

Svangerskapsdatering

Håkon K. Gjessing

*Senter for fruktbarhet og helse, Folkehelseinstituttet
Institutt for global helse og samfunnsmedisin, Universitetet i Bergen*
E-post: hakon.gjessing@uib.no Telefon +47 21078241 / +47 97522262

SAMMENDRAG

Svangerskapsdatering er et komplekst felt med mange aspekter. Minst tre svært forskjellige metoder for datering er i bruk: siste menstruasjon, ultralyd samt tidspunkt for fertilisering eller overføring av embryo ved in vitro-fertilisering. Bare innenfor ultralyddatering finnes en stor mengde forskjellige “formler” for datering. Jeg gir her en kort oversikt over de forskjellige tilnærmingene til svangerskapsdatering og viser noen av prinsippene bak populasjonsbasert ultralyddatering.

Gjessing HK. **Pregnancy dating.** *Nor J Epidemiol* 2017; 27 (1-2): 13-18.

ENGLISH ABSTRACT

Pregnancy dating is a complex and multifaceted field. At least three very different methods for dating are commonly used: last menstrual period, ultrasound, and date of fertilization or embryo transfer during in vitro fertilization. Within ultrasound dating alone there is a wide range of different “formulas” for dating. I present a short overview over the different approaches to pregnancy dating and show some of the principles behind population-based ultrasound dating.

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUKSJON

Det er en lang tradisjon for bruk av matematiske og statistiske modeller i svangerskapsprediksjon. Den (angivelig) kompetente matematikeren Sun-Tsu presenterte, rundt første århundre av vår tidsregning, følgende prediksjonsproblem med tilhørende løsning¹:

A pregnant woman, who is 29 years of age, is expected to give birth to a child in the 9th month of the year. Which shall be her child, a son or a daughter?

Take 49; add the month of her child-bearing; subtract her age. From what remains, subtract the heaven 1, subtract the earth 2, subtract the man 3, subtract the four seasons 4, subtract the five elements 5, subtract the six laws 6, subtract the seven stars 7, subtract the eight winds 8, subtract the nine provinces 9. If the remainder be odd, the child shall be a son; and if even, a daughter.

Vi får bare håpe at det var den tids jordmødre som tok seg av dateringen.

Bruken av statistiske modeller har kommet et stykke videre siden den gang, men svangerskapsdatering er et tema som er langt mer komplekst enn man kanskje i utgangspunktet skulle forvente. Minst tre svært forskjellige metoder for datering er i vanlig bruk: siste menstruasjon (last menstrual period, LMP), ultralyd (ultrasound, US) og bruk av tidspunkt for fertilisering eller overføring av embryo ved in vitro-fertilisering (IVF). Disse representerer tre helt forskjellige måter å definere starten av svangerskapet, og bare blant ultralydmetoder finnes en hærskere forskjellige modeller.

Så snart “starten” av svangerskapet er bestemt, regnes fosteralder greit nok som tid siden det estimerte

starttidspunktet. Men fosteralder er åpenbart knyttet til fødselstermin. Hvordan henger alder og termin sammen? Og hvordan definerer man egentlig en “termin”? Til tross for en omfattende internasjonal litteratur er ikke begrepet “termin” spesielt godt operasjonalisert, og man ser sjeldent spesifikke kriterier for en “god” terminprediksjon.

I det følgende vil jeg gi en kort og forenklet gjennomgang av metoder for svangerskapsdatering og noen av de utfordringene man møter. Jeg vil også beskrive noen av forskjellene mellom en full populasjonsbasert ultralydmetode (eSnurra²) og tradisjonelle utvalgsbaserte ultralydmetoder (gamle Snurra³ og Terminhjulet⁴) som alle har vært benyttet i Norge.

