

Norsk Akkreditering: Akkrediteringsdag ISO 17025 | Gardermoen 05.12.2024

# Søknad om utvidelser; krav til validering og verifikasjon.

**Anders T. Halvorsen**

Kjemiker, Haukeland Universitetssykehus

Teknisk bedømmer ISO 15189 (M04,M12)

ISO 17025 (P12)

# Agenda

- (1) Introduksjon
- (2) Faglige vurderinger
- (3) Noen valideringsparametere
- (4) Spørsmål og diskusjon

**(1)**

**Introduksjon**

# Omfanget

Denne presentasjonen er ikke eneste måten å validere på

Ikke «best practice», men et greit startpunkt

Ikke en komplett gjennomgang av valideringsparametere

Visse fagområder kan ha «pålagte» valideringsprosedyrer (f.eks. SANTE / AQC-Guidelines for pesticider)

Utgangspunktet er kvantitative kjemiske analyser

# Validering og verifisering

**Verifisering:** bekreftelse på at spesifiserte krav er oppfylt, ved å fremskaffe objektivt bevis

**Validering:** bekreftelse på at metoden er egnet til formålet, ved framskaffelse av objektivt bevis for at de angitte kravene har blitt oppfylt

# Kompetanse til ansvarlige for validering

Kjennskap til standard og NA-krav

Erfaring fra validering

Statistikk

Instrumentering

Fagområdet

# Hvilken informasjon skal sende til TB?

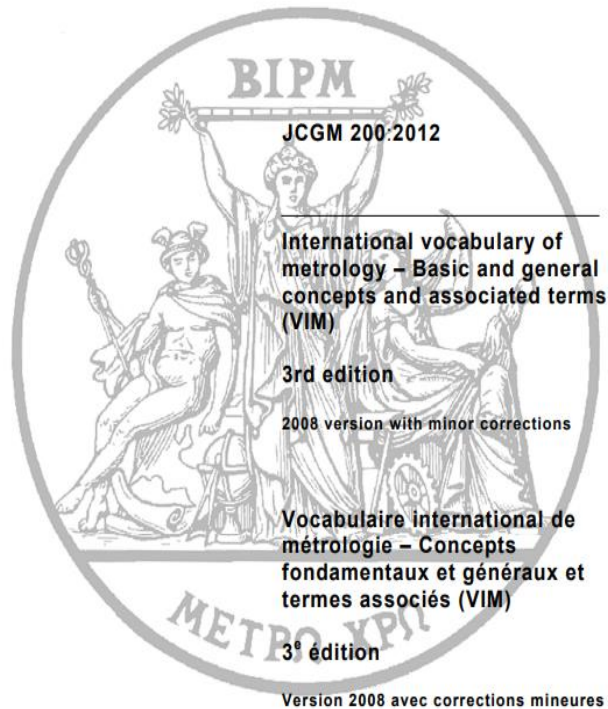
Utfylt Søknads- og akkrediteringsomfang NS-EN ISO/IEC 17025 prøving  
(dok00283)

Valideringsplan

Valideringsrapport

Rådata (beregninger)

# VIM (JCGM 200 eller ISO/IEC Guide 99)



© JCGM 2012

International vocabulary of metrology – Basic and general concepts and associated terms  
(**VIM**)

**V**ocabulaire **i**nternational de **m**étrologie

VIM4 i “final draft”

VIM er normative referanse i ISO 17025 og ISO 15189, angitt som ISO/IEC Guide 99



(2)

Faglige vurderinger

# Faglige vurderinger

Metode må være «fit for purpose». Dette krever mer enn å bare liste opp resultater. ISO 15189:2022 har fått spesifikke krav om erklæringer som skal står i rapport

**Tommelfingerregel:** Alle resultater som presenteres skal faglig vurderes/kommenteres

En valideringsplan med faglig begrunnede, **konkrete** kvalitetsmål

Systematisk rapport med relevante data og enheter (NB: %-enheter)

“ “Graphs are essential to good statistical analysis.

*Francis John Anscombe, engelsk statistiker*

# Anscombes kvartett

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,1	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,1	4	5,39	19	12,5
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89

## Samme statistikk for alle sett

N = 11

Snitt X = 9,00

Snitt Y = 7,50

Regresjonslinje  $Y = 0,500x + 3,00$

Standardfeil stigningstall: 0.118

t-score = 25,445

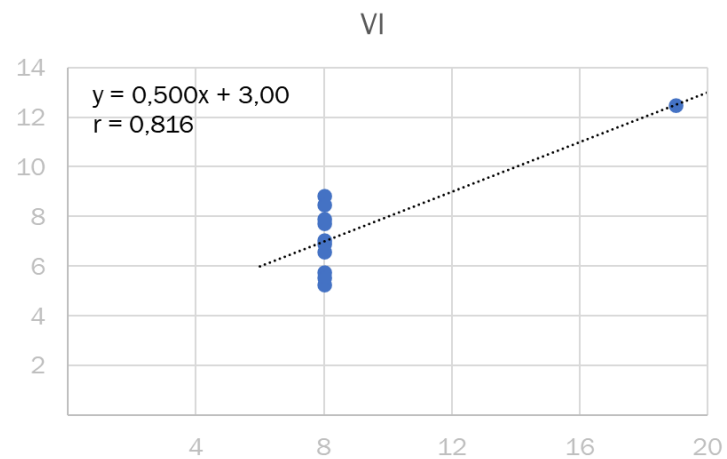
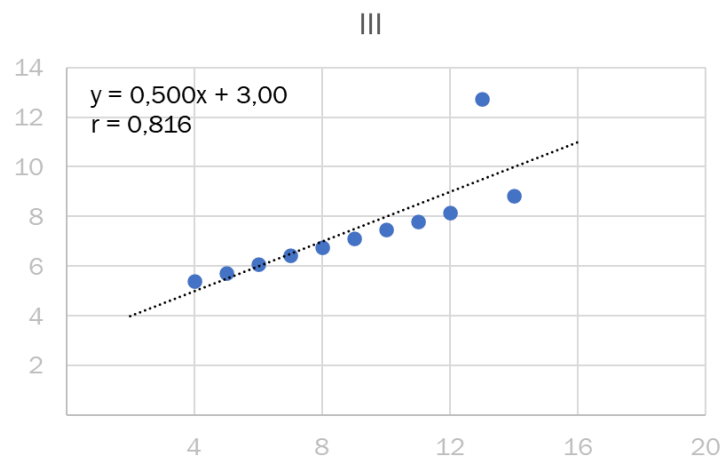
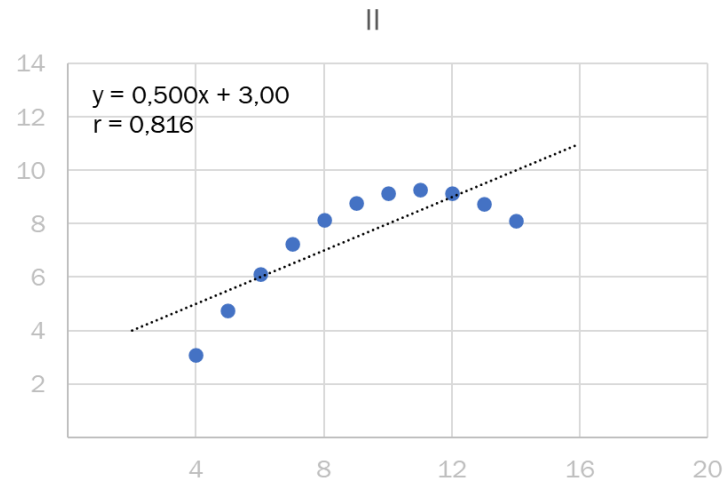
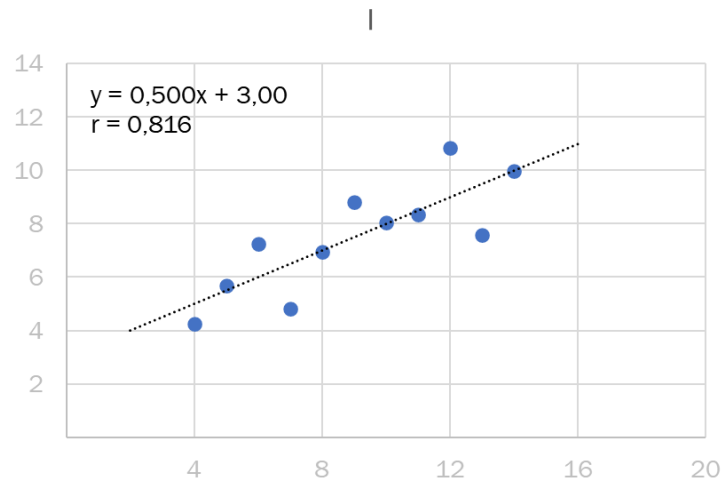
Sum av kvadrat  $X - \bar{X} = 110,0$

