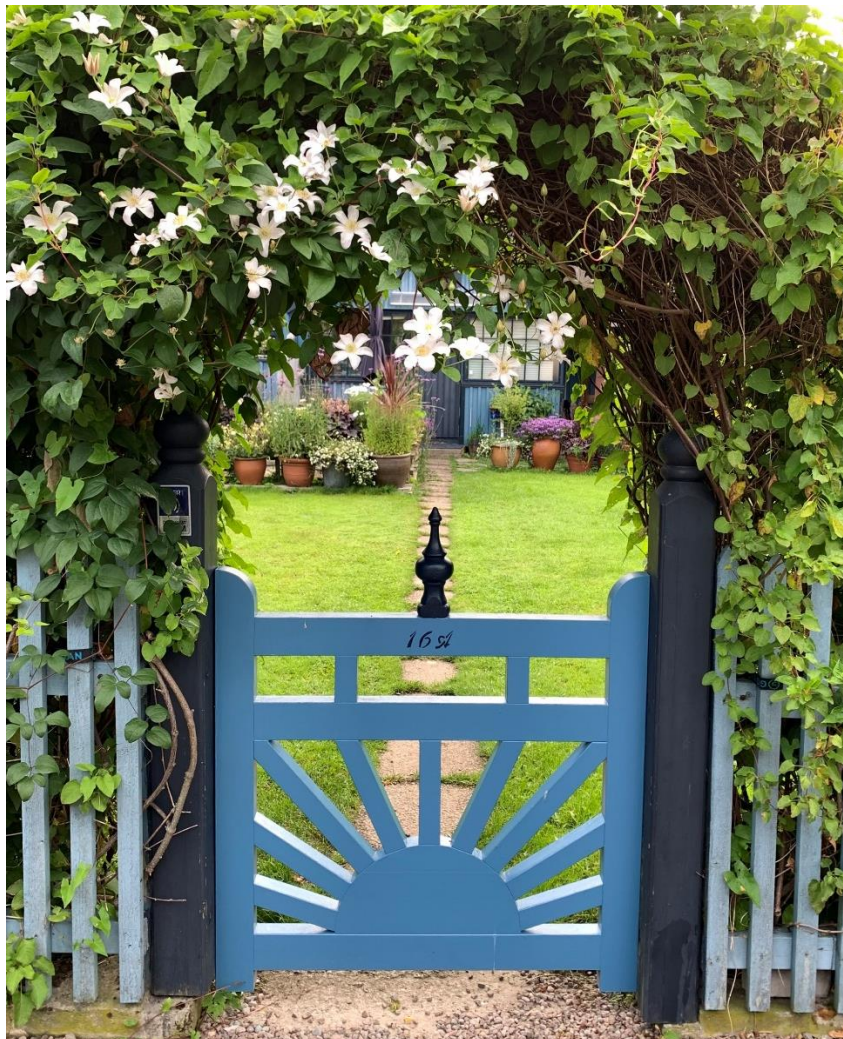




Medlemstidning för
Svensk Epidemiologisk Förening (SVEP)
Årgång 41, Nr 2-3 2023



Tema: Syntetiska data

SVEPET

Medlemstidning för Svensk Epidemiologisk förening och ges ut med fyra nummer per år.

Det är kostnadsfritt att annonsera om kurser och konferenser i SVEPET. Redaktionen förbehåller sig emellertid rätten att ändra i annonsernas layout så att de passar ihop med tidningens innehåll i övrigt.

Ansvarig utgivare

Elisabeth Strandhagen
elisabeth.strandhagen@gu.se

Redaktion

Hannah Brooke
(Redaktör)
hannah.brooke@uu.se

Hanne Krage Carlsen
(Kurser och konferenser)
hannekarlsen@gmail.com

Anton Nilsson
(Statistik och metod)
anton.nilsson@med.lu.se

Svensk Epidemiologisk Förening

SVEP är en tvärvetenskaplig sammanslutning av personer verksamma inom epidemiologi eller angränsande områden.

