

Kriterier ved validering av en metode for kostholdsundersøkelser – Når er validiteten til en metode tilfredsstillende?

Lene Frost Andersen

Institutt for ernæringsforskning, postboks 1046 Blindern, 0316 Oslo

Telefon: 22851370 telefaks: 22831531 e-post: l.f.andersen@basalmed.uio.no

SAMMENDRAG

Det finnes ingen metode for kostholdsundersøkelser som kan måle kostholdet til et individ eller en gruppe individer uten at det er feil knyttet til dataene. Dette får konsekvenser for tolkningen av resultatene fra kostholdsundersøkelser og for tolkningen av sammenhengen mellom matinntak og helse. For at mangelfulle kostholdsdata ikke skal resultere i feil konklusjoner er det viktig å undersøke validiteten til metoder for kostholdsundersøkelser, samt å vurdere i hvilke sammenhenger en kostholdsmetode gir tilstrekkelig gode data til at de kan brukes. Ved vurdering av validiteten til metoder for kostholdsundersøkelser er det tre trinn som inngår; vurdering av design og gjennomføring av en valideringsstudie, vurdering av analyseringen av kostdataene fra valideringsstudien, samt tolkning av resultatene fra valideringsstudien. I artikkelen diskuteres hvordan man kan vurdere validiteten til metoder for kostholdsundersøkelser ved å gå igjennom de tre trinn som er nevnt ovenfor. Dessuten illustreres ved hjelp av tre eksempler på forskjellige epidemiologiske studier, når validiteten til en kostholdsmetode er tilfredsstillende i en studie, samt hvilke konsekvenser validiteten kan ha for konklusjonen i en studie.

Frost Andersen L. **Criteria for the validation of dietary assessment methods. When is the validity satisfactory?** *Nor J Epidemiol* 2000; 10 (1): 17-24.

ENGLISH SUMMARY

Today there is no method available to measure dietary exposure among individuals or groups of individuals without error. The errors will affect the interpretation of the results from dietary assessment studies and have an impact on epidemiological measures of associations between intake and disease. To reduce these problems it is important to examine the validity of dietary assessment methods, and to examine in what kind of studies data from a specific dietary assessment method are valid enough to be used. Evaluation of the validity of dietary assessment methods include three levels; (1) evaluation of the design and implementation of the validation study, (2) evaluation of the data analysis from the validation study, and (3) the interpretation of the results from the validation study. This article discusses how to evaluate the validity of dietary assessment methods using the three levels mentioned above. Furthermore, three examples of epidemiological studies illustrate when the validity of a dietary assessment method is satisfactory in a study, and what kind of consequences the validity may have for the conclusion of a study.

INNLEDNING

Per i dag finnes det ingen metode for kostholdsundersøkelser som kan måle kostholdet til et individ eller en gruppe individer uten at det er feil knyttet til dataene. Dette får konsekvenser for tolkningen av resultatene fra kostholdsundersøkelser og for tolkningen av sammenhengen mellom matinntak og helse. For at mangelfulle kostholdsdata ikke skal resultere i feil konklusjoner er det viktig å undersøke kvaliteten til metodene som blir brukt til å måle matinntak – hvor godt måler

metoden det den skal? Det er også viktig å vurdere i hvilke sammenhenger en metode for kostholdsundersøkelse gir tilstrekkelig gode data til at den kan brukes.

Kvaliteten til en metode for kostholdsundersøkelse kan uttrykkes ved hjelp av dens validitet og reproduserbarhet. *Reproduserbarhet* er et uttrykk for om metoden kan gi det samme resultatet ved gjentatte målinger blant de samme personene og under samme forhold (Klaver et al., 1988). *Validitet* er uttrykk for i hvilken grad en metode kan måle det metoden er ment å skulle måle (Nelson, 1997).

Ved vurdering av validiteten til metoder for kostholdsundersøkelser er det tre trinn som inngår; vurdering av design og gjennomføring av en valideringsstudie, vurdering av analyseringen av kostdataene fra valideringsstudien, samt tolkning av resultatene fra valideringsstudien. Resultatene fra valideringsstudien, som danner bakgrunn for å vurdere om validiteten til en metode er tilfredsstillende, er et produkt av hvordan valideringsstudien er designet og gjennomført, samt hvorledes dataene fra studien er analysert. En utilstrekkelig valideringsstudie og analysering av dataene kan medføre en feil tolkning av validiteten til en metode. Ved vurdering av når validiteten er tilfredsstillende, er det en forutsetning at valideringsstudien er gjennomført tilfredsstillende og at analyseringen av dataene er gjennomført tilfredsstillende.

I denne artikkelen diskuteres når validiteten til metoder for kostholdsundersøkelser er tilfredsstillende.

HVORDAN VALIDERE EN METODE FOR KOSTHOLDSUNDERSØKELSE TILFREDSSTILLEND?