$R^2 = 0,67$

Varians Y = 4,12

etc

# Anscombes kvartett



## Samme statistikk for alle sett

$N = 11$

Snitt  $X = 9,00$

Snitt  $Y = 7,50$

Regresjonslinje  $Y = 0,500 x + 3,00$

Standardfeil stigningstall: 0.118

t-score = 25,445

Sum av kvadrat  $X - \bar{X} = 110,0$

$R^2 = 0,67$

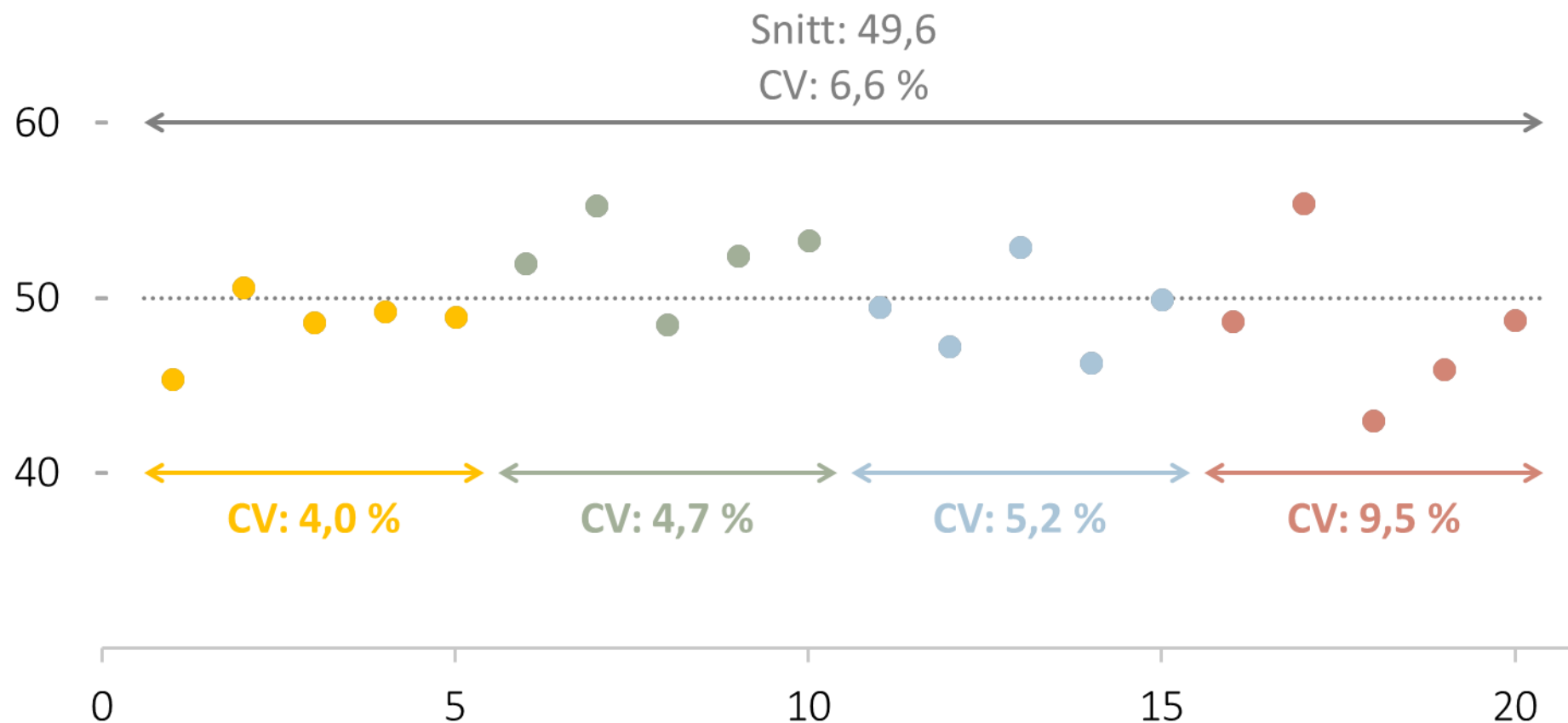
Varians  $Y = 4,12$

etc

# Usikkerhet i standard-avvik

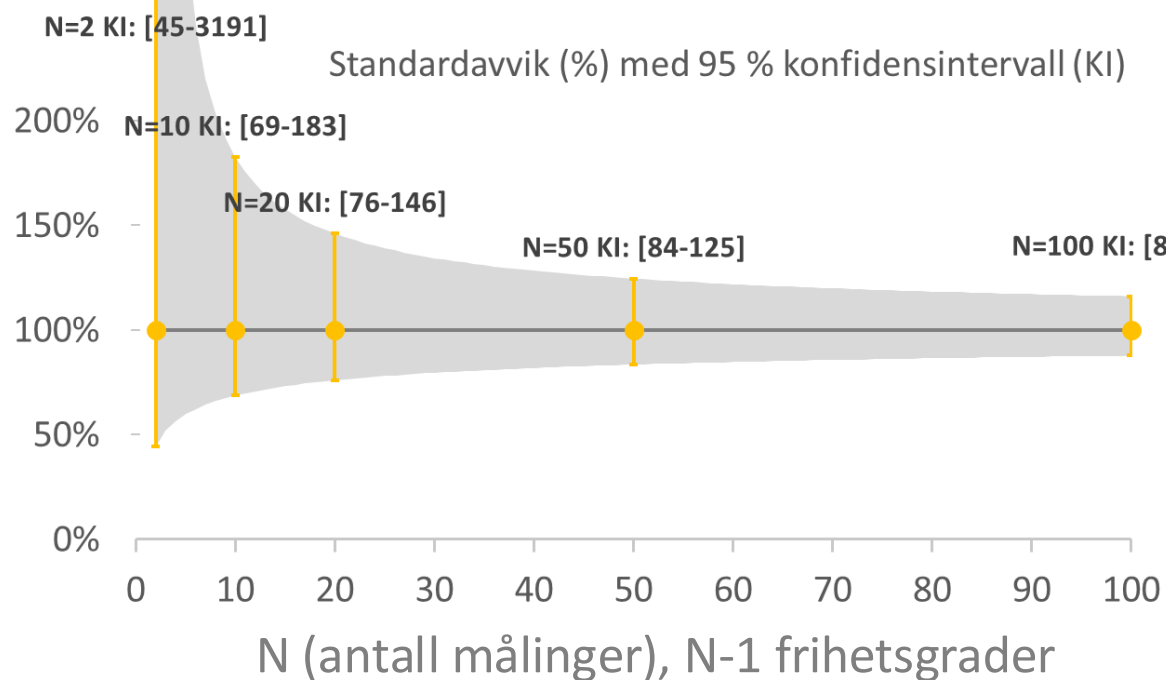
Estimering av standardavvik/CV har usikkerhet i seg selv

Eksempel: Normalfordelte data,  $N = 20$ , snitt = 50 og CV = 5,0%.



# Usikkerhet i standard-avvik

Usikkerhet i standardavvik minker med antall målinger



**Hva vil dette si for  
presisjon?**

Eksempel:

**CV: 4,5 %**

**N= 100**

**95 % KI: 4,0 – 5,2 %**

**N=50**

**95 % KI: 3,8 – 5,6 %**

(3)

Noen valideringsparametere



# Valideringsparametere

*Merk: Noen guider definerer parametere litt forskjellig.*

Selektivitet

**Presisjon**

Overdrag (Carry-over)

Deteksjonsgrense

**Riktighet**

**Metode-  
sammenligning**

Kvantifiseringsgrense

**Måleusikkerhet**

Andre

**Linearitet**

Robusthet

**Linearitet** til en analytisk prosedyre er evnen til å gi et prøveresultat som er direkte proporsjonalt med konsentrasjonen av analytt i prøven

# Linearitet

Ikke aktuelt for alle analyser

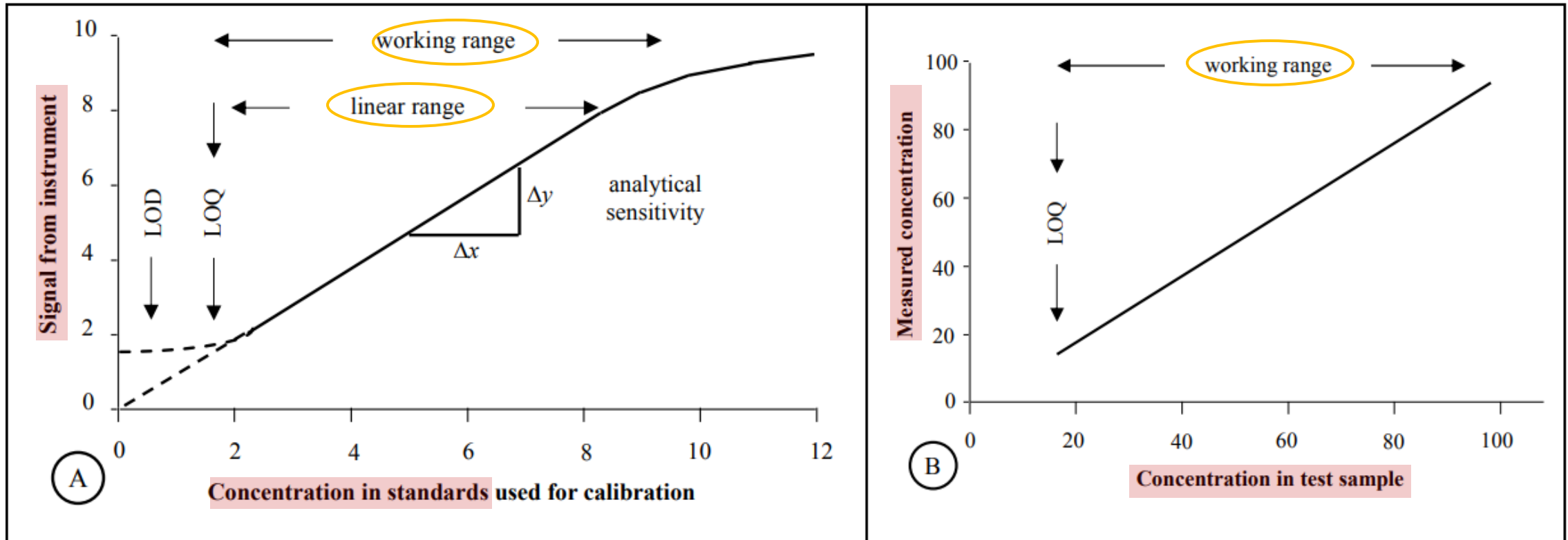
Ikke påkrevd at en metode skal være lineær

Andre kalibreringsmodeller (f.eks. kvadratisk kalibreringskurve) kan valideres under dette punkt.