Medlemsavgiften är 150 kr/år,
alternativt 600 kr för fem år.
Plusgirokonto 440 31 69–8
Swish 1230 3258 52

Hemsida: www.epidemiologi.nu

Medlemskap och adressändring

Elisabeth Strandhagen
Svensk nationell datatjänst
Göteborgs universitet
Box 463
405 30 Göteborg
elisabeth.strandhagen@gu.se

ORDFÖRANDEN HAR ORDET

Hej!

Är det möjligt att använda syntetiska data, eller simulerade data, i stället för data som är insamlade på riktigt? Vad skulle fördelarna kunna vara och när fungerar det dåligt? Dessa frågor försöker vi svara på i detta nummer av Svepet. Har du själv erfarenhet som du vill dela med Svepets läsare i ett uppföljande nummer så hör gärna av dig!

Samarrangemanget med NEON (Nätverket för epidemiologi och nutrition) i Uppsala i juni blev lyckat och välbesökt. Anton Nilsson höll en uppskattad föreläsning om DAG:s och i detta nummer kan du läsa en sammanfattning.

Behovet av ett nationellt hälsodataregister för primärvård har diskuterats i många år. Björn Wettermark redogör för nyttan och var frågan står just nu. Han rapporterar också från två intressanta konferenser inom läkemedelsepidemiologi.

Tack för alla bidrag!

Stort tack också till *Eva Andersson* för insatsen som redaktör för Svepet genom många år! Nu lämnar hon över till *Hannah Brooke* som presenterar sig själv på s 16.

Nästa nummer av Svepet ska handla om kohorter och register. Se vårt upprop på s 19!

Hälsningar Elisabeth



*Elisabeth Strandhagen,
Ordförande i SVEP*

FÖRENINGSPINFORMATION

SVEP:s styrelse

Efter årsmötet har styrelsen följande sammansättning:

Elisabeth Strandhagen, ordförande
Karin Modig, vice ordförande
Karl Mårild, kassör
Anna Axmon, ledamot
Björn Wetterstrand, ledamot
Hannah Brooke, ledamot
Hanne Krage Carlsen, ledamot
Helena Backman, ledamot
Jonas Björk, ledamot

STORT tack till Eva Andersson och Liisa Byberg för värdefullt styrelsearbete genom många år!



Medlemsavgiften

Medlemsavgiften är 150 kr/år, alternativt 600 kr för fem år.
Plusgirokonto 440 31 69 –8
Swish 1230 3258 52

INNEHÅLL

Syntetiska data: Kan vi ersätta äkta vara? Att använda syntetiska eller simulerade data	4
Syntetiska data: Using simulated data with epidemiological research	5
Syntetiska data: Utopi eller möjlighet?	8
Riktade acykliska grafer för analys av orsakssamband	10
Får vi äntligen ett nationellt primärvårdsregister?	14
Hannah Brooke – ny medlem av SVEP:s styrelse	16
Två spännande konferenser om läkemedel	17
Kurser och konferenser	19

Metod: syntetiska data

Kan vi ersätta äkta vara? Att använda syntetiska eller simulerade data

Hanne Krage Carlsen, Registercentrum Västra Götaland
E-post: hanne.carlsen@vgregion.se

I detta nummer av SVEPET beskriver vi några sätt att hantera när man av någon anledning inte kan eller vill jobba med “riktiga”, eller naturliga data.

Naturliga data, det vill säga data som samlas in från verkliga händelser och observationer, utgör grunden för många datadrivna tillämpningar och maskininlärningsalgoritmer. Dock finns det situationer där tillgång på sådana naturliga data är begränsad, dyra att samla in eller till och med etiskt problematiska att använda, till exempel vid studentprojekt. I dessa fall blir syntetiska eller simulerade data en värdefull resurs.

Simulerade data tas fram av statistikprogram där man kan ta fram dataset med önskade egenskaper och fördelningar. Man kan också ta chatGPT till hjälp, testa till exempel en prompt i stil med ”simulera en datamängd med 10 observationer med blodtryck, kön och ålder” nästa gången det behövs något slumpmässigt för att visa fram i en graf.