SVANGERSKAPETS ENDEPUNKTER

Svangerskapets start

Definisjonen av svangerskapets start er kanskje like mye et praktisk valg som et biologisk. Rent biologisk ville nok fertiliseringstidspunktet vært det mest logiske valget. Men for de fleste svangerskap er dette naturligvis ukjent. Dessuten vil det være såpass mye individuell variasjon i celledeling, implantering og videre fosterutvikling at det er vanskelig å definere et riktig tidspunkt for starten av den “faktiske” fostervekten, og uansett ville dette tidspunktet også være ukjent for de fleste graviditeter. Så den eksisterende definisjonen med siste menstruasjons første blødningsdag er nok den eneste som er praktisk mulig.

Et åpenbart problem med bruk av LMP er at moren må vite datoen. Dessuten vil follikulær fase lengde variere betydelig individuelt⁵, så LMP vil i noen grad

Figur 1. Gestasjonsalder regnet i uker fra siste menstruasjon (LMP) mot biparietal diameter i millimeter. Et typisk uniformt utvalg av målinger fra populasjonen, med ca. 10 målinger per dag gjennom svangerskapet. Lett renset ved fjerning av “outliers”, altså spesielt avvikende målinger.

Dette var åpenbart nyttig for mødre med usikre eller ikke-eksisterende LMP-verdier. Men for mødre med sikker LMP fikk man plutselig to forskjellige alders- og terminestimer, og dette åpnet naturligvis for diskusjoner om hvilken som er “rett”. Denne diskusjonen har mange viktige aspekter og er fortsatt relevant, men fra et rent prediksjonsperspektiv er det klart at man trenger et utvetydig kriterium for hva en god prediksjonsmodell er.

Kriterium for god datering

Forbausende nok er det sjeldent man ser en klar definisjon på hva en god prediksjonsmodell betyr. Stort sett konstrueres klassiske prediksjonsmodeller på et utvalgsmateriale. Deretter benyttes de klinisk i en populasjonssammenheng uten noen videre tanker angående treffsikkerhet. Fra et vitenskapelig synspunkt er dette neppe tilstrekkelig, men kun unntaksvis settes det faktisk opp en populasjonstest av metoden, og da ofte som en ettertanke, lenge etter at metoden er tatt i bruk.

I praksis er det eneste brukbare populasjonskriteriet hvor godt metoden treffer i prediksjon av termin. Som beskrevet over kan man benytte det faktiske fødsels-tidspunktet i populasjonen, korrigert for ikke-spontane fødsler. Et kriterium for en god prediksjonsmodell er altså at *metodens terminberegning skal treffe så korrekt som mulig på det naturlige, spontane fødselstidspunkt, med minimalt bias (systematisk feil) og så lite spredning som mulig*. Det er rimelig å forlange at dette skal gjelde både i totalfordelingen og når man ser over hele spennet av tillatte ultralydverdier, f.eks. alle BPD-verdier fra 24-60 mm.

Det er vanskelig å tenke seg et mer anvendelig kriterium i en populasjonssammenheng. Dette er et

statistisk veldefinert kriterium, og fra et biologisk perspektiv er det vanskelig å se et mer naturlig mål for fosterets modenhet enn det spontane fødselstidspunktet. I de tradisjonelle ultralydmetodene avledes termin direkte av alder. Derfor reflekterer kvaliteten i terminprediksjon også eventuelle problemer med aldersberegning. Av grunner vi skal se nedenfor er det dessuten i praksis nesten umulig å sette opp andre populasjonsbaserte kriterier.

Ultralyd versus siste menstruasjon

Med dette kriteriet må nok ultralyd sies å være en klar vinner over siste menstruasjon: US-estimer av termin har klart mindre spredning rundt faktisk fødsel enn det LMP-estimer har. Presisjonen er altså klart bedre⁷. Dette lyder paradoksalt: Ultralyd brukes her til å estimere LMP. Når sikker LMP likevel er tilgjengelig, hvorfor virker datering bedre når den er basert på ultralyd? Dette skyldes at ultralydestimatet er nærmere knyttet til den “sanne” starten av fostervevst. Variasjon i follikulær fase er såpass stor at den overskygger variasjonen i fostervekst fra fertilisering frem til ultralydmåling. Dermed blir ultralydestimatene av termin mer presise enn LMP-estimatene.

