

Hovedbudskap

- Produksjonsdata kan brukes til å estimere referansegrenser for friske individer hvis man bruker datamaskinprogrammer som er laget til formålet, såkalt indirekte metoder.
- Indirekte metoder har visse forutsetninger, og resultatene kan være vanskelige å kontrollere.
- Uansett må referansegrenser ikke oppfattes som beslutningsgrenser.

Sammendrag

Introduksjon: En referansepopulasjon av friske, unge kvinner uten jernmangel er vanskelig å definere. Derfor brukte vi såkalt indirekte metoder til å estimere referansegrenser for p-ferritin for denne populasjonen.

Materiale og metode: Grunnlaget var produksjonsdata, p-ferritin fra 25109 kvinner i 30-årsalderen. P-ferritin ble analysert med Abbott Alinity fra Abbott Diagnostics. Vi brukte datamaskinprogrammene refineR og TMC13 til å beregne referansegrensene. Som en kontroll brukte vi referansegrenser beregnet med publiserte fordelingsparametre av p-ferritin hos 38 år gamle kvinner med jern påvist i benmargen og p-ferritin sporbar til samme standard som Abbott Alinity-metoden.

Resultater: De to programmene kom fram til omtrent like referansegrenser. Men sammenlignet med de referansegrensene som kunne beregnes på grunnlag av publiserte fordelingsparametre, var 2,5-persentilen lavere og 97,5-persentilen høyere.

Konklusjon: Hva som er riktige referansegrenser, er usikkert. Referansegrenser må uansett ikke oppfattes som riktige beslutningsgrenser.

Nøkkelord

Ferritin, referansegrenser, indirekte metoder, beslutningsgrenser, jernmangel

■ Bioingeniøren er godkjent som vitenskapelig tidsskrift. Artikkelen er fagfellevurdert og godkjent etter Bioingeniørens retningslinjer.

Bruk av produksjonsdata til estimering av referansegrenser for p-ferritin – en blandet erfaring

Arne Åsberg og Øyvind Skadberg

Avdeling for medisinsk biokjemi, Stavanger universitetssjukehus

Kontaktforfatter: arneasb@gmail.com

Innledning

Referansegrenser er etterspurt for alle analytter (1). Slike grenser er relativt tungvinte og kostbare å estimere fordi estimatene er basert på analysing av prøver fra friske referansepersoner. Hvis vi trenger referanseverdier (prøvesvar) fra minst 120 referansepersoner i hver gruppe og har flere grupper, kan det totale behovet bli flere hundre referansepersoner. Gruppeinndelingen er som regel etter kjønn og alder, men inndeling etter andre faktorer, som for eksempel lengden av svangerskap, kan også være aktuelt. Referansepersoner blant barn er spesielt vanskelige å skaffe. For hver gruppe beregnes referansegrenser som 2,5- og 97,5-persentilen i fordelingen av referanseverdier. Dette kalles «direkte metoder». Deres motstykke, «indirekte metoder», baserer seg på bruk av gratis prøvesvar fra produksjonsdata. Slike prøvesvar kommer fra ulike pasientgrupper. Vi kan tenke på den totale fordelingen av prøvesvar som en blanding av ulike fordelinger fra syke og friskere pasi-

