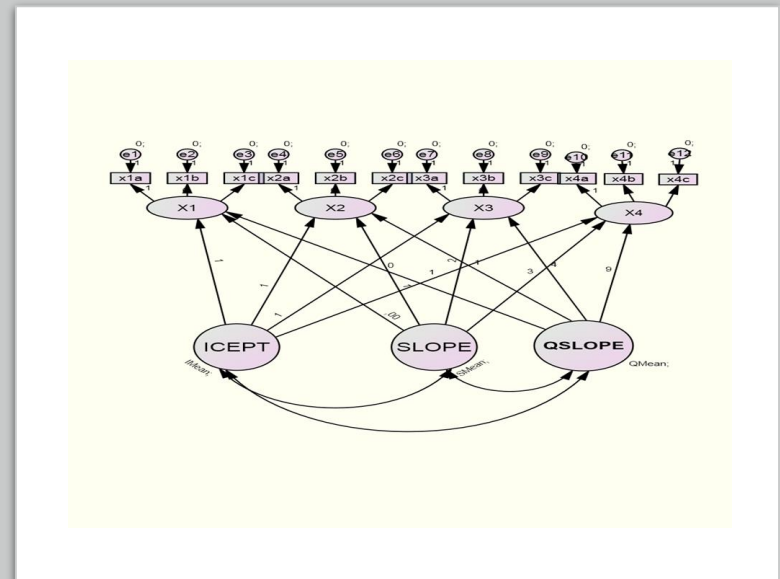
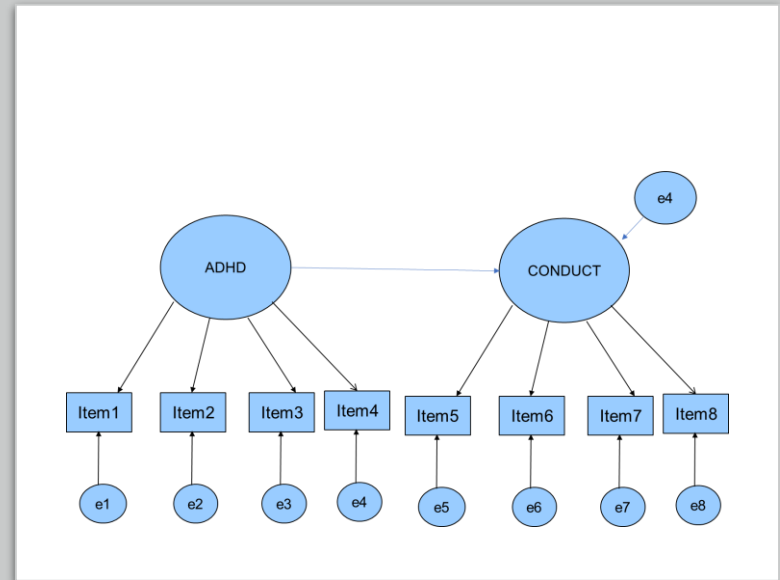


30% av pasient tilfredshet er ikke forklart av indikatorene

Alle indikatorene predikerer pasient tilfredshet. God info fra lege var den viktigste, å kunne velge mat var minst viktigst

# Konsekvenser for statistiske analyser:

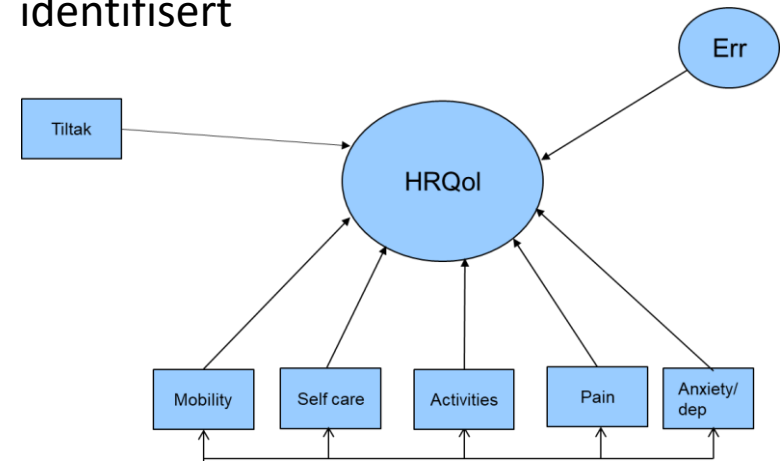
- Bruk av variabler som inneholder mye feilvariasjon kan føre til både bias og mindre statistisk styrke i f.eks. korrelasjonsanalyser, multipl regresjonsanalyser, longitudinelle og stimodeller (f.eks. Cole & Preacher, 2014)
- Om målet inneholder reflekseve indikatorer kan disse problemene minimeres ved å:
  - benytte SEM analyser hvor en representerer variablene som latente variabler og dermed fjerne residualvariasjonen fra analysene
  - Påse at en bruker observertede variabler med så høy indre konsistens (f.eks. Chronbach's alpha) som mulig. En kan da justere (f.eks. korrelasjonene for dette)
- Merk at begge disse bygger på en reflekseve indikatormodell. Dersom en bruker en reflekseve indikatormodell på formative variabler for å fjerne residual variasjon skaper en bias (Rhemtulla et. al., 2020)



