



Spå eller forstå?

MEDISIN OG TALL

KATHRINE FREY FRØSLIE

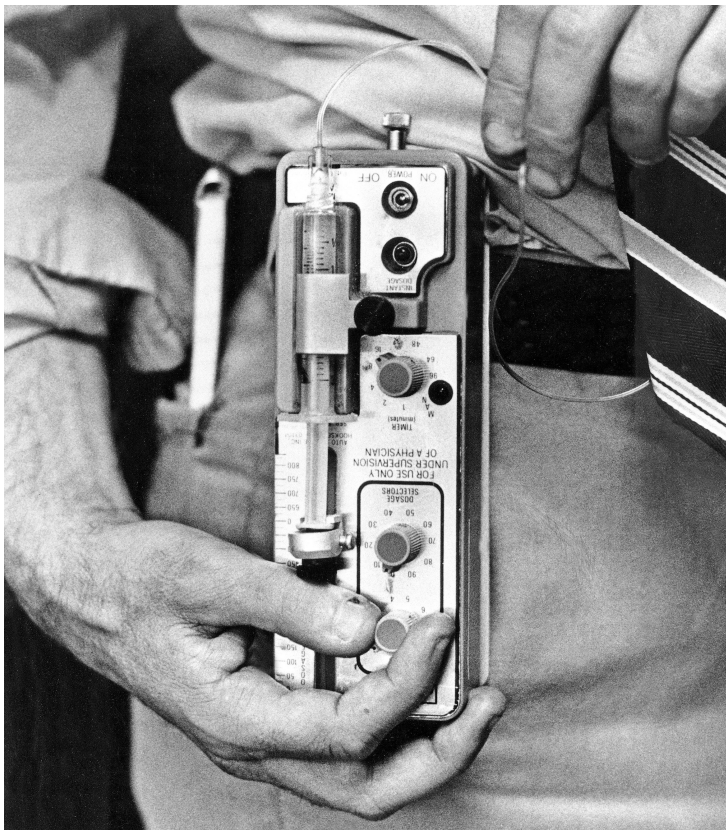
Kathrine Frey Frøslie (f. 1971) er statistiker ved Nasjonal kompetansetjeneste for kvinnehelse, Oslo universitetssykehus, Rikshospitalet. Hun har en ph.d.-grad i biostatistikk, er en erfaren foreleser og forskningsveileder og driver den populærvitenskapelige strikkebloggen Statistrikk.no. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

JO RØISLIEN

E-post: jo@joroislien.no

Jo Røislien (f. 1972) er professor i medisinsk statistikk ved Det helsevitenskapelige fakultet ved Universitetet i Stavanger og førsteamanuensis II ved Institutt for matematiske fag ved Norges teknisk-naturvitenskapelige universitet. Han er en profilert fagformidler og programleder på TV. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

For bare hundre år siden predikerte man en snarlig og smertefull død for dem med unormalt stor vannlating og urin som luktet eller smakte søtt. Dette var før man ante noe om mekanismene bak diabetes. Diabetes var en prediktor for død, men det var lite man kunne gjøre. Man visste rett og slett ikke nok om sykdommen.



Prediksjon i praksis: En av verdens første insulinpumper i bruk i 1986. Insulinpumper «vet» ingenting om mekanismene bak diabetes, men ved hjelp av målinger av den intracellulære væsken i underhuden injiserer de riktige mengder med insulin. Foto: Associated Press/NTB scanpix.

Et av hovedmålene med statistisk analyse er å avdekke sammenhenger mellom variabler, og et av de mest allsidige statistikkverktøyene vi har er regresjonsanalyse.

Med regresjon kan vi finne ut om en variabel kan brukes til å predikere en annen. Sukker i urinen predikerer død, været i dag predikerer været i morgen og postnummer predikerer skoleresultater.

Søt urin – en prediktor

En statistisk prediksjonsmodell er demokratisk og inkluderende. Alle variabler kan være en prediktor: urinlukt, været, bosted eller hva som helst. Og den tilhørende matematiske ligningen kan være så komplisert som bare det. Det eneste som teller er om vi klarer å spå korrekt om fremtiden med det vi har av informasjon.

Hele diskusjonen rundt stordata (big data), datautvinning (data mining) og maskinlæring (machine learning) handler om å tråle store datamengder etter statistiske sammenhenger og så bruke disse sammenhengene for å lage treffsikre reklamer på Facebook-siden din eller finne ut hvilken genetisk subgruppe som har nytte av den nye, dyre kreftmedisinen. Det hele har en aura av science fiction over seg, men det er ofte bare helt vanlig regresjonsanalyse.