Det er ikke dermed sagt at ultralyddatering er å foretrekke i enhver sammenheng (se senere), men i en klinisk sammenheng gir de klart bedre presisjon, og man unngår de meget usikre dateringene man noen ganger ser med LMP.

Kart versus landskap. Populasjonsbasert datering

Siden enhver dateringsmetode til syvende og sist skal benyttes i en populasjonssammenheng er det åpenbart at et avgjørende kriterium for kvalitet er hvordan metoden oppfører seg anvendt på kliniske populasjonsdata. Dette står sentralt i utviklingen av eSnurra: når kriteriet for god prediksjon er hvordan den predikerer (median) termin i en populasjon er det mest naturlig å utvikle modellen slik at den *faktisk optimaliseres* til dette. Til syvende og sist er det kartet som må stemme med landskapet, ikke motsatt.

Basert på dette ble eSnurra laget ved å predikere *gjenværende tid av svangerskapet* utfra BPD, altså tid fra målingen ble gjort til svangerskapet endte spontant. Dette ble gjort på et stort populasjonsmateriale som er godt representativt for store deler av Norge. Evalueringen av eSnurra på tre store geografiske regioner i Norge viste tydelig at den populasjonsbaserte modellen treffer med veldig lite bias, mens de tradisjonelle utvalgsmateriale modellene viser bias i de samme testene^{10,11}.

Kilde til bias i aldersberegning

Hva er så kilden til bias i terminberegningene fra de tradisjonelle modellene? Feilkilden ligger allerede i aldersberegningen. Problemet er statistisk: Som vist i Figur 1 antar man i utvikling av metoden at målingene gjøres jevnt fordelt på alder gjennom hele svangerskapet. Imidlertid anvendes disse metodene i ettertid

på en populasjon som helt systematisk kalles inn til ultralydundersøkelser rundt uke 18. I praksis har dermed populasjonsfordelingen av undersøkelser en helt annen fordeling enn den kunstige uniforme fordelingen som legges til grunn i utvikling av de tradisjonelle metodene. Dette vises i Figur 2, med populasjonsdata fra Trondheimsområdet². Utvelgingen gjøres altså på LMP-alder, som er Y-variabelen, eller den avhengige variabelen, i regresjonen. Dette er kilden til et bias i *aldersberegning* i de tradisjonelle modellene, og dette reflekteres videre i et bias i terminberegningen.

Tilbake til alder

I den populasjonsbaserte modellen beregnes altså terminen først, direkte i populasjonsdata. For å predikere alder fra samme modell regnes så alder = 283 dager – estimert tid til termin. Dette skulle tilsynelatende sette aldersestimatet i annen rekke i eSnurra. Fordelen med denne beregningsmåten er imidlertid at den fullstendig omgår seleksjonsproblematikken beskrevet ovenfor. Prediksjonsmodellen benytter seg ikke av LMP-alder i populasjonsdata da disse har et samplingsfokus rundt uke 18. Derimot benytter den fødselstidspunkt, som er uaffisert av denne samplingen.

Med denne metoden unngår man altså bias i aldersberegningen. Det viser seg også at aldersestimatet fra eSnurra er det estimatet som samsvarer best med LMP-aldersestimatet i populasjonen, selv om de tradisjonelle modellene faktisk tar utgangspunkt i LMP i konstruksjonen¹².