Syntetiska data skapas eller genereras av datorprogram och efterliknar mönster och egenskaper som finns i en riktig, eller naturlig datamängd. Dessa genereras vanligtvis genom att använda avancerade algoritmer och modeller som tar hänsyn till de statistiska egenskaperna hos de naturliga data som de försöker efterlikna.

I artikeln om simulerade data beskriver Hannah Brooke och Thomas Vogt hur man simulerar data, och också hur detta kan vara praktiskt för att jämföra metoder i datamängder med olika distributioner för att jobba med praktiska exempel och tillämpningar.

I den andra artikeln berättar Erik Bülow om ett försök att ta fram ett syntetiskt dataset baserat på riktiga data från ett kvalitetsregister, men där det fanns juridiska och tekniska utmaningar som försvårade processen.

Syntetiska eller simulerade data kan lösa problemet med brist på tillgängliga naturliga data, särskilt i nya eller nischade områden där datainsamling är svår, men de kan också användas för att öka datamängden och därigenom förbättra prestanda och generalisering av maskininlärningsalgoritmer. Dessutom kan syntetiska data användas i situationer där det är oetiskt, kostsamt eller riskabelt att använda verkliga data, som det ibland kan vara med data från medicinsk forskning som innehåller känsliga personuppgifter.



Illustration Hanne Krage Carlsen (logga från författarens egen t-shirt från phdcomics.com)

Även om det kan låta lockande att använda syntetiska eller simulerade data har dessa även sina utmaningar och måste noggrant kalibreras och valideras för att säkerställa att de är representativa för de verkliga data de efterliknar. Dessutom kan det vara svårt att helt efterlikna komplexiteten och variationen hos verkliga data som det beskrivs i de följande artiklar på sidorna 5-7 och 8-9.

Metod: syntetiska data

Using simulated data within epidemiological research

Hannah Brooke, Inst för kirurgiska vetenskaper, Medicinsk epidemiologi, Uppsala universitet
E-post: hannah.brooke@uu.se

Thomas Vogt, Inst för kirurgiska vetenskaper, Medicinsk epidemiologi, Uppsala universitet
E-post: thomas.vogt@uu.se

Hannah Brooke has asked Thomas Vogt some questions about his research and why and how he uses synthetic data in his work.

Who are you and what are you working on?

I am a PhD student in epidemiology at Uppsala University with a background in sociology and public health. My doctoral studies focus on complications of type 2 diabetes. At the moment, my supervisors and I aim to use register data to investigate the relationship between lifestyle trajectories and the risk of neuropathy-related complications.

What are you using synthetic/simulated data for in your work? Why are you using synthetic/simulated data for this? Advantages?

I use simulations for mainly three purposes. Firstly, to become more familiar with statistical software. Secondly, as a tool to learn about statistical models used to analyse data. Thirdly, to gauge how those statistical models would be affected by assumptions' violations that are likely to happen when used with real data. Such situations can arise for example when an outcome variable is not continuous but collected on an ordinal scale.

More specifically, I have been interested in joint latent class models to simultaneously model longitudinal and survival processes. These models are composed of multinomial, mixed, and survival sub-models. Simulating data with some covariates, a longitudinally collected variable, and a time-to-event variable, and then fitting a joint latent class model on the simulated data is a helpful tool for understanding how the model works, and it is a useful addition to reading the related literature.

Because using simulated data allows us to know what the true values are (because they are simulated), we can then assess if and by how much parameters estimated by a model deviate

from the simulated truth. In the case of joint latent class models, I can look at the results using different ways of selecting the starting values, as those models can frequently converge to local maxima. In this case, with simulations, I can see that indeed, at least under some conditions, the models often converge to local maxima. This emphasizes the need to try models with different starting values if one wants to estimate a joint latent class model that converges to the global maximum.

Ultimately, given our research questions and our data, simulations might aid in determining whether a method is adequate, or if we need to favour alternatives.

What sort of results do you get from using synthetic/simulated data?