Valideringsstudier gjennomføres for å gi kunnskap om hvor god metoden er til å måle det sanne inntak samt for å gi kunnskap om hvilken type feil som er knyttet til metoden. Validiteten kan undersøkes ved å sammenligne resultatene fra den metode som skal testes (testmetode) med resultatene fra en referansemetode. Den ideelle referansemetode, som måler det sanne inntak (dvs. en gullstandard), eksisterer ikke. I stedet kan resultatene fra testmetoden sammenlignes med resultatene fra en annen kostholdsundersøkelsesmetode som antas å måle matinntaket mer nøyaktig enn testmetoden. Resultatene fra testmetoden kan også sammenlignes med biologiske markører.

Det er flere kriterier som må være oppfylt for at designen og gjennomføringen av valideringsstudien kan vurderes som tilfredsstillende:

a. Referansemetoden må anses å være mer nøyaktig enn testmetoden. Veid registrering av kosten i flere dager er ofte blitt brukt som referansemetode da den blir ansett for å være mer nøyaktig enn andre tradisjonelle kostholdsmetoder (Willett et al., 1985; Block et al., 1990; Hankin et al., 1991). Forskning gjennom de senere år har imidlertid vist at det er en tendens til underestimering av energiinntak med denne metoden, og at graden av underestimering varierer mellom individer (Schoeller, 1990; Black et al., 1991). Med hensyn til biologiske markører finnes det per i dag to valide markører som kan være markør for det kvantitative inntak. Dobbelmerket vann metoden som måler energiforbruk, kan være markør for energiinntak når personene er i energibalanse (Prentice et al., 1990). Dessuten er nitrogen i døgnurin en god markør for inntak av protein (Isaksson, 1980; Bingham & Cummings, 1985). Det finnes andre biologiske markører som blir anvendt i forbindelse med validering av metoder for kostholdsundersøkelser, for eksempel fettsyre

i blod og fettvev. Disse er ikke gode markører for det totale kvantitative inntak, men er mer brukbare som markører for det relative inntak (Katan et al., 1997; Hunter, 1990; Bates et al. 1997).

b. Referansemetoden skal måle inntaket på samme nivå som testmetoden. Skal testmetoden gi et godt mål for det vanlige inntaket eller inntaket over en bestemt periode, må referansemetoden gi samme type data. Skal testmetoden gi inntaket på gruppe eller individnivå må det velges en referansemetode som gir et mål på samme nivå.

c. Feilene knyttet til referansemetoden skal være uavhengig av feilene knyttet til testmetoden. For eksempel har et frekvensspørreskjema om kosthold ofte lukkede svar alternativer og det avhenger av hukommelse. Ved validering av et frekvensskjema vil veid registrering være en referansemetode med delvis uavhengige feil ettersom metoden er åpen, uavhengig av hukommelse og måler porsjonsstørrelser direkte. De to metodene har imidlertid også felles feil, for eksempel kan tendensen til å feilrapportere inntak være den samme med begge metoder, og næringsberegningen av dataene fra begge metoder er basert på samme matvaretabell. En av fordelene med biologiske markører som referanse, er at feilene knyttet til disse ikke er relatert til feilene knyttet til en kostholdsmetode.

d. Ideelt å velge to referansemetoder. Valideringsstudien styrkes når dataene fra testmetoden sammenlignes med både data fra en annen metode for kostholdsundersøkelse og en biologisk markør. Det er ofte et ressursspørsmål om to referansemetoder kan inkluderes i en valideringsstudie.

e. Rekkefølgen av referansemetoden og testmetoden er viktig, referansemetoden bør ikke påvirke testmetoden.

f. Deltagerne i valideringsstudien bør være et underutvalg av populasjonen der testmetoden skal brukes. Det er vanskelig å få representative underutvalg da deltagelse i en valideringsstudie krever en høy grad av motivasjon og aktivitet fra deltagerne. Derfor blir de fleste valideringsstudier gjennomført i utvalg som er mer motivert enn den populasjonen hvor metoden skal brukes, med de tolkningsfeil dette kan medføre. Hvor mange deltagere som bør være i en valideringsstudie avhenger av formålet med valideringsstudien (Carroll et al., 1997; Willett, 1998b). På bakgrunn av forskjellige beregninger kommer Willett fram til at et utvalg på 100-200 personer vil være tilstrekkelig stort for å kunne si noe om validiteten, og være tilstrekkelig stort til at man kan miste noen deltagere i løpet av studien (Willett 1998b).

HVORDAN KAN VI ANALYSERE DATAENE FRA VALIDERINGSSTUDIEN TILFREDSSTILLEND?

Analyseringen av dataene fra valideringsstudien skal danne bakgrunn for vurdering av validiteten til testmetoden. Før dataene analyseres må de kvalitetskon-

trolleres. Dataene bør sjekkes for mangelfulle data, for alle typer kodefeil og for ekstreme verdier ("outliers").