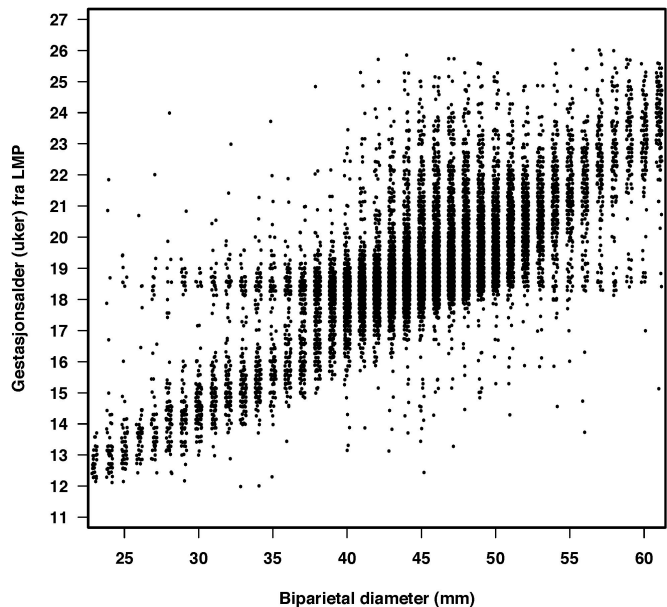
Interessant nok kan man teste de to utviklingsstrategiene på uniformt fordelte data, og finner da at de gir tilnærmet identiske estimater for termin og alder. Forskjellen oppstår først når populasjonsdata ikke er uniforme, men innkalt rundt en LMP-alder, per i dag uke 18. Da oppstår et bias i metodene som først estimerer alder og avleder termin fra denne.

IVF-SVANGERSKAP

In vitro-fertiliserte (IVF) svangerskap er dateringsmetodenes “stebarn”. Deres betydning er flittig diskutert, og utallige IVF-studier, på varierende grunnlag, mener å kunne danne en “gullstandard” for datering siden tidspunkt for overføring av embryo er kjent. Stadig bedre IVF-registre gjør naturligvis slike studier mer interessante og relevante. Men her er det to spørsmål å ta hensyn til:

1. Er virkelig IVF-svangerskap en gullstandard for datering og fostervekst?
2. Og delvis uavhengig av dette, hvordan bør IVF-svangerskapene selv faktisk dateres i klinikken?

Når det gjelder det første spørsmålet vil det etter forfatterens mening være nesten umulig å fremheve IVF-svangerskap som en gullstandard for en hel populasjon. Det finnes mange forskjellige metoder for å utføre IVF. Mødrene er en meget selektert del av populasjonen. Embryo får en behandling som er helt



Figur 2. Gestasjonsalder regnet i uker fra siste menstruasjon (LMP) mot biparietal diameter i millimeter. Figuren viser uselekterte data fra populasjonen, med ca. 31 000 målepunkter i området. Man kan se utvalget av undersøkelser i uke 18 som en tydelig horisontal linje av datapunkter.

annerledes enn ved en naturlig fertilisering. Det er høyere andel flerfødsler, og mødrene får en annen oppfølging og andre rutiner for inngrep i svangerskapet. Det er naturligvis interessant å detaljstudere IVF-svangerskap for å kunne sammenligne dem med den øvrige populasjonen, men det er vanskelig å se dem som en gullstandard.

I datering av IVF-svangerskap virker svaret gitt: Man kjenner den eksakte dagen for overføring av embryo, og hva er mer naturlig enn å bruke dette til datering. Det er imidlertid en ytterligere komplikasjon i dette: IVF-dateringen teller fra fertilisering eller overføring av embryo, ikke fra LMP. Det er vanskelig å si nøyaktig hvor dette plasserer seg i forhold til den naturlige syklusen. For å kunne komme på samme skala som normalgraviditeter må man legge til – sånn omtrent – median lengde av follikulær fase. Av bekvemmelighetshensyn settes denne ofte til 14 dager, som er blitt vanlig praksis og sjeldent stilt spørsmål ved. Imidlertid kan man godt argumentere for at dette tallet skal settes noen dager annerledes, gjerne større. Uansett er det vanskelig å vite hva som er et “korrekt” tall å legge til for at dateringen skal bli parallell med vanlig ultralyddatering.

Bruk av ultralyddatering, derimot, følger samme standard for alle graviditeter, og fanger opp den faktiske tidlige fosterveksten, nettopp med tanke på å gi et godt estimat av når den egentlige veksten startet. Den forutsetter imidlertid at IVF-svangerskap viser samme vekst i første halvdel (frem til datering) som normale svangerskap, hvilket ikke nødvendigvis er riktig.