The type of results one gets will depend on the reason for running simulations. For example, one might want to learn more about the influence of sample size on sampling variation and on the subsequent estimated parameters of a model (assuming everything else to be equal and the model to be correctly specified). Example code in R is shown below, for linear regression. We simulate an outcome (y) that depends on a continuous covariate (x), on a categorical covariate (z), on an intercept, and on a random term. We fit a linear model to the simulated data, and we perform this operation 10,000 times for four different sample sizes ($n=50$, $n=100$, $n=500$, and $n=1000$). Then, we plot the results as four histograms, where each histogram displays the distribution of the parameters for each of the four scenarios with different sample sizes (here, only for the coefficient for x).

```

set.seed(354)

# FUNCTION TO SIMULATE THE DATA -----

sim_fun <- function(n) {
  x <- rnorm(n, 20, 6.5)
  z <- rbinom(n, size = 1, prob = c(0.3, 0.7))
  y <- 3 + 0.1*x + 1.2*z + rnorm(n, 0, 5)
  return(coef(lm(y ~ x + z)))
}

# REPEAT THE SIMULATION 10,000 TIMES FOR DIFFERENT SAMPLE SIZES -----

sample_size <- c(50L, 100L, 500L, 1000L)

results <- lapply(
  sample_size,
  function(i) {
    res <- matrix(as.numeric(NA), nrow = 10000, ncol = 3)
    for (j in 1:10000) res[j, ] <- sim_fun(n = i)
    return(res)
  }
)

# PLOT THE RESULTS FOR THE X COEF -----

par(mfrow = c(2, 2))

for (i in seq_along(results)) {
  hist(
    results[[i]][, 2],
    main = paste("Distribution of the coefs for x when the sample size is",
                 sample_size[i]),
    sub = paste0("(standard deviation = ",
                 round(sd(results[[i]][, 2]), digits = 3),
                 ")"),
    xlab = "x",
    xlim = c(-0.4, 0.6)
  )
  abline(v = mean(results[[i]][, 2]), lwd = 3)
}

```

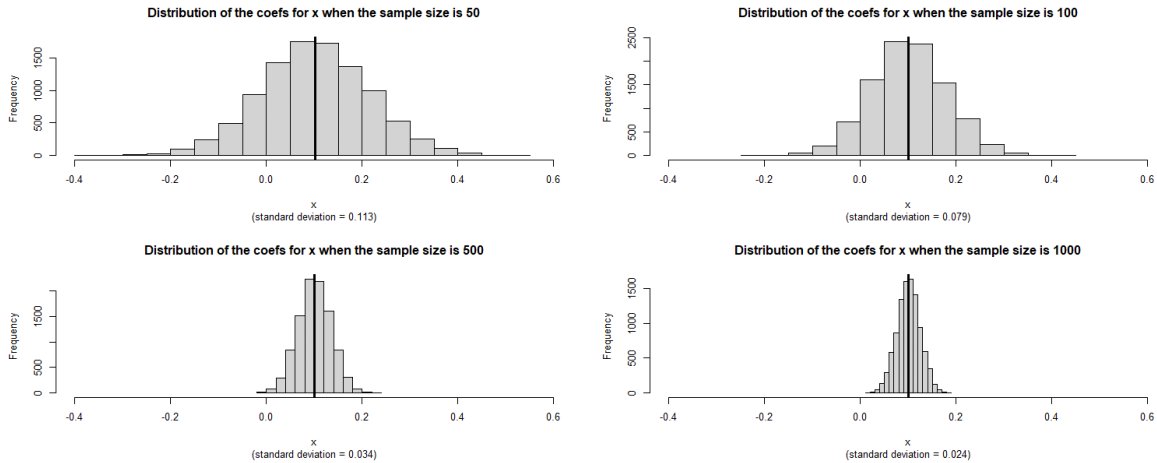


Fig: The resulting plot, where the darker vertical line is the mean of all coefficients for the specific sample size. The “true population value” here is 0.1.