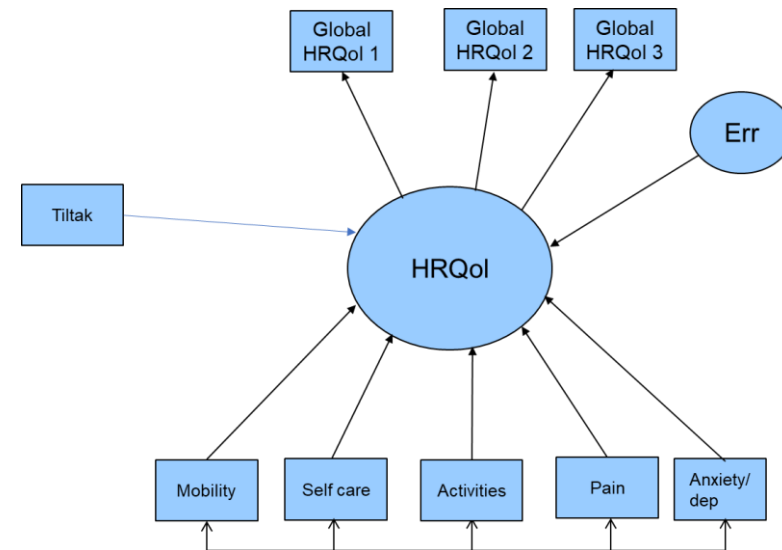
# Bruk av formative variabler i statistiske analyser

- Om en benytter en kausal indikator SEM modell må denne ha med to eller helst flere refleksive indikatorer for at modellen skal bli identifisert
- HRQol blir da i realiteten en reflesiv indikator modell målt med Global HRQol 1, HRQol 2 & HRQol 3
- Spørsmål:
  - a) Hvorfor trenger en da de spesifikke indikatorene i tillegg til de globale målene?
  - Hva gjør en om en ikke har de globale målene på HRQol?
- Mulig løsninger:
  - Bruke en vektet/uvektet summeskåre av de kausale indikatorene som proxy for HRQol
  - Behandle domenene som separate prediktorer/utfallsvariabler istedenfor kausale indikatorer

Ikke matematisk identifisert



Ok



# Composite model

Om en bruker en composite model er det utviklet egne avanserte analyser for dette

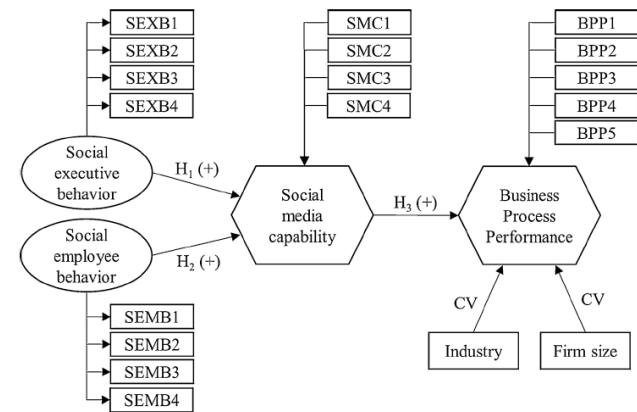
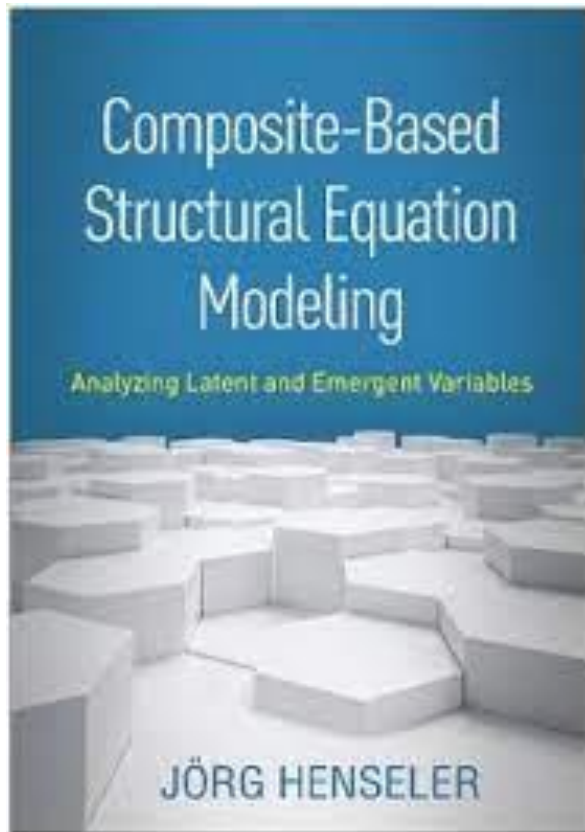


Fig. 1. Research model (CV = Control variables).

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, 57(2), 103168.