Dessuten må kvaliteten til dataene fra referansemetoden vurderes. Dette er viktig da vurderingen av validiteten til testmetoden tar utgangspunkt i at referansemetoden måler tilnærmet det sanne inntak. Hvis referansemetoden er en annen kostholdsmetode finnes det en enkel metode for å vurdere om metoden har en tendens til å over- eller underestimere energiinntaket. Denne metoden, som ble publisert av Goldberg og medarbeidere, krever kjennskap til vekt og høyde (Goldberg et al., 1991). Er det en stor grad av over- eller underestimering kan det være et problem å anvende dataene fra referansemetoden som et bilde på det sanne inntak. Hvis problemet med over- eller underestimering kun gjelder få personer, bør det overveies om disse skal ekskluderes fra den videre beregningen. Er referansemetoden en biologisk markør, er det viktig å kontrollere at prøvematerialet for analyse er behandlet korrekt, samt at presisjonen og nøyaktigheten til analyseteknikken er best mulig (Bates et al., 1997; Hunter, 1998).

Før den statistiske analyseringen av kostdataene er det vanlig å undersøke fordelingen til dataene, om de er normalfordelte eller ikke. I mange tilfelle er kostdata ikke normalfordelte, og det må vurderes om man vil analysere på dataene som de er, om dataene skal transformeres før analyse så de blir tilnærmet normalfordelte, eller om man vil analysere på begge typer data. Hvilken statistikk (parametrisk eller ikke-parametrisk statistikk) som er best egnet til å analysere dataene fra valideringsstudien må vurderes i den enkelte studie med utgangspunkt i fordelingen av dataene og antall deltagere.

En gjennomgang av publiserte valideringsstudier viser at det i en stor del av artiklene er brukt de samme metodene til å vurdere kvaliteten til testmetoden. I det følgende diskuteres kort hvordan forskjellen mellom data fra testmetoden og referansemetode ofte blir presentert (tabell 1). Det er ikke noen fullstendig oversikt over metoder til å presentere forskjellen mellom testmetoden og referansemetoden, det finnes andre relevante metoder (Nelson 1997; Willett 1998a).

Tabell 1. Oversikt over analyser til å uttrykke forskjellen mellom test- og referansemetoden.

Datanivå ¹	Analyser
Gruppenivå	Sammenligne gjennomsnitt og spredningsmål mellom test- og referansemetoden
Individnivå	<ul style="list-style-type: none"> ▪ Altman-Bland plot (plott av differansen mellom test- og referansemetode mot gjennomsnitt av de to metodene) ▪ Andel som har mindre enn en definert størrelse for eksempel 10% i differanse mellom test- og referansemetode
Rangering	<ul style="list-style-type: none"> ▪ Korrelasjoner; Pearson korrelasjon (parametrisk) Spearman's rangkorrelasjon (ikke-param.) Intra-class korrelasjon Kappa ▪ Klassifisering i tertiler, kvartiler eller kvintiler

¹Det datanivå som man ønsker å validere testmetoden i forhold til.

For å vurdere om testmetoden kan gi et godt mål på matinntaket til en gruppe, kan man sammenligne gjennomsnitt (eller median) og spredningsmål fra henholdsvis test- og referansemetoden. "Altman-Bland plot" (plot av differansen mellom test- og referansemetode mot gjennomsnitt av dataene fra de to metodene) kan brukes til å vurdere hvor godt samsvar det er mellom de to metodene for det enkelte individ (Altman & Bland, 1983; Bland & Altman, 1986) (Figur 1). Plottet kan avsløre ekstreme verdier ("outliers"), samt avsløre om det er systematiske feil ved dataene fra testmetoden, for eksempel økende differanse mellom de to kostholdsmetodene med økende inntak. Dessuten kan plottet vise om forskjellen mellom test- og referansemetoden er systematisk eller tilfeldig mellom individer. Differansen mellom test- og referansemetode for de enkelte individer kan også brukes til å vurdere hvor godt samsvar det er mellom de to metodene på individnivå. En beregning av andelen av deltagerne i en valideringsstudie som har en differanse større enn 10% eller 20%, kan også være en hjelp ved vurdering av validiteten (tabell 2).

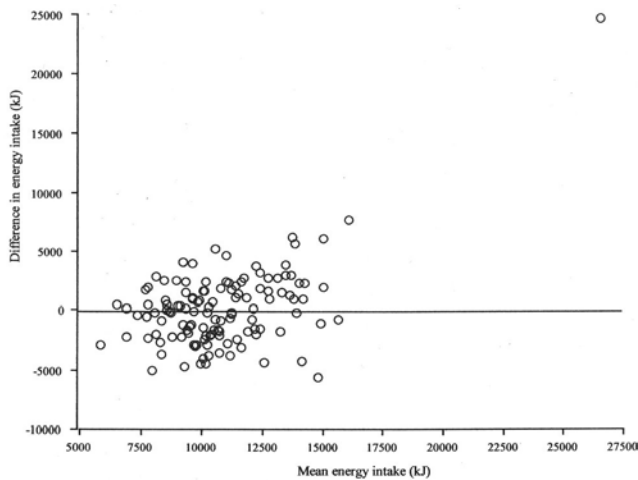
Tabell 2. Det daglige inntak av energigivende næringsstoffer estimert fra frekvensskjemaet (FS) og 7-dagers veid registrering (VR) (n=49) (Etter Frost Andersen et al., 1995).