Beyond sampling variation from simulation to simulation, we can see that when the sample size is small, even if the true model is specified, the estimated parameter can suggest a relation between x and the outcome that is in the opposite direction compared to the true simulated value:

```
perc_below_zero <- sapply(results, function(x) sum(x[, 2] < 0) / nrow(x))
names(perc_below_zero) <- paste0("n=", sample_size)
perc_below_zero

##   n=50  n=100  n=500  n=1000
## 0.1807 0.0985 0.0015 0.0000
```

Here, this is the case in 18.07 % of the simulations when $n=50$, and in 9.85 % of the simulations when $n=100$. These proportions would become larger if the standard deviation of the error term was increased when generating y , for example with:

```
y <- 3 + 0.1*x + 1.2*z + rnorm(n, 0, 20)
```

Depending on the application, one or the other might be more plausible and/or interesting.

However, it should be noted that even if data are simulated to resemble a practical application and a real-world dataset, it will likely be impossible to simultaneously assess all potential sources of systematic and random errors that may be affecting the output of a statistical model. In addition, if the simulated data does not resemble the real-world data or the true data-generating process in some aspects, its usefulness will be limited. Nevertheless, it will still likely remain an interesting addition to other ways of learning about statistical software, statistical models, and epidemiologic concepts.

Some references related to the use of simulations in epidemiology and medicine can be of interest, for example:

1. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statistics in Medicine*. 2006;25(24):4279-92.
2. Fox MP, Nianogo R, Rudolph JE, Howe CJ. Illustrating How to Simulate Data From Directed Acyclic Graphs to Understand Epidemiologic Concepts. *American Journal of Epidemiology*. 2022;191(7):1300-6
3. Rudolph JE, Fox MP, Naimi AI. Simulation as a Tool for Teaching and Learning Epidemiologic Methods. *American Journal of Epidemiology*. 2021;190(5):900-7.

Syntetiska data: Utopi eller möjlighet?

Erik Bülow, Institutionen för medicin, Sahlgrenska Akademin, Göteborgs universitet
E-post: erik.bulow@gu.se

Att dela känsliga persondata är svårt. Syntetisering innebär att man i stället efterliknar en verklig datamängd men utan att enskilda individer i efterhand går att identifiera. Tyvärr tycks det vara mer komplicerat än hur det ibland beskrivs. Utmaningarna är både statistiska, tekniska, etiska och juridiska. Vårt försök fick tyvärr läggas på is.

Enligt det amerikanska konsultföretaget Gartner kommer maskininlärningsmodeller inom sju år att använda mer syntetiska än riktiga data. Varför då? Det finns flera skäl men det kanske mest relevanta avseende epidemiologi rör legala och integritets-känsliga aspekter.

För egen del kom jag i kontakt med metodiken i samband med att Registercentrum Västra Götaland önskade tillgängliggöra syntetiska data från de kvalitetsregister för vilka man har ett centralt personuppgiftsansvar. Tanken var inte att begränsa tillgången till den äkta varan men att kunna erbjuda forskare, studenter och allmänhet en smidig ”förhandsvisning”, utan att först kräva etikgodkännande, sekretess- och menprövning, datahanteringsplan, säkra tekniska lösningar och i vissa fall flertalet juridiska avtal. I detta fall rörde det sig alltså om så kallad ”tabulärdata” (även om syntetiseringsmetoder finns även för bild och text).

Vi tänker oss således en datamängd med ett antal variabler och observationer. Att bara plocka bort direkt identifierbar information såsom namn och personnummer innebär att eventuellt ovanliga kombinationer av tider, platser, diagnoser och demografiska data ändå kan ge ledtrådar till ”vem som är vem”. Data är då ”pseudonymiserade” men inte nödvändigtvis ”anonymiserade”. Är syftet bara att ge tillgång till en datamängd utan individdata, men med korrekta variabelnamn och kategorier (för att exempelvis underlätta forskares inledande programmering inför framtida bearbetning), är ett alternativ att betrakta varje variabel som oberoende. Att då bara slumpa fram en ny ordning (permutation) av respektive variabel låter sig lätt göras, men ovanstående problem med ovanliga kombinationer kvarstår och kan potentiellt identifiera individer trots att desammas uppgifter inte längre förekommer på samma rad. Pondera att vi i en variabel ser att det finns en individ med yrke ”tunnbindare”, och att det enligt en annan variabel finns en boende i