Det er derfor fordeler og ulemper med begge strate-

giene. Selv om bruk av overføringstidspunkt virker som den åpenbart rette løsningen for IVF-svangerskap har man likevel vanskeligheten med skala-plasseringen, som er avgjørende for at svangerskapsdateringen skal bli parallell til vanlige svangerskap.

ANDRE ASPEKTER

Jeg har her kun beskrevet noen viktige hovedprinsipper ved svangerskapsdatering. Det er en lang rekke andre elementer som også kommer i betraktning i det fulle bildet. Dette gjelder f.eks. diskusjonen rundt hvilken ultralydparameter som bør brukes som primærparameter i prediksjonsmodellen. Videre legges rutineundersøkelse med ultralyd i Norge i dag til rundt uke 18. Man kan se fordeler med å flytte denne til et tidligere tidspunkt, f.eks. rundt uke 12. Internasjonalt gjøres dette også i stor grad. En tidlig datering er også interessant ut fra en større mulighet til å gjøre vurdering av tidlig vekst. Både ved valg av beste ultralydparametre og det beste tidspunkt for rutineundersøkelse kan den populasjonsbaserte modellen i stor grad bidra til en enhetlig evaluering av disse valgene slik at metoden er tilpasset en norsk populasjon.

Ideelt sett kunne man sett for seg at dateringsmodeller ble utviklet spesielt med fokus på å redusere forekomst av sykkelighet og dødelighet hos barna. Kanskje kunne termin vært definert som f.eks. det optimale fødselstidspunktet i forhold til å unngå fosterdød. Dessverre er nok en slik strategi ikke en farbar vei. Den ville kreve svært store datamengder for å kunne oppnå rimelige presise estimater. Dessuten ville den medføre en komplisert diskusjon av hvilke dødsfall og uheldige utfall som faktisk kunne vært forebygget ved å endre datering tilsvarende. Det virker derfor rimelig at en dateringsmetode må representere den biologiske klokken i svangerskapet og måle hvor modent fosteret er. Så må anbefalinger om inngrep baseres på det faste rammeverket dateringen gir.

I klinisk bruk er det i stor grad akseptert at ultralyddatering er mer presis og stabil enn datering fra siste menstruasjon. En stor forskjell mellom US-termin og LMP-termin kan oppfattes som et varsel, men ultralyddateringen er den offisielle. I en epidemiologisk forskningssammenheng er det mindre åpenbart hvilken av dateringene som er den "korrekte". Siden eksponeringsvariable kan påvirke fosterstørrelse rundt uke 18 kan de også påvirke ultralyddatering på måter som blir misvisende¹³. Det kan derfor argumenteres for at i epidemiologiske populasjonsstudier kan man ofre

noe av presisjonen i ultralydestimatene til fordel for LMP-estimatene, som er uten systematiske bias i dette henseende.

Det har i de senere årene vært en stor innsats på å utvikle internasjonale standarder for fosterdatering og biometri^{14,15}. En slik innsats er naturligvis svært interessant og av stor verdi spesielt for land som ikke har tilstrekkelig gode nasjonale data. Det kan også se ut som at fosterstørrelse tidlig i svangerskapet er mindre betinget av etnisitet enn man kanskje skulle tro. Imidlertid er det liten grunn til å tro at dette arbeidet vil føre til full internasjonal konsensus. Prosjektene har heller ikke lagt opp til systematisk populasjonstesting av metodene men kun fokusert på utviklingen av dem.

OPPSUMMERING

The end of a pregnancy is easily determined, but its beginnings are obscure.

Allen J. Wilcox¹⁶

Som vi har sett er svangerskapsdatering et mer kompleks tema enn man skulle forvente. Starten av svangerskapet må defineres fra LMP, men vil i praksis ofte bli estimert fra ultralyd. Selv slutten av et svangerskap må defineres klart for å ha enighet om hva en "termin" er. I den internasjonale litteraturen har det vært liten tradisjon for å faktisk teste dateringsmodeller på populasjonen de er ment for, og når det er gjort er dette ofte som en ettertanke. Den populasjonsbaserte modellen for ultralyddatering har fordelene at den baserer seg på klare kvalitetskriterier definert i den populasjonen den faktisk skal benyttes på.