	FS Median (p ₂₅ , p ₇₅) ¹	VR Median (p ₂₅ , p ₇₅) ¹	Antall personer ² diff. < 20%
Energi, MJ	10.7 (9.0, 14.1)	8.6 (7.2, 10.8) ^{***}	13
Protein, g	86 (76, 111)	68 (56, 84) ^{***}	15
Total fett, g	79 (64, 115)	67 (57, 88) ^{***}	19
Total karbohydrat, g	355 (291, 428)	284 (235, 332) ^{***}	13
Sukker, g	75 (47, 114)	73 (57, 105)	20

* p ≤ 0.05 ** p ≤ 0.01 *** p ≤ 0.001

¹ p₂₅ = 25th persentil, p₇₅ = 75th persentil.

²Antall personer med en differanse mellom inntak målt med FS og VR på mindre enn 20% av inntaket målt med FS.



Figur 1. Differansen mellom energiinntaket målt med hhv. et frekvensspørreskjema og 14-dagers veid registrering plottet mot gjennomsnitt av energiinntaket målt med de to metodene (Etter Andersen et al., 1999).

I epidemiologiske studier er det ofte det relative inntaket til individer som er av interesse (Kohlmeier & Bellach, 1995). Det er derfor viktig å vite hvor god kostholdsmetoden er til å rangere individer i forhold til matinntak. Korrelasjonen mellom inntak målt med henholdsvis test- og referansemetoden, er brukt av mange som basis for å vurdere metodens evne til å rangere individer (Nelson, 1997). Spearman rang korrelasjon og Pearson korrelasjon er mest brukt, men også intra-class korrelasjon og Kappa er inkludert i enkelte studier (Willett, 1998a). Det har gjennom årene vært en del kritikk ved bruken av korrelasjoner i denne sammenheng, da det kan være komplisert å tolke hva størrelsen på korrelasjonskoeffisienten betyr. For eksempel er størrelsen på korrelasjonen avhengig av spredningen i matinntak (mellom-person variasjonen) (Bland & Altman, 1986; Delcourt et al., 1994). Det betyr at i en populasjon med stor spredning i inntak er det lettere å få høye korrelasjoner enn i en med liten spredning. Signifikante korrelasjoner kan heller ikke direkte tolkes som at metoden er god til å rangere individer (Delcourt et al., 1994). En lav korrelasjon på for eksempel 0,20 kan være signifikant når det er et stort antall deltagere. Dessuten overestimeres korrelasjonen hvis det er korrelerte feil mellom test- og referansemetoden, og korrelasjonen svekkes når referansemetoden ikke måler det sanne inntak (Nelson, 1997; Kohlmeier et al., 1997).

Som nevnt tidligere finnes det ingen referansemetode som måler det sanne inntak, men det er utviklet flere modeller til å beregne korrelasjonen mellom inntaket målt med testmetoden og det sanne inntak. Disse beregningene justerer for feil som er knyttet til referansemetoden. I appendiks 1 og 2 er det vist to modeller for beregning av justerte korrelasjoner.

I tillegg til korrelasjoner kan klassifisering av individer i tertiler, kvartiler eller kvintiler i forhold til

inntak målt med test- og referansemetode brukes til å illustrere hvor god kostholdsmetoden er til å rangere individer i forhold til inntak (Nelson, 1997). Ideelt sett bør alle deltagerne klassifiseres i samme tertil, kvartil eller kvintil med begge metodene (100% korrekt klassifisert) og 0% være feilklassifisert. Ved vurdering av testmetodens evne til å rangere individer ut i fra andel som klassifiseres korrekt eller feilklassifiseres, er det viktig å være klar over at bare med sjanse vil en viss andel av deltagerne være korrekt klassifisert og en viss andel være feilklassifisert. Ved bruk av for eksempel kvintiler vil det ved sjanse være 20% som blir korrekt klassifisert og 8% vil være grovt feilklassifisert (dvs. klassifisert i motsatte kvartiler med de to metodene).

TOLKNING AV RESULTATENE FRA VALIDERINGSSTUDIER – NÅR ER VALIDITETEN TILFREDSSTILLENDEN?

Den største utfordringen i forbindelse med valideringsstudier er ofte tolkningen av resultatene fra dataanalysene. Ved tolkningen er det viktig å huske at kostdataene fra testmetoden sammenlignes med data fra en referansemetode som ikke gir det sanne bilde av inntaket. Dessuten er det viktig ikke å tolke mer ut av resultatene enn det studien var designet for. Når studien er designet for å vurdere det gjennomsnittlige inntaket i en gruppe målt med testmetoden, er det ikke sikkert valideringsstudien gir grunnlag for å vurdere om testmetoden kan rangere individer i forhold til matinntak.