”Byhåla”. Känner vi då en tunnbindare från Byhåla så kan vi gissa oss till att densamme nog fanns med i originaldata. Detta ger förvisso inte per definition några känsliga uppgifter om individen, men om datamängden som sådan är tänkt att emulera ett register över individer med hösnuva. Ja, då har vi plötsligt fått en indikation på att personen i fråga lider av just en sådan diagnos, vilket i sig är en känslig personuppgift.

En metod kallad *k*-anonymisering finns för att undvika sådana fadäser och problematiken beskrevs redan 1986 av svenske Tore Dalenius.¹ Med rakt av simulerade data (se intilliggande artikel) kan vi förstås också först anta att en viss variabel följer en på förhand given fördelning. Individens längder kan antas vara normalfördelade. Skattar vi medelvärde och standardavvikelse från originaldata kan vi lätt simulera sådan data. För en könsvariabel gör vi motsvarande med en binomialfördelning där andelen män (kvinnor) skattas på samma sätt. Nackdelen blir förstås att vi bryter associationen mellan dessa två variabler. Resultatet blir därmed tämligen ointressant för den som vill träna varken studenter eller statistiska modeller avseende associationer, trender eller prediktioner. För att fånga längdfördelningens bimodala natur utifrån kön krävs i stället en betingad fördelning. Råkar vi därutöver veta att få individer under två år är längre än 175 centimeter et cetera så krävs en alltmer komplex multivariat modell. Hur denna bäst formuleras har vi sällan koll på men R-paketet ”synthpop”² kan underlätta processen för den som vill testa.

Går vi ett steg längre hittar vi metoden ”Conditional Tabular Generative Adversarial Network” (CT-GAN).³ Då handlar det om neurala nätverk och faller inom ramen för vad som ofta beskrivs som artificiell intelligens. För sådana metoder finns spår av R-implementeringar på nätet men främst är det Python som gäller.⁴ För stora

datamängder är metoden dock beräkningsintensiv.

Under vårt försök med syntetisering på Registercentrum hade vi tillgång till VGR:s AI-plattform. Försöket blev därmed också ett pilotprojekt för att testa de möjligheter denna nya superdator har att erbjuda. Noteras bör emellertid att kraften i denna monstermaskin främst ligger i dess grafiska processorer (GPU:er). Desamma kan göra underverk för maskininlärningsmodeller kopplade till bildbehandling men är dessvärre inte till lika stor hjälp för tabulärdata av den typ vi avsåg syntetisera.

Ett skäl till regionens satsning på denna plattform är möjligheten att bearbeta integritetskänsliga hälsodata inom organisationens egna väggar. För även om resultatet av syntetiseringen i bästa fall blir en datamängd utan identifierbara individer, så krävs förstås tillgång till originaldata dessförinnan. Med tanke på den eventuellt stora beräkningskraften som krävs blir molnlösningar i andra sammanhang ett alternativ, men detta är inte lämpligt för känslig hälsodata av datasäkerhetsskäl. The National Academic Infrastructure for Super-computing in Sweden (NAISS) ger forskare tillgång till ett alternativt beräkningskluster "NAISS SENS Bianca"⁵ men det juridiska ansvaret för datasäkerheten åligger fortsatt forskningshuvudmannen.

En av principerna med CT-GAN är att metoden låter två modeller tävla mot varandra. I ett första steg skapas en syntetiserad datamängd som i största möjligaste mån ska efterlikna originaldata. Därefter försöker den andra modellen hitta brister i anonymiseringsprocessen som gör att individer trots allt kan identifieras. En datamängd kan utgöra en närmast perfekt syntetisering men blir samtidigt olämpligt ur integritetssynpunkt och vice versa.