Mange begreper («method working range», «instrument working range», kalibreringsområde, måleområde) om delvis samme parameter

## Instrument

## Metode



**Figure 3 – A) Typical example of response curve obtained with an instrumental method. The performance characteristics ‘working range’, ‘linear range’, ‘analytical sensitivity’, ‘LOD’ and ‘LOQ’ are identified. B) Typical example of a curve obtained with a measurement procedure where the test sample concentration is plotted versus measured concentration.**

## Linearitet – Hva valideres?

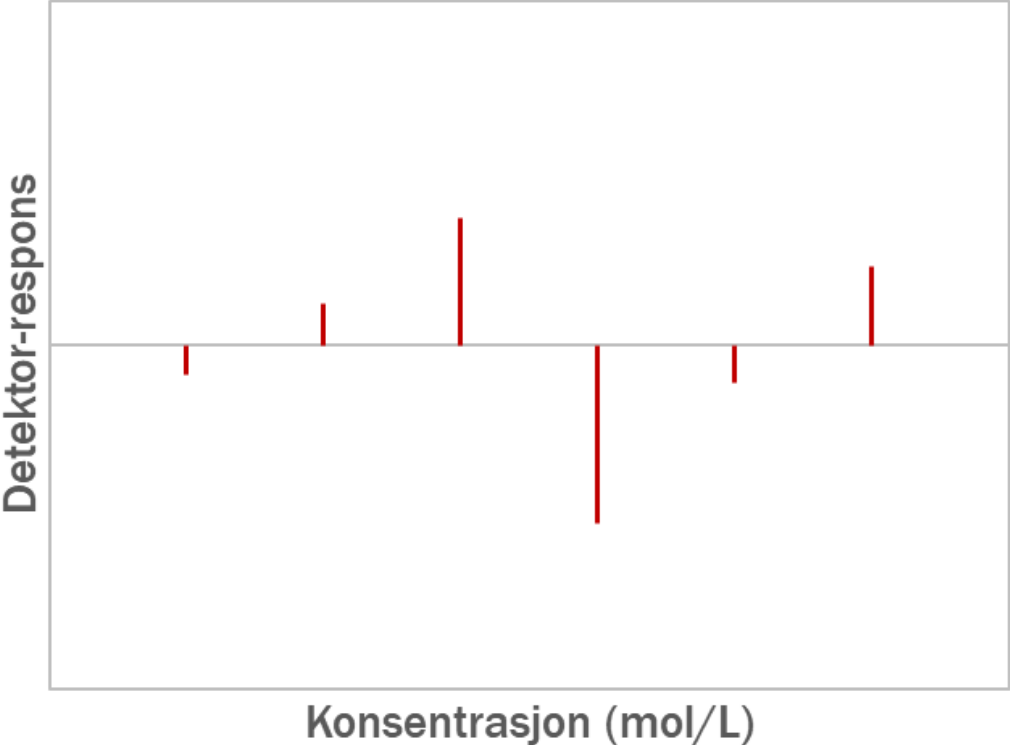
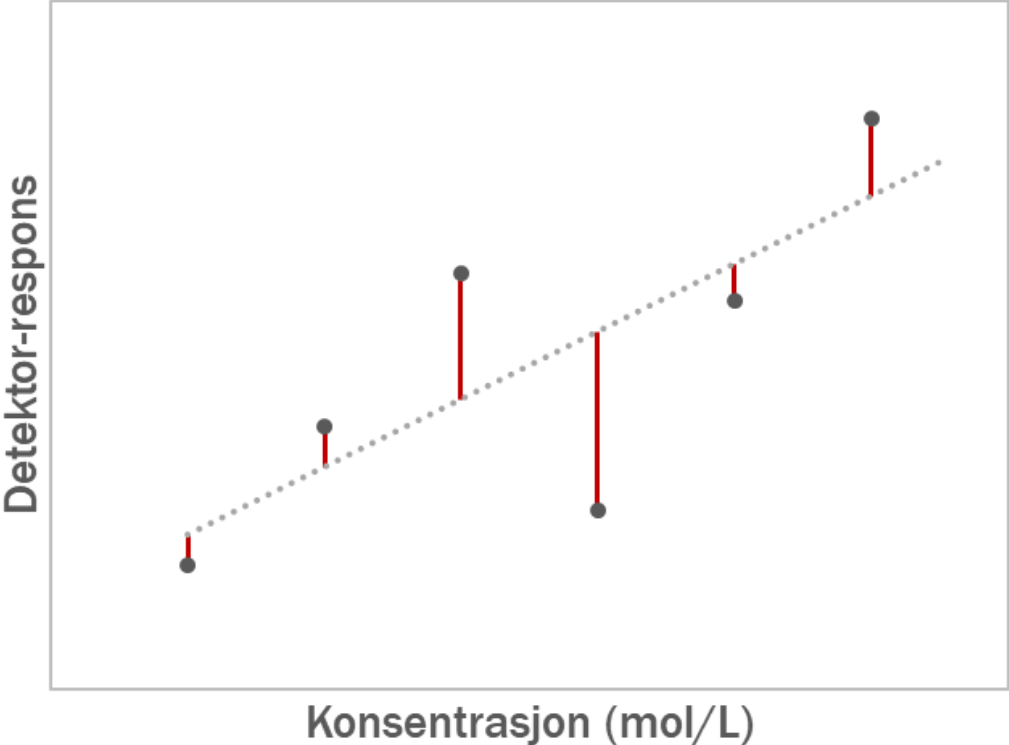
Korrelasjonskoeffisient er **ikke** et mål på linearitet (jamfør Anscombes kvartett).

Minimum: Visuell vurdering av spredningsdiagram med regresjonslinje og residualplott

Vær ekstremt nøyaktig og få minst mulig usikkerhet i prøver til linearitetstest

Minimum 6 konsentrasjonsnivå, jevnt fordelt. Utvide gjerne 20% utover området som skal testes

# Linearitet – spredningsdiagram og residualplott



**Presisjon:** grad av overenstemmelse mellom indikasjoner eller målte verdier oppnådd ved gjentatte målinger på samme eller lignende objekter under gitte betingelser.

# Presisjon

Presisjon beskriver spredning av resultater som følge av tilfeldig feil i metoden.