Generelt sett er validiteten til en metode for kostholdsundersøkelser tilfredsstillende når metoden måler matinntaket så nøyaktig at det er styrke nok til å finne/oppdage de svar/sammenhenger som eksisterer uten å komme med feil konklusjoner. I det følgende vil tre eksempler illustrere når validiteten til en kostholdsmetode er tilfredsstillende.

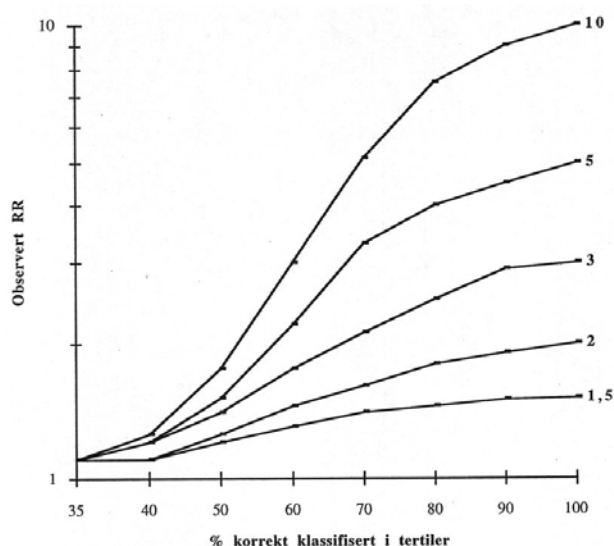
I tre studier med forskjellige formål skal man vurdere om validiteten til et kvantitativt frekvensspørreskjema om kosthold er tilfredsstillende for bruk i studiene. Vi tar utgangspunkt i at frekvensspørreskjemaet er blitt validert mot 14-dagers veid registrering og energiforbruk målt med dobbeltmerket vann metoden i en gruppe på 100 deltagere. Valideringsstudien er designet og gjennomført etter alle kriteriene som er nevnt tidligere i artikkelen, og studien blir vurdert som tilfredsstillende. Energiinntaket fra den veide registreringen (referansemetoden) er sammenlignet med energiforbruk målt med dobbeltmerket vann metoden. I mange valideringsstudier vil det faktisk målte energiforbruk ikke være tilgjengelig for evaluering av energiinntak, da anvendes den tidligere nevnte metoden til Goldberg og medarbeidere (1991) til evaluering av energiinntak. Energiinntaket målt med veid registrering blir i dette tilfelle vurdert til å være tilfredsstillende, og det er ikke noen utpreget tendens til over- eller underestimering blant deltagerne.

Eksempel 1: I en tverrsnittsstudie skal frekvensspørreskjemaet gi tilfredsstillende estimater for det

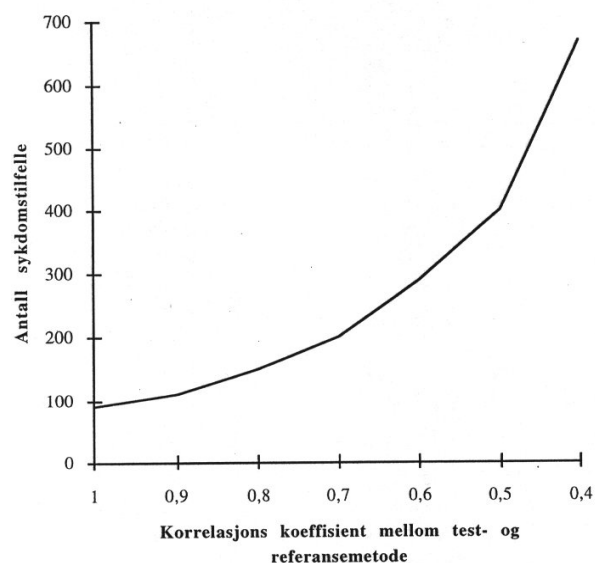
gjennomsnittlige inntak av energi og andelen av energi fra fett. For å vurdere om dette er tilfelle sammenlignes det gjennomsnittlige energiinntaket fra frekvensskjemaet med det gjennomsnittlige energiforbruket målt med dobbeltmerket vann metoden samt energiinntak fra den veide registreringen. Dessuten sammenlignes den gjennomsnittlige andel av energi fra fett målt med frekvensspørreskjema med andelen av energi fra fett målt med den veid registreringen. Validiteten til frekvensskjemaet vil sannsynligvis bli vurdert som tilfredsstillende i forhold til formålet i denne studien hvis ikke det er signifikante forskjeller i de ovennevnte sammenligninger.