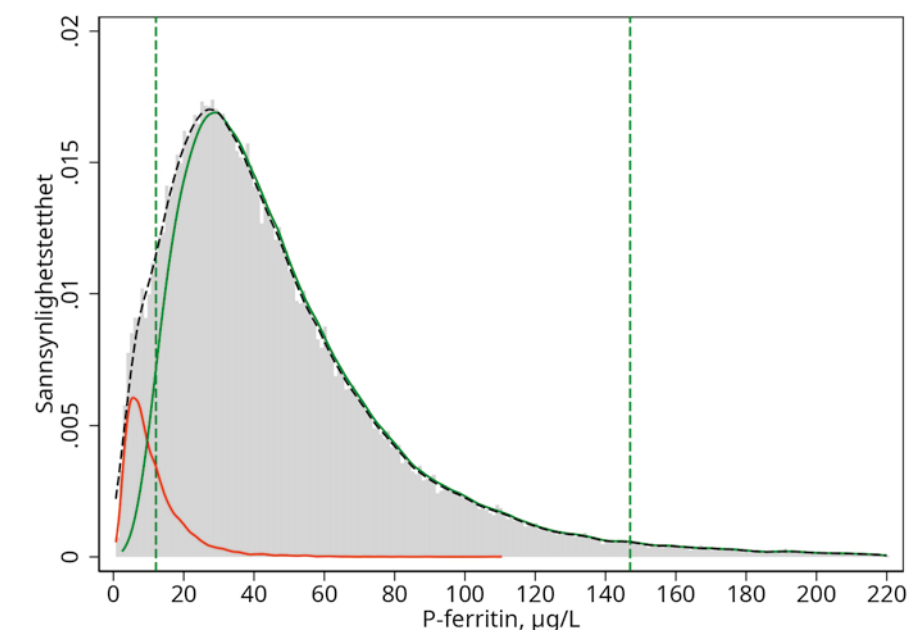
enter. De indirekte metodene skal finne den fordelingen av prøvesvar som kommer fra friske individer. Ved hjelp av litt ulike teknikker finner de den mest sannsynlige, matematisk definerbare, største fordelingen (den som rommer de fleste prøvesvarene). Skal dette være prøvesvar fra friske, må selvsagt mesteparten av prøvesvarene i den totale fordelingen stamme fra friske individer. Et annet krav er at den sentrale delen av den totale fordelingen må bestå nesten utelukkende av prøvesvar fra friske. I flere tilfeller kan dette gi troverdige resultater (1), men ikke alltid (1, 2). Hvordan kan vi stole på resultatene? Gir de indirekte metodene omtrent samme referansegrenser som de direkte, tror vi gjerne at resultatene er til å stole på (1, 3). Vi mener at de indirekte metodene er til mest hjelp når vi ikke kan eller vil bruke de direkte metodene. Ta for eksempel referansegrenser for p-ferritin for yngre kvinner. Et referansemateriale bestående av yngre kvinner uten jernmangel er vanskelig å identifisere klinisk, siden mange

med tomme jernlagre ikke føler seg syke. Friske blodgivere er heller ikke representative da de har økt tap av jern, så her kunne indirekte metoder være til hjelp. Vi har prøvd to datamaskinprogrammer for indirekte metoder, refineR (4) og TMC13 (5). Som fasit brukte vi publiserte data om fordelingen av p-ferritin hos yngre kvinner med påvist jern i benmargen.

Metoder

Fra perioden 3. januar 2022 til 1. september 2025 ble det samlet inn data om p-ferritin fra pasienter i primærhelsetjenesten, kun ett prøvesvar (det eldste) fra hver pasient. Av disse var 25109 kvinner i alderen fra og med 30 år til 40 år. P-ferritin ble analysert med Abbott Alinity fra Abbott Diagnostics, Abbott Park, Illinois, USA, med reagenser fra produsenten. Analysemetoden var sporbar til den internasjonale standarden 80/602, også kalt «First WHO Human Liver Ferritin International Standard» (6). Den totale variasjonskoeffisienten var omtrent 5 %. Vanlige prosedyrer for intern og ekstern statistisk kvalitetskontroll ble fulgt. Laboratoriet var akkreditert for analysen. På grunnlag av disse p-ferritin-data ble referansegrenser (2,5- og 97,5-persentilene) estimert med datamaskinprogrammene refineR (versjon 1.6.2) og TMC13.

Som en kontroll brukte vi data om persentiler i fordelingen av p-ferritin hos 38 år gamle kvinner med og uten jern i benmargen (7). Av 203 kvinner ble 105 klassifisert som jernsuffisiente (uten jernmangel) fordi de hadde «clearly visible reticuloendothelial iron» i benmargen, mens 69 ble klassifisert med jernmangel fordi de hadde «no stainable iron in their bone marrow smears». I tillegg var det 29 kvinner som ikke kunne klassifiseres. Hos kvinner uten jernmangel var 50- og 97,5-persentilene henholdsvis 42 µg/L og 145 µg/L. Hos kvinner med jernmangel var 50- og 97,5-persentilene henholdsvis 9 og 35 µg/L. Analysemetoden for p-ferritin skulle være sporbar til den internasjonale standarden 80/602 (7). Vi forutsatte at log(p-ferritin) er normalfordelt, siden det er en rimelig antakelse for mange analytter (8). Som gjennomsnitt i fordelingen av log(p-ferritin) brukte vi log(50-persentilen i fordelingen av p-ferritin). Avstanden fra