Estimeres på bakgrunn av gjentatte målinger

Presisjon  $\neq$  måleusikkerhet (men nært beslektet)

Uttrykt som et **standardavvik**



# Presisjon

*Merk: Forskjellige guider har forskjellig navn/definisjoner på typer presisjon. Som et minimum bør valideringen inkludere en presisjon for hele metode og en presisjon for innen-serie.*

## Innen serie presisjon (**repeterbarhet**)

*Samme måling gjort under like omstendigheter*

## Total-presisjon (**reproduserbarhet**)

*Samme måling gjort over tid, av forskjellige operatører osv.*

# Presisjon

Bør inkludere flere konsentrasjonsområder

Bør minimum inkludere (nær) kvantifiseringsgrense og øvre del av kalibrerings-/måleområde

## Innen-serie-presisjon (repeterbarhet)

Enten gjentatte målinger i samme serie (minimum 6)

Eller få paralleller i en prøveserie gjentatt over flere dager (2 x 20 dager)

Formel for duplikate målinger («par») i en serie over flere dager:

$p$  = antall par,  $d$  = differanse

(Egentlig: Gjennomsnittet av **daglig varians**)

$$S = \sqrt{\frac{\sum_{i=1}^p (d_i^2)}{2p}}$$

## Total-presisjon (reproduserbarhet)

Flere målinger over flere dager, forskjellige operatører, etc

En måling fra hver serie/dag

Noen guider tillater noen målinger fra samme dag

Alternativ: ANOVA-analyse på balansert design

# Hvordan rapportere presisjonsdata

*Enheter Konsentrasjonsnivå Uteligger, hvilken påvirkning har de?*

**Tabell 1:** Presisjonsdata for to nivåer (batch L1-1901 og L2-2101). 95% KI i parentes.

		Lav kontroll		Høy kontroll
		L1-1901	L1-1901 (eks. uteligger)*	L2-2101
<b>Snitt</b>	µmol/L	0,208	0,208	0,415
<b>Innen serie</b>	N	56	54	56
	SD, µmol/L	0,009 (0,0078 - 0,0114)	0,007 (0,0060 - 0,0088)	0,011 (0,0095 - 0,0139)
	CV, %	4,5 (3,76 - 5,49)	3,4 (2,87 - 4,22)	2,7 (2,30 - 3,36)
<b>Totalpres.</b>	N	56	54	56
	SD, µmol/L	0,010 (0,0087 - 0,0127)	0,009 (0,0079 - 0,0116)	0,015 (0,0122 - 0,0179)
	CV, %	5,0 (4,19 - 5,34)	4,5 (3,79 - 5,56)	3,5 (2,95 - 4,30)

*Konfidensintervall og/eller antall*

*Kilde: Egne data*

# Hvordan rapportere presisjonsdata

Kan bruke diagram, som oftest er oppsummeringstabell nok

Faglig vurdering av resultatene (uteliggere, som forventet, etc)

Er kvalitetsmål nådd? Godt innenfor eller akkurat innenfor?

Delkonklusjon: Hva er presisjonen, i hvilke nivå?

**Riktighet:** grad av overensstemmelse mellom gjennomsnittet av et uendelig antall replikater av målte kvantitative verdier og en kvantitativ referanseverdi

## Riktighet (trueness)

Kan ikke gjøre et uendelig antall replikater, så riktighet kan ikke bestemmes nøyaktig

Uttrykt som en bias mot en referanseverdi.

Fravær av bias kan ikke bestemmes, en bias på 0 vil alltid ha en usikkerhet i seg



# Riktighet (trueness)

- 1) Mot (sertifisert) referansemateriale (SRM/RM)
- 2) Mot SLP-resultater
- 3) Mot en referansemetode
- 4) Ved spiking (tilsetning) av kjent mengde til prøvemateriale

## Riktighet (trueness)

I praksis: Kan være vanskelig med nok resultater til et robust estimat av bias ved «førstegangsvaldiering»

Husk å ta hensyn til usikkerhet i «sann verdi» i tillegg til avvik

Vurder om bias er konsentrasjonavhengig

Oppgi alle resultater i tabell: Eget resultat, SLP tillagt verdi/ SRM, eget bias, samt usikkerhet i «sann verdi» og egne avvik (+SLP akseptgrense)

**Måleusikkerhet:** ikke-negativ parameter som kjennetegner spredningen av kvantitative verdier som har blitt tildelt en målestørrelse, basert på informasjonen som er benyttet

# Måleusikkerhet

**Vilkårsdokumentet** angir at laboratoriet skal oppgi følgende:

Usikkerhet i prøving skal angis med **dekningsfaktor  $k = 2$**

Måleprinsipp som benyttes for hver aktivitet samt aktuelle **måleområder med tilhørende beregnet/estimert måleusikkerhet**

I praksis: Usikkerhet beregnet i hver ende av måleområdet som et minimum, med dekningsfaktor 2.

# Måleusikkerhet

Vanlig å estimere måleusikkerhet basert på presisjon og riktighet

Merk: VIM (ISO/IEC Guide 99) angir to måter å estimere usikkerhet på

*Nordtest Report TR 537 Handbook for calculation of measurement uncertainty in environmental laboratories* («Trollboka») gir en god oversikt over beregning av måleusikkerhet

Ved validering: Vurder også usikkerhet i kalibrator dersom kun en kalibrator er benyttet under presisjonsforsøk

**Metodesammenligning** bør inkluderes ved overgang til nytt instrument, vesentlig forandring av metode, men også ved flytting til ny lokaliteter og for å sammenligne seg mot eksterne laboratorier

# Metodesammenligning

Rapport bør ha både diagrammer og oppsummerings-tabeller, pga komplekse data

Kan være aktuelt å dele inn i forskjellige konsentrasjonsområder

Bør bestå av relativt mange resultater

# Metodesammenligning

Plan bør ha krav til både gjennomsnittlig forskjell mellom metoder, for eksempel: Gjennomsnittlig forskjell skal være  $\leq 5\%$

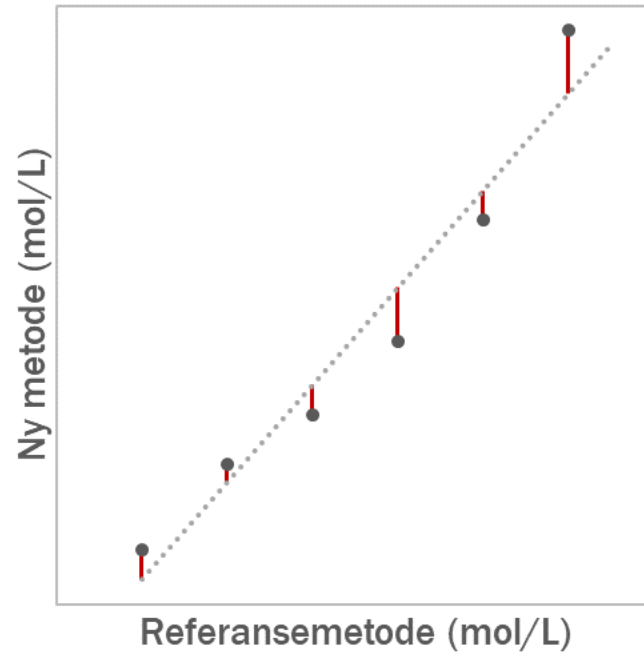
Fordel med krav på maks antall prøver med avvik utenfor gitt grenser, for eksempel: 95% av prøvene skal ha et avvik på  $\leq 15\%$

Rapport bør ha både diagrammer og oppsummerings-tabeller, pga komplekse data.