Eksempel 2: I en kohortstudie er formålet å undersøke om det er en sammenheng mellom andel av energi fra fett og brystkreft. Data fra valideringsstudien viser at frekvensspørreskjemaet klassifiserer 55% av individene i den korrekte tertilen med hensyn til andel av energien fra fett. Dessuten antas det her at den sanne relative risiko (RR) for brystkreft er 3 mellom høyeste og laveste tertil for andel av energi fra fett. Ettersom frekvensskjemaet er beheftet med målefeil vil den observerte RR mellom høyeste og laveste tertil være mindre enn den sanne RR. Figur 2 viser hvilken betydning misklassifisering har for den observerte RR. I denne studien med en korrekt klassifisering på 55% og en sann relativ risiko på 3, vil det være mulig å observere en RR på rundt 1,5 (figur 2). Hvis denne RR på 1,5 er signifikant ved en test for trend, betyr det at frekvensspørreskjemaets validitet er tilfredsstillende i forhold til formålet i kohortstudien. Hadde den sanne RR vært lavere enn 3, noe som ofte vil være tilfelle for sammenhenger mellom kosthold og sykdom, hadde det vært nødvendig med en bedre klassifisering for å kunne finne en sammenheng. Det å finne en signifikant sammenheng mellom kosthold og sykdom avhenger ikke bare av validiteten til kostholdsmetoden, men også av mellom-person-variasjonen i inntak, og om kostholdet måles i den tidsperiode som er relevant i forhold til sykdomsutviklingen (Margetts og Nelson, 1997).

Det fins flere muligheter til å kompensere eller justere for en utilstrekkelig validitet ved kostholdsmetoden, for eksempel er det flere tilgjengelige metoder for å justere den relative risikoen for målefeil (Spiegelman et al., 1997; Willett, 1998c). Dessuten er det mulig å kompensere for validiteten ved å øke antall deltagere (Kohlmeier og Bellach, 1995). Figur 3 viser antall sykdomstilfeller det er nødvendig å inkludere for å kunne identifisere en signifikant trend når den sanne RR er 3 mellom høyeste og laveste kvintil. Hvis valideringsstudien i vårt eksempel viser en korrelasjon på 0,60 mellom andel av energi fra fett estimert fra henholdsvis frekvensskjema og veid registrering, vil det være nødvendig å inkludere omtrent 300 brystkrefttilfeller i kohortstudien for å kunne observere en sammenheng. Jo lavere korrelasjon, desto flere brystkrefttilfeller må inkluderes.



Figur 2. Kurvene viser den observerte relative risiko (RR) mellom høyeste og laveste tertil ved forskjellige grader av nøyaktighet for en metode og forskjellige verdier for sann RR. Den sanne RR som ligger bak de fem kurvene, varierer fra 1,5 til 10 (Figur etter Margetts og Nelson, 1997).



Figur 3. Antall sykdomstilfeller som er nødvendig for å kunne identifisere en signifikant trend for en sann relativ risiko på 3 mellom høyeste og laveste kvintil i forhold til validiteten (styrke på 80%) (Figur etter Elmståhl et al., 1997).

Eksempel 3: Frekvensspørreskjemaet skal brukes i en tverrsnittsstudie hvor formålet er å undersøke sammenhengen mellom inntak av mettet fett og serumkolesterol. Eksperimentelle studier har vist en sammenheng mellom inntak av mettet fett og serumkolesterol, men man har ikke klart å påvise denne sammenhengen i populasjonsstudier (Marshall et al., 1998). Korrelasjonen som observeres mellom inntak av mettet fett og serumkolesterol vil være svekket av

målefeil ved estimatene fra frekvensskjemaet, feil ved måling av serumkolesterol, og liten spredning i inntaket av mettet fett. Størrelsen på avviket mellom den observerte og den sanne korrelasjonen avhenger bl.a. av hvor godt frekvensskjemaet måler det sanne inntak av mettet fett. Den observerte korrelasjonen kan uttrykkes som $r_o = r \cdot c$ (Cole, 1997), hvor c er en "attenuation" koeffisient som det er mulig å beregne ved hjelp av dataene fra valideringsstudien. Beregningen av denne koeffisienten er vist i Appendix 1. r er den sanne korrelasjonen, dvs. den sammenheng som det er mellom mettet fettinntak og serumkolesterol hvis det ikke var noen målefeil tilstede. I vårt eksempel antas det at "attenuation" koeffisienten for mettet fett er 0,67 beregnet på grunnlag av valideringsstudien, og at den sanne korrelasjon mellom inntaket av mettet fett og serumkolesterol er 0,50. Det betyr at på grunn av usikkerheten ved kostdataene vil vi observere en korrelasjon på 0,34 ($0,50 \cdot 0,67$) istedenfor 0,50. Sammenhengen svekkes med 33%. I dette eksemplet er det kun feilen ved kostholdsmetoden som det er tatt hensyn til, men også feil ved måling av serumkolesterol vil redusere muligheten for å finne en sammenheng. I forhold til denne studien vil validiteten av frekvensspørreskjemaet sannsynligvis ikke være tilfredsstillende. Det er også her mulig å kompensere for utilfredsstillende validitet ved å øke antall deltagere (Cole, 1997).