FIGUR 1: Fordeling av p-ferritin fra kvinner med og uten jernmangel. Figuren er basert på kunstige data med fordelingsparametre som beskrevet av Hallberg et al. (7) for kvinner i 38-årsalderen. Det ble simulert 90 000 data for kvinner uten jernmangel (grønn kurve) og 10 000 for kvinner med jernmangel (rød kurve). I dette materialet var 2,5- og 97,5-persentilen henholdsvis 12,1 og 147 µg/L for kvinner uten jernmangel (grønne, stiplede vertikale linjer). Bare 66,1 % av verdiene for kvinner med jernmangel var lavere enn 12,1 µg/L.

gjennomsnittet til 97,5-persentilen i fordelingen av log(p-ferritin) er 1,96 standardavvik i den fordelingen, siden det er en normalfordeling. Som standardavvik i fordelingen av log(p-ferritin) brukte vi derfor $[\log(97,5\text{-persentilen i fordelingen av p-ferritin}) - \log(50\text{-persentilen i fordelingen av p-ferritin})] / 1,96$. Referansegrenser ble estimert som gjennomsnitt $\pm 1,96$ standardavvik i fordelingen av log(p-ferritin) for kvinner med påvist jern i benmargen, og deretter transformert til måleskala.

Til en ekstra prøve av de indirekte metodene laget vi en samling av 10 000 p-ferritin-verdier med gjennomsnitt og standardavvik i fordelingen av log(p-ferritin) som for kvinner med jernmangel, og 90 000 med gjennomsnitt og standardavvik i fordelingen av log(p-ferritin) som for kvinner uten jernmangel. Programmene refineR og TMC13 fikk så i oppgave å finne referansegrenser for p-ferritin basert på denne samlingen av kunstige data fra kvinner med og uten jernmangel. Stata (versjon 16.1) ble brukt til å lage kunstige data og til grafisk

framstilling av deres fordelinger.

Pasientenes identitet var ukjent for oss. Studien var godkjent av Regionale komiteer for medisinsk og helsefaglig forskningsetikk under søknadsnummer 155292, som gir oss anledning til å bruke produksjonsdata til beregning av referansegrenser.

Resultater

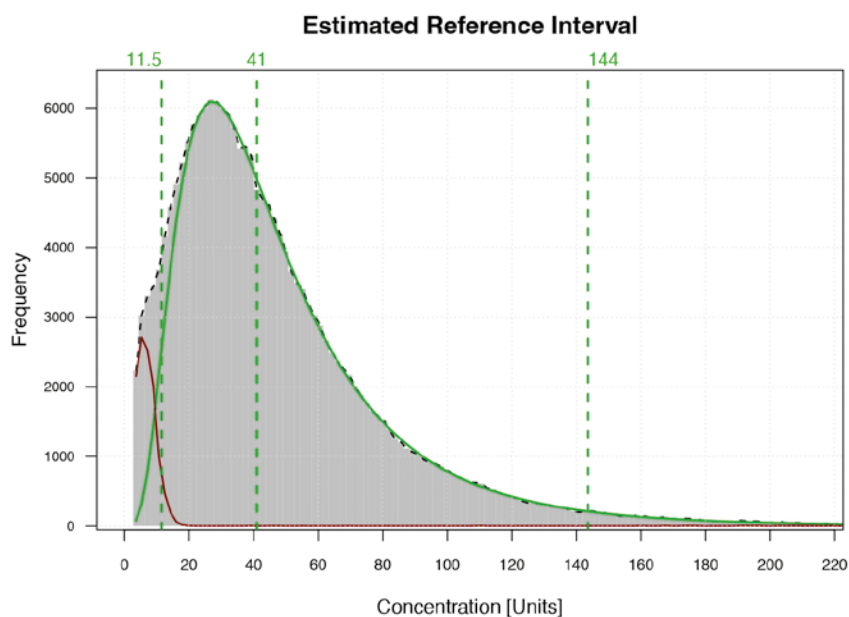
Basert på produksjonsdata estimerte refineR referansegrensene 8,53 µg/L (95 % konfidensintervall 7,84-11,1) og 181 µg/L (95 % konfidensintervall 174-209), mens medianen var 50,7 µg/L. TMC13 kom ut med grensene 8,71 µg/L (95 % konfidensintervall 8,40-9,02) og 180 µg/L (95 % konfidensintervall 177-184). Ingen median var oppgitt. I utskriften fra TMC13 var det et varsel om at estimeringen kanskje ikke var optimal. Basert på fordelingsparametre hos de 105 kvinnene med jern i benmargen var referansegrensene 12,2 og 145 µg/L.