I alle eksisterende metoder er alder og termin koblet matematisk gjennom den totale lengden av et svangerskap, som i den populasjonsbaserte modellen er satt til median 283 dager. Ved å først estimere termin og deretter alder unngår den populasjonsbaserte modellen det bias som oppstår ved å først estimere alder.

Selv om ultralyddatering er standarden i klinisk praksis kan man likevel argumentere for å bruke LMP-datering i f.eks. epidemiologiske studier. Spesielt i studier av eksponering som kan påvirke både svangerskapslengde og fostervekst kan dette være naturlig. Faktorer som påvirker fosterstørrelse ved uke 18, f.eks. barnets kjønn, kan tilsvarende forstyrre en ultralyddatering som baseres på fosterstørrelse. Det kan derfor noen ganger være naturlig å ofre ultralydens presisjon til fordel for et redusert bias i LMP-basert datering, spesielt i store registerstudier.

REFERANSER

1. Beckmann P. A History of Pi. St. Martin's Press, 1971.
2. Gjessing HK, Grøttum P, Eik-Nes SH. A direct method for ultrasound prediction of day of delivery: a new, population-based approach. *Ultrasound Obstet Gynecol* 2007; **30**: 19-27.
3. Eik-Nes SH, Grøttum P. Graviditetskalenderen Snurra. 1983.
4. Johnsen SL, Rasmussen S, Sollien R, Kiserud T. Fetal age assessment based on ultrasound head biometry and the effect of maternal and fetal factors. *Acta Obstet Gynecol Scand* 2004; **83**: 716-23.

5. Baird DD, McConaughey DR, Weinberg CR, et al. Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. *Epidemiology* 1995; **6**: 547-50.
6. Aalen OO, Borgan Ø, Gjessing HK. Survival and event history analysis: a process point of view. Springer Verlag, 2008.
7. Tunón K, Eik-Nes SH, Grøttum P. A comparison between ultrasound and a reliable last menstrual period as predictors of the day of delivery in 15,000 examinations. *Ultrasound Obstet Gynecol* 1996; **8**: 178-85.
8. Smith GC. Use of time to event analysis to estimate the normal duration of human pregnancy. *Hum Reprod* 2001; **16**: 1497-500.
9. Altman DG, Chitty L. New charts for ultrasound dating of pregnancy. *Ultrasound Obstet Gynecol* 1997; **10**: 174-91.
10. Økland I, Gjessing HK, Grøttum P, Eik-Nes SH. Biases of traditional term prediction models: results from different sample-based models evaluated on 41 343 ultrasound examinations. *Ultrasound Obstet Gynecol* 2010; **36**: 728-34.
11. Økland I, Gjessing HK, Grøttum P, Eggebø TM, Eik-Nes SH. A new population-based term prediction model vs. two traditional sample-based models: validation on 9046 ultrasound examinations. *Ultrasound Obstet Gynecol* 2011; **37**: 207-13.
12. Økland I, Nakling J, Gjessing HK, Grøttum P, Eik-Nes SH. Advantages of the population-based approach to pregnancy dating: results from 23 020 ultrasound examinations. *Ultrasound Obstet Gynecol* 2012; **39**: 563-8.
13. Henriksen TB, Wilcox AJ, Hedegaard M, Secher NJ. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology* 1995; **6**: 533-7.
14. Papageorghiou AT, Kemp B, Stones W, et al. Ultrasound-based gestational-age estimation in late pregnancy: International late pregnancy dating. *Ultrasound Obstet Gynecol* 2016; **48**: 719-26.
15. Kiserud T, Piaggio G, Carroli G, et al. The World Health Organization Fetal Growth Charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLOS Med* 2017; **14**: e1002220.
16. Gjessing HK, Skjærven R, Wilcox AJ. Errors in gestational age: evidence of bleeding early in pregnancy. *Am J Public Health* 1999; **89**: 213-18.