



Statistisk modellering

Experterna på Aktuellt och i sjukvården använder ofta uttrycket "modellering" i diskussioner runt hur covid-19-epidemin kommer att utveckla sig i Sverige och världen. Vad betyder egentligen "statistisk modellering" – verkligheten är väl som den är och varför skall den då "modelleras"? Det känns som att man smyger in ett moment av subjektivitet och omformar verklighet till en statistisk modell. I nedanstående artikel reder statistiker Anna Törner ut begreppet "modellering" och varför statistiska modeller är viktiga för att förstå verkligheten.

Modellering för att kunna säga något om framtiden ...

Först är det viktigt att förstå att har man ett "dataset" – till exempel information om antalet covid-19-sjuka i Stockholm vid en given tidpunkt – så kan man inte säga något om hur det kommer att se ut om en vecka eller två veckor om man inte är villig att sätta upp en statistisk modell över smittspridningen och göra olika antaganden om hur epidemin kommer att utvecklas. All statistisk modellering bygger på olika antaganden. För covid-19-pandemin handlar det mycket om att planera för sjukhusvård och samhällsåtgärder de närmaste veckorna, men också att försöka förstå vilka preventiva åtgärder som är lämpliga.

Modellen behöver inte vara komplicerad, men utan antaganden kan vi inte säga något om hur framtiden kommer att se ut. När det gäller covid-19 är vi såklart intresserade av att deskriptivt ha koll på läget, det vill säga hur många som smittas, hur många som vårdas på sjukhus och hur många som dör, och så vidare. Men, framför allt är vi intresserade av att kunna hantera situationen framåt. Just därför behöver vi statistisk modellering, för att kunna predicera utvecklingen framåt. Vi är också intresserade av att förstå hur olika preventiva åtgärder, till exempel social distansering och stängning av delar av skolväsendet påverkar smittspridningen. Detta för att förstå vilka åtgärder som är lämpliga och när olika restriktioner kan lyftas bort.

Reproduktionstalet – hur mycket smittar en infektion?

För infektionssjukdomar gäller att utan åtgärder smittar varje infekterad person ett antal nya personer, för covid-19 har det så kallade reproduktionstalet R i början av epidemin angivits till mellan 2 och 3 i media. En förutsättning för att en epidemi skall kunna etablera sig överhuvudtaget är att varje smittad i genomsnitt infekterar fler än en person. Reproduktionstalet behöver naturligtvis inte vara ett heltal även om det i verkligheten faktisk är så att varje person smittar just 0, 1, 2 ... eller flera personer. Hur många varje

person smittar beror på många olika faktorer, till exempel smittsamhet just för den individen och hur det sociala kontaktmönstret ser ut. Reproduktionstalet förändras också över tid, dels som en konsekvens av olika åtgärder för att begränsa smitta, dels när en större del av befolkningen har blivit immuna. Det betyder att när vi uppnått det som kallas flockimmunitet, det vill säga när en stor andel av befolkningen har haft infektionen och inte längre är mottagliga, då avtar epidemin eftersom smittan inte längre har samma möjligheter att sprida sig. Olika åtgärder som social distansering och stängning av arbetsplatser och skolor påverkar också reproduktionstalet så att det blir lägre.

De som följt nyhetssändningar de senaste veckorna har alla hört om dag 0 och också mått på hur snabbt antal fall dubblats i olika länder. Dubblingstakten är naturligtvis kopplad till hur många nya personer varje smittad infekterar, men också hur fort detta sker. För en utveckling av antal fall över tid, och för att kunna jämföra länder behövs också en tydlig startpunkt. Det finns säkert fler orsaker till att man valt dagen för första dödsfallet till startpunkt, men en är säkert att det är en dag som går att fastställa med ganska stor säkerhet och som också antyder att epidemin nått en viss mognadsgrad i det aktuella landet. Modeller över hur infektioner sprider sig kan förstås göras hur komplicerade som helst, men detta är viktiga grundbegrepp.

- Några patienter blir friska nästan omedelbart, medan andra har ingen effekt alls av att få C-vitamin ("allt eller inget"-effekt).

- Hälften av patienterna får en förkortning av sin förkylningsepisod på mellan 3–5 dagar, övriga patienter responderar inte alls (hälften har effekten men storleken varierar mellan individer).

Det finns fler möjligheter! Poängen är att utan underliggande antaganden om en statistisk modell kan vi inte dra några slutsatser som helst. Statistiska modeller är inget hokus pokus utan enkla redskap för att göra verkligheten tolkningsbar.

Den kliniska studien – en modell av verkligheten

Av förståeliga skäl kan vi inte inkludera ett oändligt antal patienter i kliniska studier, ibland visserligen tusentals, men oftast betydligt färre, kanske hundratals, för godkännande av ett nytt läkemedel. Om vi skall vara väldigt krassa så handlar en klinisk studie om att utvärdera effekten av en ny behandling för det oändliga antal patienter som inte deltog i den kliniska studien – vi vill extrapolera resultaten från den kliniska studien för att kunna säga något meningsfullt om effekten av en ny behandling. Den kliniska

– för att predicera och tolka!

Statistik modellering i kliniska studier

Jag tror vi alla förstår att det behövs statistiska modeller för att predicera utvecklingen av en epidemi och vidta lämpliga åtgärder. Faktum är dock att för nästan allt vi studerar och där vi vill dra slutsatser utöver de faktiska data vi observerar, så behövs någon form av modell. Den kanske absolut enklaste statistiska modell vi använder är att vi antar ofta att observationer (effekter) är normalfördelade. Om vi till exempel drar slutsatsen att en studie visar att behandling med C-vitamin förkortar en förkylning med i genomsnitt två dagar, så har vi en intuitiv förståelse att på individuell basis kan behandling ge både mindre och större förkortning av sjukdomsförloppet, men att den "typiska" eller "genomsnittliga" patienten kan förvänta sig att tillfriskna två dagar snabbare om de får behandling med C-vitamin. Vi tänker så här eftersom vi vet vad genomsnitt och normalfördelning är och att just de här begreppen är djupt förankrade i vår vetenskapliga själ. Hade vi inte haft den bakgrundskunskapen så skulle alternativa tolkningar kunna vara:

- Behandlingen förkortar sjukdomsförloppet med exakt 2 dagar för alla patienter (deterministisk eller absolut effekt).

studien tjänar som en "proxy" för att vi skall kunna uttala oss om den sanna effekten av en ny behandling – om den ges till alla tänkbara patienter med den aktuella diagnosen. Statistiska modeller är med andra ord inget trick eller fusk där vi omformar verkligheten, utan enkla redskap för att göra verkligheten förståelig och för att vi skall kunna tänka bortom endast de få patienter vi har data för, oavsett om det gäller hur en epidemi utvecklas eller hur vi skall tolka resultat från en klinisk studie.

Statistisk modellering handlar ibland i vanrykte – intrycket är ju att man kan bevisa vad som helst med statistik. Och visst är det så; gör man helt fel antaganden och väljer en modell som inte beskriver data på ett bra sätt ligger vägen öppen för feltolkningar och galna slutsatser. För statistik gäller som för alla andra komplexa discipliner – det krävs expertkunskaper för att göra själva arbetet. När sedan resultat och preciseringar föreligger är tolkningen ofta rättfram och enkel.

ANNA TÖRNER
Statistiker och verkställande direktör
Scandinavian Development Services