Fordelingen av de kunstige data er framstilt i figur 1. For kvinner uten jernmangel ➤

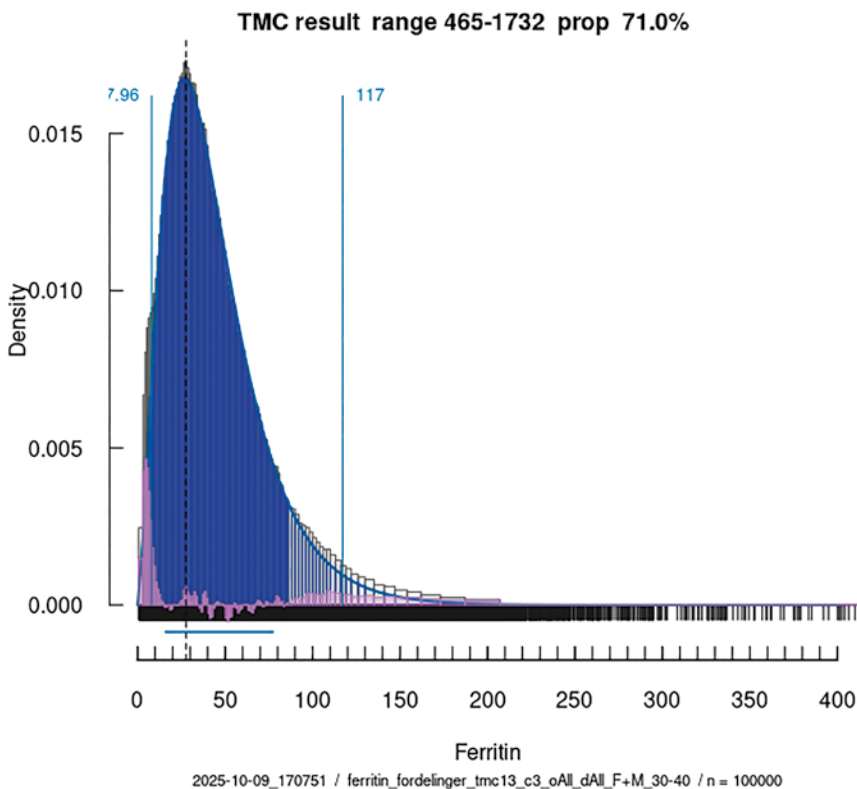
var 2,5- og 97,5-persentilen henholdsvis 12,1 og 147 $\mu\text{g/L}$. Andelen uten jernmangel var 90 %. Basert på de samme data estimerte refineR referansegrensene 11,5 og 144 $\mu\text{g/L}$ (figur 2), mens andelen ikke-patologiske verdier ble estimert til 92,4 %. TMC13 kom fram til referansegrensene 8,0 og 117 $\mu\text{g/L}$ (figur 3) og en andel ikke-patologiske verdier på 93,8 %.