# Målemodellens innvirkning på skåring av indikatorer i en summeskåre

- I de fleste analyser (f.eks. multippel regresjon, ANOVA/ANCOVA, t-test, flernivåanalyser) bruker vi summeskåre som proxy for konstrukt
- Hvilken metode vi bruker til å summere spørreleddene til en summeskåre er viktig og bør være basert på hvilken målemodell en bruker (Avila et. al., 2015)
- Om en benytter en refleksiv indikator modell er det som regel ok å bruke en uvektet summeskåre om den er unidimensjonal.
  - Residualene ansett som tilfeldige hvor summen er lik 0.
  - Vekting basert på f.eks. størrelsen på faktorladningene har vist seg å ha lite praktisk nytteverdi – spesielt om det er relativt høy korrelasjon mellom indikatorene

# Målemodellens innvirkning på skåring av indikatorer i en summeskåre

- Formative indikator modeller

- Kausal indikator modell: Siden indikatorene er multidimensjonal og ofte ikke høyt korrelerte kan det ofte være hensiktsmessig å vekte indikatorene (f.eks. stressende livshendelser) mht. hvilken sammenheng de har med et globalt mål på det en ønsker å måle (f.eks. opplevd stress)
- Composite indikator modell. I og med at dette ofte har til hensikt å måle et kunstig konstrukt som har pragmatisk nytteverdi står forskerne relativt fritt til å velge vekting
- En kan f.eks. velge å vekte indikatorer ulikt i apgarskåre basert på hvor godt de predikerer barns helse.

# Målemodellens innvirkning på skåring av indikatorer i en summeskåre

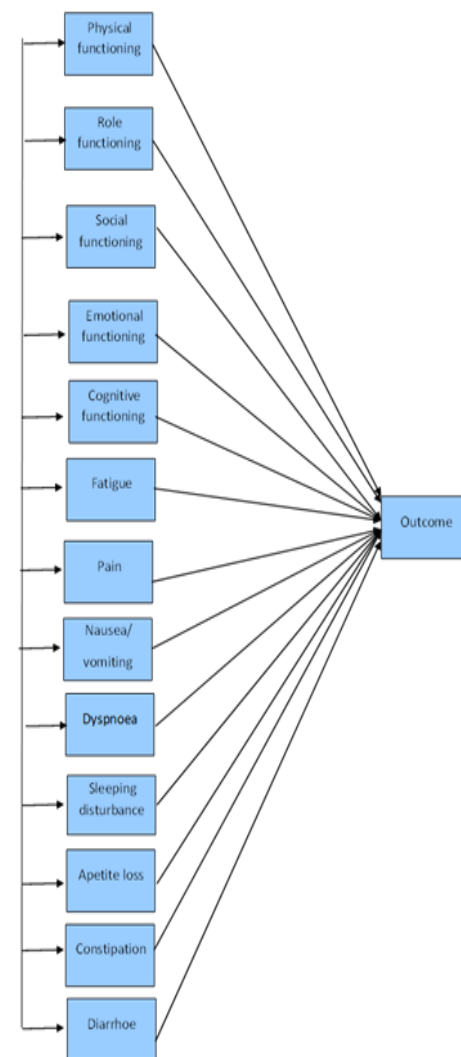
- Utility skåre
  - Preference/utility skårer som ofte brukes som grunnlag i helseøkonomiske evalueringer og prioritering av ressurser i helsetjenesten.
  - Vektet på bakgrunn av preferanser for hypotetiske helsetilstander i populasjonen
- F.eks. EQ5D har fem spørsmål som dekker gange(mobilitet), personlig stell, vanlige gjøremål, smerte/ubehag og angst/depresjon
  - Et utvalg har da blitt å vurdere ulike helsetilstander f.eks. 22221, 31321 som danner grunnlaget for indeksskåren

# Noen potensielle problemer med summeskårer av komplekse multidimensjonale konstrukt

- Flere forskere har pekt på at bruken av komplekse multidimensjonale konstrukt kan være uheldig
- Kan være vanskelig å tolke skåren ettersom individer med samme skåre på en skala kan ha helt ulike profiler på de ulike dimensjonene som inngår
- Uvist om det er noen av dimensjonene som predikerer eller blir påvirket mer eller mindre enn de andre dimensjonene
- Prediksjonsevnen (forklart varians) er større om en inkluderer de ulike dimensjonene som separate variabler i en regresjonsanalyse enn om en slår dem sammen i summeskårer



versus





# Konklusjon

- Pasientrapporterte data er viktig og nyttig
- En bør helst ha med både generiske og sykdomsspesifikke skjema i registrene ettersom de begge har sine styrker og svakheter
- Viktig å bruke spørreskjema som passer til formålet og har gode måleegenskaper
- Viktig å bruke korrekt målemodell både når en vurderer måleegenskapene til et spørreskjema og når en benytter det i analyser
- Stort behov for større enighet om konseptuelle modeller på sentrale konstrukt som f.eks. helse relatert livskvalitet – hvordan de skal måles og analyseres
- Bruken av komplekse multidimensjonelle konstrukt må vurderes i forhold hva en vinner/taper i forhold til å bruke mer spesifikke konstrukt