# Tre vanlige regresjonsmodeller til metodesammenligning

## Minste kvadrater



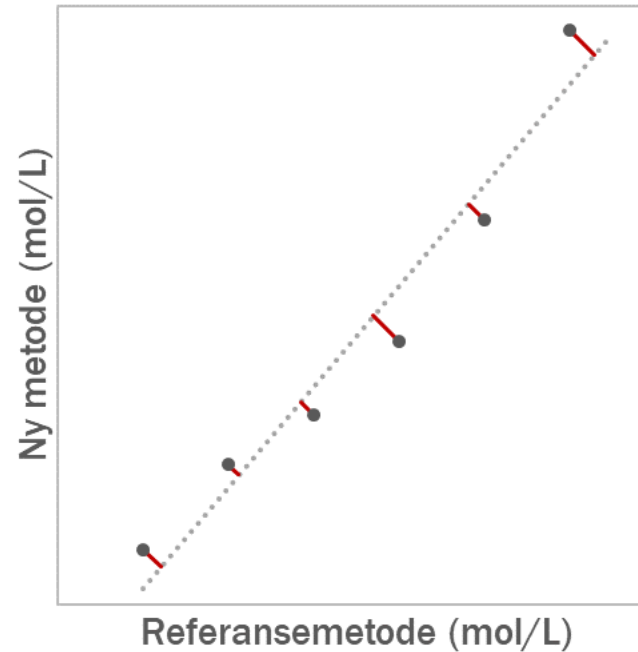
*Inkludert i Excel*

*Kun usikkerhet i Y*

*Kan vektes (ikke i Excel)*

*Ikke anbefalt*

## Deming



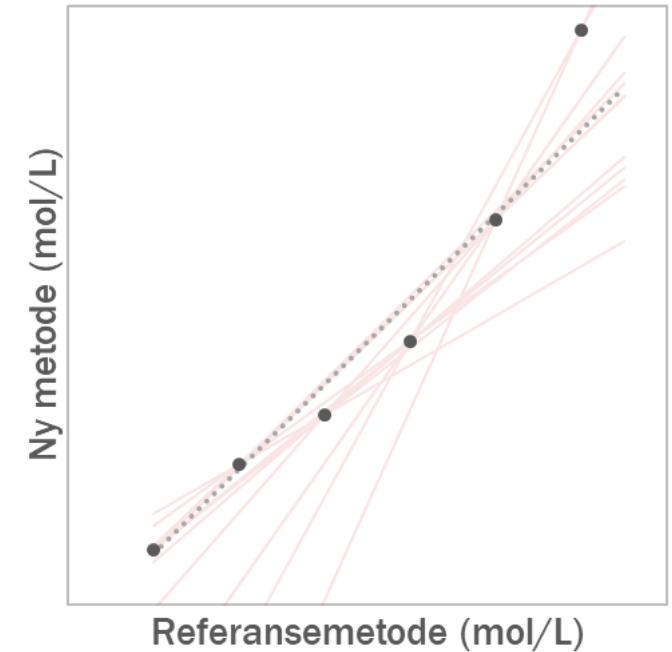
*Ikke i Excel*

*Kan angi usikkerhet X og Y*

*90° ved lik usikkerhet*

*Kan vektes.*

## Passing-Bablok

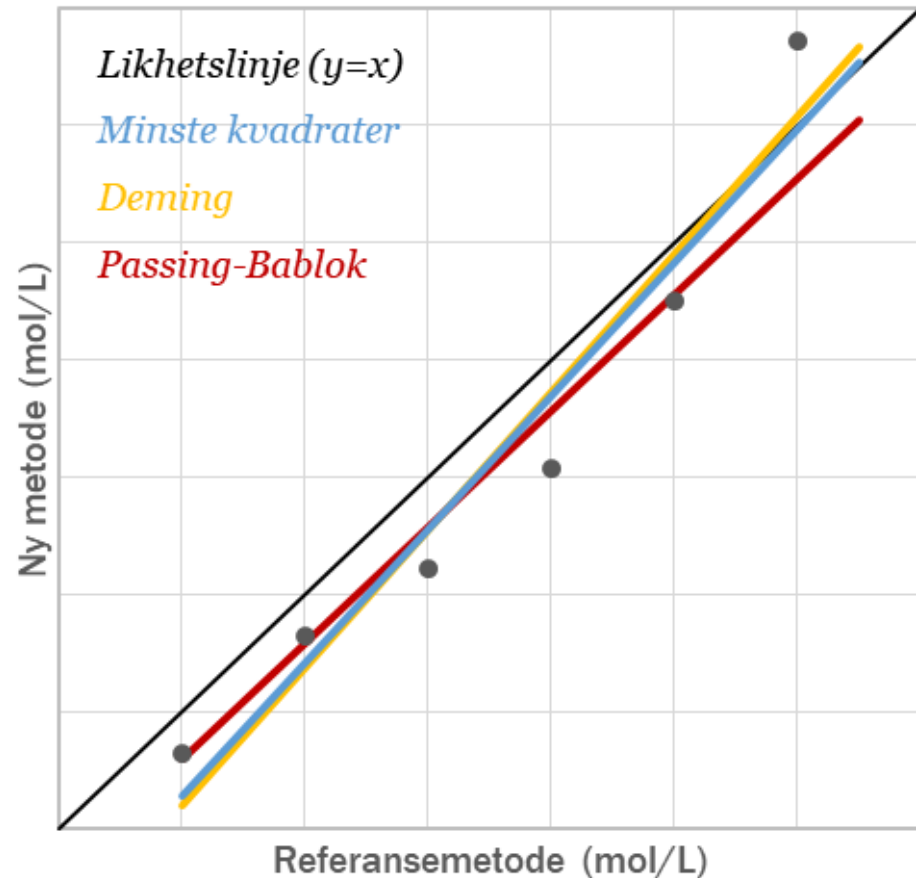


*Ikke i Excel*

*Ikke-parametrisk, «median»  
stigningstall.*

*Lite følsom for uteliggere.*

# Tre vanlige regresjonsmodeller til metodesammenligning



Modellene vil gi forskjellige regresjonslinjer

Når  $u$  i ny metode er mye større enn referansemetode, er Deming og minste kvadrater tilnærmet like