Diskusjon

Vi valgte ut produksjonsdata fra kvinner i 30-årsalderen fordi vi hadde kjennskap til fordelingen av p-ferritin hos friske kvinner på 38 år (7), og kunne kontrollere resultatene mot referansegrenser utregnet fra disse. Basert på produksjonsdata var programmene refineR og TMC13 nokså samstemte om referansegrensene for p-ferritin. Punktestimatene fra refineR var innenfor 95 % konfidensintervallene for TMC13 sine grenser. De nedre referansegrensene (8,5 og 8,7 $\mu\text{g/L}$) var litt høyere enn median nedre referansegrense for premenopausale kvinner (6 $\mu\text{g/L}$) i en oversiktsartikkel fra 2024 (9), men vesentlig lavere enn den (12,2 $\mu\text{g/L}$) som ble estimert fra fordelingsparametre for kvinner uten jernmangel (7). Det kan hevdes at p-ferritin-metoden til Hallberg et al. (7) og Abbott Alinity-metoden ikke var likt kalibert, og det kan ikke motbevise, for prøvene til Hallberg et al. kan ikke reanalyseres med Abbott Alinity-metoden. Median p-ferritin for kvinner uten jernmangel var 42 $\mu\text{g/L}$ i Hallberg-materialet og 50,7 $\mu\text{g/L}$ i den «friske» delen av produksjonsdata. Det indikerer at Abbott Alinity-metoden måler litt høyere, så ulik kalibrering kan ikke forklare at de lave referansegrensene estimeres lavere med de indirekte metodene som brukte Abbott Alinity-resultater. Ulik kalibrering kan uansett ikke forklare at referanseområdene er bredere med estimatene fra produksjonsdata. Kanskje kvinnene i Hallberg-materialet var ekstra friske og at de bredere referanseintervallene basert på produksjonsdata er et bedre uttrykk for fordelingen av p-ferritin hos friske kvinner i 30-årsalderen. Det vet vi ikke. Benmargsundersøkelser av friske kvin-



FIGUR 2: Kopi av figur laget av refineR etter at programmet hadde analysert det kunstige datasettet som er beskrevet i figur 1. Rød kurve: Jernmangel. Grønn kurve: Ikke jernmangel. De grønne tallene over figuren angir 2,5-, 50- og 97,5-persentilene i den grønne fordelingen.



FIGUR 3: Kopi av figur laget av TMC13 etter at programmet hadde analysert det kunstige datasettet som er beskrevet i figur 1. Patologiske verdier er markert med en lys lilla farge. De blå tallene øverst i figuren er de estimerte referansegrensene.

ner er heldigvis sjeldne. I en oversiktsartikkel fra 2018 nevnes bare Hallberg-materialet og tre langt mindre (10). En svakhet ved Hallberg-materialet er likevel størrelsen, det vil si at fordelingsparametrene for p-ferritin baseres på bare 69 og 105 kvinner med og uten jernmangel. Dessuten kan manglende aldersvariasjon gjøre at materialet blir mindre representativt for alle kvinner i 30-årsalderen. Det er selv sagt også mulig at forutsetningen om at log(p-ferritin) er normalfordelt i Hallberg-materialet ikke holder. I mangel av rådata kan det ikke kontrolleres.

Da vi prøvde refineR og TMC13 med det kunstige datasettet, kom refineR rimelig godt fra det. De estimerte referansegrensene var ikke langt fra fasit, henholdsvis 11,5 mot 12,1 µg/L og 144 mot 147 µg/L. Sammenligner vi figur 2 (refineR) med figur 1 (fasit), er fordelingen av p-ferritin fra individer uten jernmangel (grønne kurver) ganske lik i de to figurene, mens fordelingen av p-ferritin fra de med jernmangel (røde kurver) er litt for smal i figur 2. TMC13 greide ikke prøven så bra som refineR, med estimerte referansegrenser på 8,0 og 117 µg/L. Programmet fant også noen patologisk høye verdier, se figur 3. På forhånd hadde vi trodd at det kunstige datasettet med bare to veldefinerte fordelinger skulle være en lett oppgave for både refineR og TMC13.

Vi er fortsatt usikre på hvilke referansegrenser for p-ferritin som er de riktige for kvinner i 30-årsalderen. Det er ingen stor ulykke, for legene bør ikke bruke referansegrensene til diagnostikk av verken jernmangel eller jernoverskudd. For å stille en diagnose ved hjelp av en sykdomsmarkør må man kjenne fordelingen av markøren hos pasienter både med og uten sykdommen. De to delpopulasjonene må til sammen utgjøre en klinisk relevant populasjon. Når vi skal finne den riktige beslutningsgrensen, må vi også ta hensyn til pretestsannsynlighet og de kliniske konsekvensene av diagnosen (11). I den forbindelse har referansegrenser ingen selvstendig betydning. Valgte vi nedre referansegrense i figur 1 (12,1 µg/L) som beslutningsgrense for jernmangel, ville