Selv uten kunnskap om de bakenforliggende mekanismene er prediksjonsmodeller nyttige. En kvinne som føder et barn på over 4 500 g havner i høyrisikogruppen for senere å få type 2-diabetes, rett og slett fordi det er høy forekomst av type 2-diabetes hos kvinner som har født store barn. Dette er nyttig kunnskap fordi det forteller kvinnen at hun bør sjekke blodsukkernivået sitt oftere enn hun kanskje ellers ville gjort.

Men det å føde et stort barn fører ikke til at man får type 2-diabetes. Det er en sammenheng, men det er ikke en årsakssammenheng. Uten kjennskap til årsakssammenhenger kan vi som regel ikke intervensere. Og vi kan ikke fortelle kvinnen hvordan hun kan unngå å utvikle type 2-diabetes. Til det trenger vi å vite noe om mekanismene bak diabetes.

En forklaringsvariabel må kunne forklares

For diabetes kom vendepunktet da det på begynnelsen av 1900-tallet ble avdekket hvilken rolle pancreas og insulinet spilte. Med kunnskap om viktigheten av insulinregulering kunne legene gripe inn og påvirke overlevelsen. «God eller dårlig insulinregulering» er noe mer enn en prediktor for død, slik som «søt urin». Den er en *forklaringsvariabel* for død, fordi den inneholder informasjon om mekanismene bak sykdommen.

I 1954 postulerte den danske legen Jørgen Pedersen (1914–78) at kvinner med diabetes føder store barn fordi de overfører fosteret med sukker under graviditeten (1). I 2008 ble Pedersen-hypotesen utvidet til også å gjelde ikke-diabetiske mødre, da man ved regresjonsanalyse fant en tilnærmet lineær sammenheng mellom kvinnens blodsukkernivå og barnets fødselsvekt (2). Ved hjelp av nitid forskning vet vi nå at det ikke bare er en sammenheng mellom kvinnens blodsukkernivå og fødselsutfall som høy fødselsvekt og hypoglykemi hos den nyfødte – det er en årsakssammenheng.

I jakten på årsakssammenhengen mellom kvinnens blodsukkernivå og fødselskomplikasjoner må vi bruke all vår fysiologiske og kliniske ekspertkunnskap og vurdere hvilke variabler som er aktuelle å ha med i de tilhørende regresjonsmodellene.

I motsetning til det som er tilfellet for en prediksjonsmodell, kan vi i en regresjonsmodell der vi jakter på mekanismer og kausalitet ikke velge fritt hva vi inkluderer. Mens en prediktor kan være hva som helst, må forklaringsvariablene være en del av den antatte kausale kjeden, og vi må spesifisere hva som er hovedeksponering, konfunderende

variabler, mediatorer eller variabler som påvirkes kausalt av minst to variabler i den kausale kjeden (collider). I motsetning til i en prediksjonsmodell er det også viktig at det matematiske uttrykket som gir oss estimater for sammenhengene er enkelt nok. Det hjelper lite med en forklaringsmodell man ikke forstår.

Å predikere er ikke å forklare

Regresjonsanalyse er et allsidig statistisk verktøy – så allsidig at det kan brukes til fundamentalt forskjellige ting. Hvorvidt formålet med den statistiske analysen er å predikere et utfall eller å estimere antatt kausale sammenhenger, er det de samme statistiske regresjonsmodellene som brukes. Det er kun ordene vi bruker som forteller leseren – og oss selv – hvor i kunnskapskjeden vi befinner oss.

I medisinsk forskning er det vanlig å bruke ordet prediktor om enhver uavhengig variabel i en statistisk regresjonsmodell. Men denne ordbruken visker ut skillet mellom to prinsipielt ulike måter å forholde seg til medisin og helse på. Det er forskjell på sammenheng og årsakssammenheng. Å spå er ikke å forstå. Å predikere er ikke å forklare.

LITTERATUR:

1. Pedersen J. Weight and length at birth of infants of diabetic mothers. *Acta Endocrinol (Copenh)* 1954; 16: 330 - 42. [PubMed]
2. Metzger BE, Lowe LP, Dyer AR et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008; 358: 1991 - 2002. [PubMed][CrossRef]

Publisert: 28. mai 2018. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.18.0086

Mottatt 23.1.2018, godkjent 19.3.2018.

© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no