Passing-Bablok skiller seg ofte fra de andre ved uteliggere

# Metodesammenligning: Spredningsdiagram

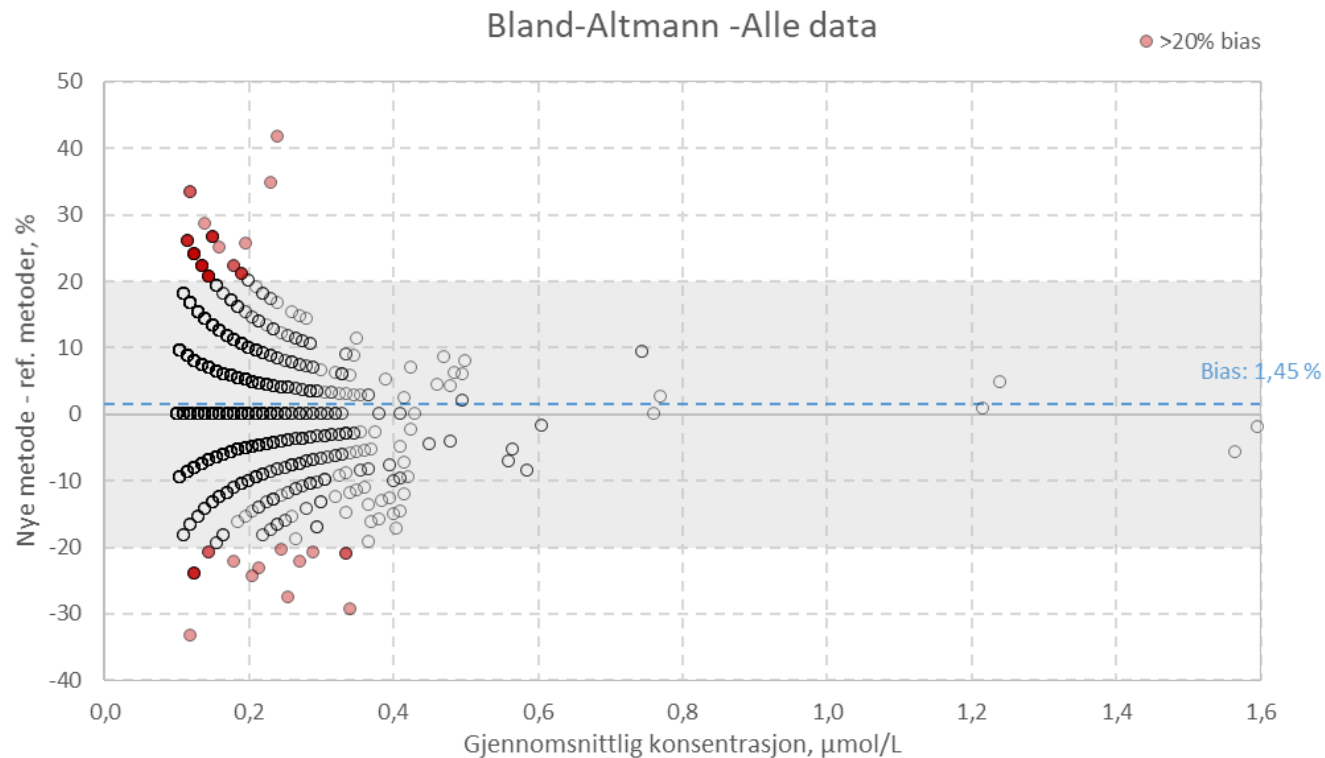
Minste kvadraters regresjon er akseptabel, men tolk med forsiktighet

Stigningstall angir proporsjonal bias (prosentvis forskjell), skjæringspunkt  
konstant bias

Dersom tilgjengelig, rapporter konfidensintervall på stigningstall og  
skjæringspunkt (forskjell fra hhv 1 og 0 angir signifikant forskjell)

Kombiner med Bland-Altman og statistiske tester (t-test)