Datakvaliteten i de tre ovennevnte studieeksemplene avhenger ikke bare av validiteten til kostholdsmetoden, men også av deltagelsesprosenten.

KONKLUSJON

Det er arbeidskrevende og det kan være komplisert å vurdere om validiteten til en kostholdsmetode er tilfredsstillende. Det finnes dessverre ikke et enkelt sett med kriterier som kan brukes til denne vurderingen. Hankin og medarbeidere foreslo i 1991 å bruke krite-

riene vist i tabell 3, men det er enighet om at disse er utilstrekkelige (Hankin et al., 1991).

Det er ikke mulig å snakke om validiteten til en kostholdsmetode som et generelt begrep, ettersom validiteten avhenger av design og gjennomføring av valideringsstudien, og av populasjonen inkludert i valideringsstudien. Dessuten må validiteten vurderes i forhold til den studien som kostholdsmetoden skal brukes i. Validiteten kan være tilfredsstillende i forhold til én problemstilling, men ikke til en annen.

Et enkelt sett med kriterier for vurdering av validiteten eksisterer altså ikke, men å gjennomgå og forsøke å svare på de tre nedenstående spørsmål kan være en hjelp ved vurdering av validiteten til en kostholdsmetode;

- Er design og gjennomføring av valideringsstudien tilfredsstillende?
- Er dataanalysene tilfredsstillende?
- Er validiteten tilfredsstillende i forhold til målgruppen og problemstillingen i studien hvor kostholdsmetoden skal brukes?

Selv om valideringsstudier er arbeidskrevende og kostbare å gjennomføre, er det viktig å validere en metode for kostholdsundersøkelse slik at mangelfulle kostholdsdata i en undersøkelse ikke resulterer i feil konklusjoner. Dessuten kan kunnskapen fra valideringsstudien brukes i epidemiologiske studier for å estimere utvalgsstørrelse og til å justere for målefeil ved kostholdsmetoden.

Tabell 3. Kriterier for korrelasjonskoeffisienter foreslått av Hankin og medarbeidere (1991), som kan brukes ved vurdering av om en metode for kostholdsundersøkelse er tilfredsstillende.

Korrelasjonskoeffisient > 0,50	"Satisfactory"
Korrelasjonskoeffisient 0,30–0,49	"Fair"
Korrelasjonskoeffisient < 0,30	"Poor"

REFERANSER

- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; 307-17.
- Andersen LF, Solvoll K, Johansson L, Saminen I, Aro A, Drevon CA. Evaluation of a food frequency questionnaire with weighed records, fatty acids and alpha-tocopherol in adipose tissue and serum. *Am J Epidemiol* 1999; **150**: 75-87.
- Bates CJ, Thurnham DI, Bingham SA, Margetts BM, Nelson M. Biochemical markers of nutrient intake. I: Margetts B, Nelson M. *Design Concepts in Nutritional Epidemiology*. Oxford:Oxford University Press 1997, 170-240.
- Black AE, Goldberg GR, Jebb SA, Livingstone MB, Cole TJ, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of published surveys. *Eur J Clin Nutr* 1991; **45**: 583-99.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307-10.