spesifisiteten bli 97,5 %, men sensitiviteten ville bare være 66,1 %. Det finnes ingen grunn til å velge akkurat den verdien som beslutningsgrense. Selv for pasienter uten betennelse er riktige beslutningsgrenser for p-ferritin i diagnostikk av jernmangel høyere enn nedre referansegrense (9, 10). For pasienter med betennelse er riktige beslutningsgrenser høyere enn for pasienter uten betennelse (10).

Konklusjon: Referansegrenser lar seg lett estimere med indirekte metoder, men riktigheten av grensene er vanskelig å kontrollere. Uansett må referansegrenser ikke oppfattes som riktige beslutningsgrenser. ■

Forfatternes bidrag

Arne Åsberg: Idé, statistiske analyser, utkast til artikkel.

Øyvind Skadberg: Idé, uthenting av data fra laboratoriedatasystemet, utforming av artikkelteksten.

Interessekonflikter

Ingen.

Referanser

1. Dirks NF, den Elzen WPJ, Hillebrand JJ, Jansen HI, Boekel ET, Brinkman J, et al. Should we depend on reference intervals from manufacturer package inserts? Comparing TSH and FT4 reference intervals from four manufacturers with results from modern indirect methods and the direct method. *Clin Chem Lab Med.* 2024;62:1352-61.
2. Åsberg A, Mikkelsen G, Løfblad L. CgA100 – eGFR-adjusted serum chromogranin A. *Scand J Clin Lab Invest.* 2025;85:133-7.
3. Bohn MK, Nyholt D, Balion C, Cembrowski G, Collier C, Guire V, et al. Best practice guidelines on reference interval harmonization in Canada: Evidence-based recommendations from the CSCC working group on reference interval harmonization (CSCC WG-hRI). *Clin Biochem.* 2025;139:110986.
4. Ammer T, Schützenmeister A, Rank CM, Doyle K. Estimation of reference intervals from routine data using the refineR algorithm-a practical guide. *J Appl Lab Med.* 2023;8:84-91.
5. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. *Clin Chem Lab Med.* 2019;57:1933-47.

6. Braga F, Pasqualetti S, Frusciante E, Borrillo F, Chibireva M, Panteghini M. Harmonization status of serum ferritin measurements and implications for use as marker of iron-related disorders. *Clin Chem.* 2022;68:1202-10.
7. Hallberg L, Bengtsson C, Lapidus L, Lindstedt G, Lundberg PA, Hultén L. Screening for iron deficiency: an analysis based on bone-marrow examinations and serum ferritin determinations in a population sample of women. *Br J Haematol.* 1993;85:787-98.
8. Haeckel R, Wosniok W. Observed, unknown distributions of clinical chemical quantities should be considered to be log-normal: a proposal. *Clin Chem Lab Med.* 2010;48:1393-6.
9. Truong J, Naveed K, Beriault D, Lightfoot D, Fralick M, Sholzberg M. The origin of ferritin reference intervals: a systematic review. *Lancet Haematol.* 2024;11:e530-9.
10. Garcia-Casal MN, Pasricha SR, Martinez RX, Lopez-Perez L, Peña-Rosas JP. Are current serum and plasma ferritin cut-offs for iron deficiency and overload accurate and reflecting iron status? A Systematic Review. *Arch Med Res.* 2018;49:405-17.
11. Åsberg A, Bolann BJ. The optimal cut-off value. *Clin Chim Acta.* 2025;565:119953.

Om forfatterne



Arne Åsberg er spesialist i medisinsk biokjemi, pensjonist, men jobber deltid som overlege ved Avdeling for medisinsk biokjemi på Stavanger universitetssjukehus.



Øyvind Skadberg er spesialist i medisinsk biokjemi og jobber som avdelingsoverlege ved Avdeling for medisinsk biokjemi på Stavanger universitetssjukehus. Han er også overlege i bistilling ved Nasjonalt kompetansesenter for porfyri-sykdommer (NAPOS), Haukeland universitetssjukehus.