# Metodesammenligning: Bland-Altman plott



*Prosentvis og/eller absolutt forskjell på Y-akse*

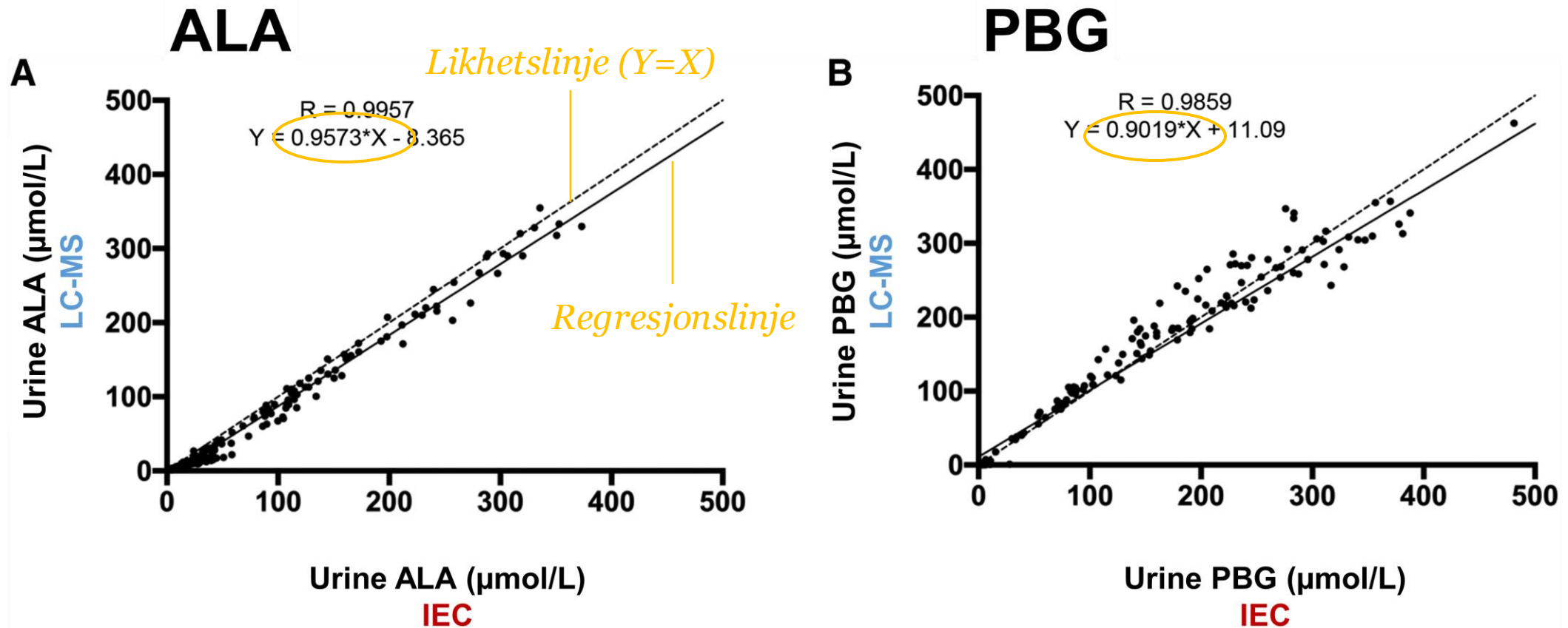
*Gjennomsnitt av metodene på X-akse*

*Linje med gjennomsnittlig forskjell*

*Grenser, gjerne krav på enkeltresultater, eller 95% konfidensintervall.*

# Eksempel: Ionebyttekolonne og UV (IEC) mot LC-MS.

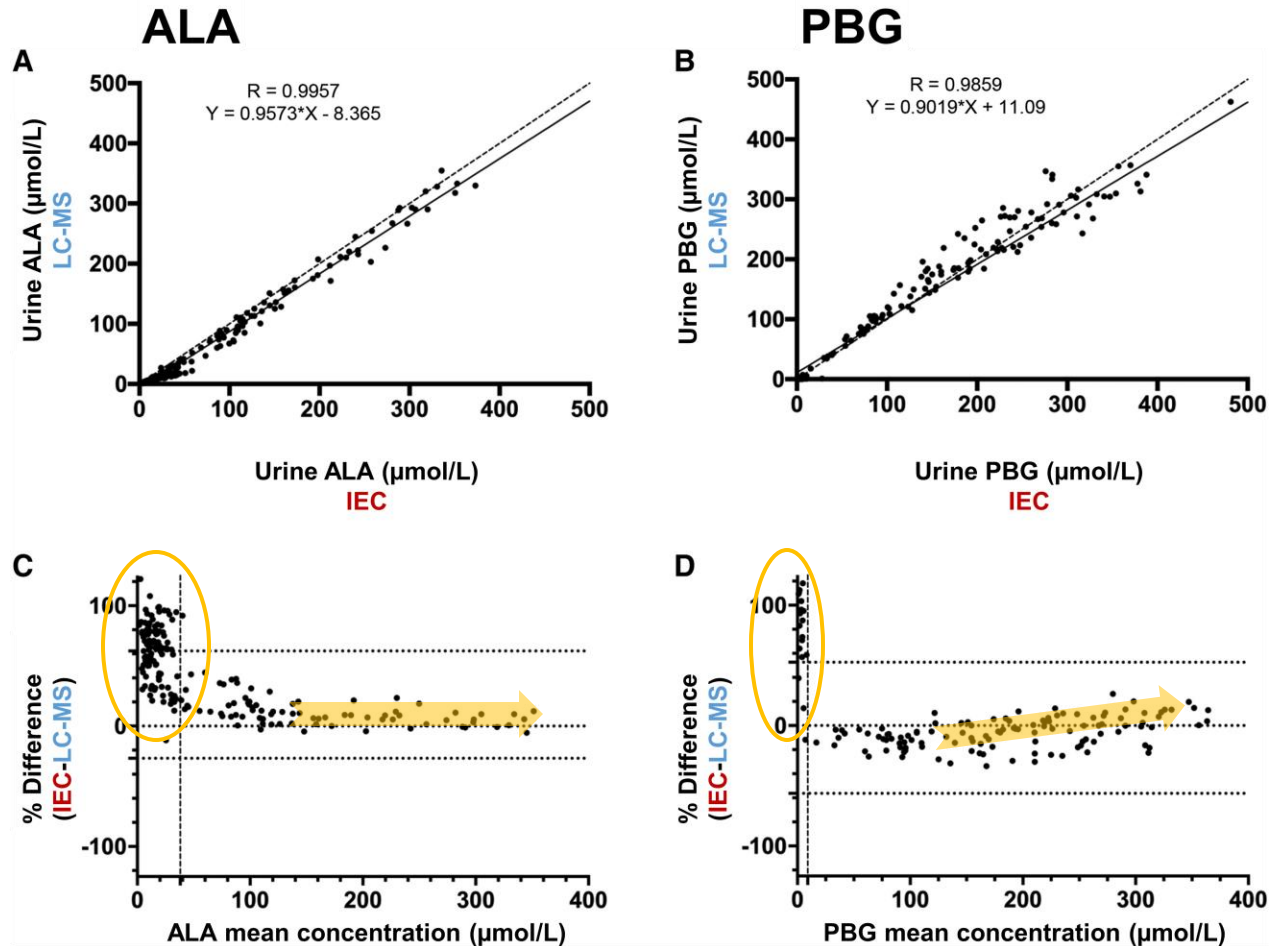
Aminolevulinsyre (ALA) og porfobilinogen (PBG)



Kilde: Poli et al, Clin Chem. 2023 Oct 3;69(10):1186-1196

# Eksempel: Ionebyttekolonne og UV (IEC) mot LC-MS.

Aminolevulinsyre (ALA) og porfobilinogen (PBG)



Spredningsdiagram

Bland-Altman

# Hvordan rapportere metodesammenligning?

Tabell med oppsummering:

*Antall prøver*

*Maks og min resultat per metode*

*Gjennomsnitt for hver metode, gjennomsnittlig bias*

*Antall enkeltresultater med avvik større enn krav.*

Spredningsdiagram, f.eks. Passing-Bablok eller Demming (+ statistikk).

Bland-Altman plott, med grenser

# Hvordan rapportere metodesammenligning?

Vurder om det er nødvendig å dele inn i flere konsentrasjonsnivåer

Husk at eventuell bias påvist kun er gyldig for utvalget

Parvis statistisk test, for eksempel t-test.

Delkonklusjon, hva er påvist bias, er den reell og akseptabel.



(4)

Spørsmål og diskusjon

# Kilder

ISO/IEC 17025:2017

ISO 15189:2022

Norsk Akkreditering - Vilkår for å være akkreditert (versjon 6.00)

ISO/IEC Guide 99:2007 - International vocabulary of metrology - Basic and general concepts and associated terms (VIM)

Eurachem: The Fitness for Purpose of Analytical Methods - A Laboratory Guide to Method Validation and Related Topics

# Kilder

M. Thompson, S. L. R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC technical report), *Pure Appl. Chem.*, 2002, 74(5), 835.

Bertil Magnusson, Teemu Näykki, Håvard Hovind, Mikael Krysell, Eskil Sahlin, Handbook for calculation of measurement uncertainty in environmental laboratories (NT TR 537 – Edition 4)

Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician*, 27(1), 17–21.  
<https://doi.org/10.1080/00031305.1973.10478966>

Poli A, Manceau H, Nguyen AL, et al. Quantification of Urine and Plasma Porphyrin Precursors Using LC-MS in Acute Hepatic Porphyrrias: Improvement in Routine Diagnosis and in the Monitoring of Kidney Failure Patients. *Clin Chem*. 2023;69(10):1186-1196. doi:10.1093/clinchem/hvad117