- Bingham SA, Cummings JH. Urine nitrogen as an independent validity measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr* 1985; **42**: 1276-89.
- Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology* 1990; **1**: 58-64.
- Carroll RJ, Pee D, Freedman LS, Brown CC. Statistical design of calibration studies. *Am J Clin Nutr* 1997; **65** (suppl): 1187S-9S.
- Cole TJ. Sampling, study size, and power. I: Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1997: 64-86.
- Delcourt C, Cubeau J, Balkay B, Papoz L, Group TC-I-ZPS. Limitations of the correlation coefficient in the validation of diet assessment methods. *Epidemiology* 1994; **4**: 518-24.
- Elmståhl S, Gullberg. Bias in diet assessment methods – consequences of colinearity and measurement errors on power and observed risks. *Int J Epidemiol* 1997; **26**: 1071-9.
- Frost Andersen L, Nes M, Lillegaard IT, Sandstad B, Bjørneboe G-E Aa, Drevon CA. Evaluation of a quantitative food frequency questionnaire used in a group of Norwegian adolescents. *Eur J Clin Nutr* 1995; **49**: 543-54.
- Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* 1991; **45**: 569-81.
- Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. *Am J Epidemiol* 1991; **133**: 616-28.
- Hunter D. Biochemical indicators of dietary intake. I: Willett W. *Nutritional Epidemiology*. New York: Oxford University Press 1998, 174-243.
- Isaksson I. Urinary nitrogen output as a validity test in dietary surveys (letter). *Am J Clin Nutr* 1980; **33**: 4-5.
- Katan MB, Deslypere JP, van Birgelen APJM, Penders M, Zegwaard M. Kinetics of the incorporation of dietary fatty acids into serum cholesteryl esters, erythrocyte membranes, and adipose tissue: an 18-months controlled study. *J Lipid Res* 1997; **38**: 2012-22.
- Klaver W, Burema J, van Staveren WA, Knuijman JT. Definitions of terms. I: Cameron ME, van Staveren WA. *Manual on Methodology for Food Consumption Studies*. Oxford: Oxford University Press, 1988.
- Kohlmeier L, Bellach B. Exposure assessment error and its handling in nutritional epidemiology. *Annu Rev Public Health* 1995; **16**: 43-59.
- Kohlmeier L, Mendez M, McDuffie J, Miller M. Computer-assisted self-interviewing: a multimedia approach to dietary assessment. *Am J Clin Nutr* 1997; **65**: 1275S-81S.
- Margetts BM, Nelson M. Overview of the principles of nutritional epidemiology. I: Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1997: 3-38.
- Marshall JA, Scarbro A, Shetterly SM, Jones RH. Improving power with repeated measures: diet and serum lipids. *Am J Clin Nutr* 1998; **67**: 934-9.
- Nelson M. The validation of dietary assessment. I: Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1997: 241-72.
- Ockè MC, Kaaks RJ. Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 1997; **65** (Suppl): 1240S-5S.
- Prentice AM. The doubly-labelled water method for measuring energy expenditure – technical recommendations for use in humans. Vienna: International Atomic Energy Agency, 1990.
- Schoeller DA. How accurate is self-reported dietary energy intake? *Nutr Rev* 1990; **48**: 373-8.
- Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr* 1997; **65** (suppl): 1179S-86S.
- Willett W. Issues in analysis and presentation of dietary data. I: Willett W. *Nutritional Epidemiology*. New York: Oxford University Press 1998a, 321-46.
- Willett W. Reproducibility and validity of food-frequency questionnaires. I: Willett W. *Nutritional Epidemiology*. New York: Oxford University Press 1998b, 101-47.
- Willett W. Correction for the effects of measurement error. I: Willett W. *Nutritional Epidemiology*. New York: Oxford University Press 1998c, 302-20.
- Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985; **122**: 51-65.

Appendiks 1. Beregning av korrelasjon mellom inntak målt med testmetoden og det sanne inntak (Nelson, 1997).

$$r_{ST} = r_{RT} / r_{SR}$$

Hvor:

r_{ST} = korrelasjon mellom inntak målt med testmetoden og det sanne inntak

r_{RT} = korrelasjon mellom inntak målt med testmetoden og inntak målt med referansemetoden

r_{SR} = korrelasjon mellom inntak målt med referansemetoden og det sanne inntak

r_{SR} kan beregnes:

$$r_{SR} = \sqrt{\frac{d}{d + s_w^2/s_b^2}}$$

Hvor:

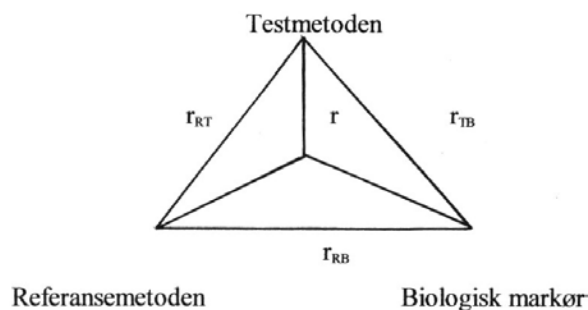
d = antall dager kostholdet er registrert med referansemetoden

s_w = innen-person-variasjon (kan beregnes på bakgrunn av kostholdsdata fra registreringen)

s_b = mellom-person-variasjon (kan beregnes på bakgrunn av kostholdsdata fra registreringen)

OBS! Denne metode tar hensyn til tilfeldige feil, men ikke systematiske feil.

Appendiks 2. Triangelmetode. En modell for beregning av korrelasjonen mellom inntak fra testmetoden og det sanne inntak, foreslått av Ockè og Kaaks (1997).



Hvor:

r_{RT} = korrelasjon mellom inntak målt med testmetoden og inntak målt med referansemetoden

r_{TB} = korrelasjon mellom inntak målt med testmetoden og biologisk markør

r_{RB} = korrelasjon mellom inntak målt med referansemetoden og biologisk markør

Alle disse korrelasjonene kan beregnes på grunnlag av valideringsstudier som inkluderer to referansemetoder, for eksempel en annen kostholdsundersøkelsesmetode og en biologisk markør.

Korrelasjonen mellom inntak fra testmetoden og det sanne inntak kan da beregnes:

$$r = r_{TB} \cdot r_{RT} / r_{RB}$$

OBS! En forutsetning for at denne modellen kan brukes er at feilene mellom test- og referansemetoden er uavhengige.