Processen bollas fram och tillbaka till dess man hittat en lämplig "trade-off" mellan en perfekt syntetisering som samtidig håller måttet för dataintegritet.

Tyvärr visade emellertid Stadler med flera⁶ förra året att denna metod, ibland beskriven som en "silver-bullet", i själva verket inte håller måttet: *"synthetic data does not provide a better tradeoff between privacy and utility than traditional anonymisation techniques."* Vidare menade man att det kan vara nästintill omöjligt att verkligen garantera att ingen individ från originaldata får sin identitet röjd.

Dessa svårigheter berör både teknik, etik och juridik. I det senare fallet testade Integritetsskyddsmyndigheten (IMY) under 2022 ett nytt arbetssätt. Utöver den oftast retroaktiva rollen

som tillsynsmyndighet inleddes under 2022 ett pilotprojekt med regulatorisk testverksamhet. Syftet var då att på förhand ge VGR och Region Halland vägledning kring möjliga lagtolkningar på området.⁷ Tyvärr visade sig frågan dock för svår även för detta forum.

Vart ledde då vårt försök med syntetiserade data från ett kvalitetsregister? Kontentan blev tyvärr att vi inte just då vågade ta risken att öppet tillgängliggöra den datamängd vi lyckats syntetisera. Förmodligen hade det varit relativt riskfritt om endast det slutgiltiga resultatet publicerats. Här finns dock frågor kring huruvida man som offentlig myndighet även kan tvingas lämna ut själva de underliggande algoritmerna för syntetiseringen till den som begär detta inom ramen för offentlighetsprincipen. Med tillgång till både syntetiserade data och den genererande modellen såg vi tyvärr en teoretisk möjlighet till att en illasinnad "hacker" på något sätt ändå kan lyckas avslöja information som egentligen borde föregås av just den mycket rigorösa process som idag tillämpas inför utlämnande av känsliga personuppgifter inom EU. Projektet ligger därmed på is. Men frågan huruvida det hela är en bra idé eller inte är knappast avgjord.

Noteras bör att det även finns flertal företag som erbjuder datasyntetisering som tjänst för den som inte själv har tillgänglig beräkningskapacitet. Några garantier för total anonymitet tycks dock inte ges heller av dessa pionjärer på området.⁸

Referenser

1. Tore Dalenius, "Finding a Needle In a Haystack", Journal of Official Statistics, Vol. 2, No. 3, 1986, pp. 326–336.
2. <https://www.synthpop.org.uk/index.html> besökt 2023-09-20.
3. Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni. Modeling Tabular data using Conditional GAN. NeurIPS, 2019
4. Exempelvis The Synthetic Data Vault (SDV) <https://docs.sdv.dev/sdv> besökt 2023-09-21
5. https://supr.naiss.se/round/open_type/?type=NAISS+SENS besökt 2023-09-21
6. <https://www.usenix.org/conference/usenixsecurity22/presentation/stadler> besökt 2023-09-21
7. <https://www.imy.se/nyheter/nytt-arbetsatt-testas-for-att-ge-innovationsprojekt-vagledning-om-dataskydd/> besökt 2023-09-21
8. <https://youtu.be/hcuu29M3Ac4?si=F6DwVnQp mFVBfGP1> besökt 2023-09-21

Metod: DAG:s

Riktade acykliska grafer för analys av orsakssamband

Anton Nilsson, EPI@LUND, Avdelningen för arbets- och miljömedicin, Lunds universitet
E-post: anton.nilsson@med.lu.se

Riktade acykliska grafer (eng: directed acyclic graphs; DAG:s) har på senare år blivit ett allt vanligare redskap i den epidemiologiska verktygslådan. Genom att rita upp en eller flera DAG:s kan forskaren klargöra på vilket sätt hen tänker sig att exponeringen och utfallet av intresse hänger ihop. Vidare kan forskaren genom att tillämpa en uppsättning relativt enkla regler på sin graf dra slutsatser om vilka variabler som en analys av orsakssamband behöver kontrollera för, och vilka variabler som analysen bör undvika att kontrollera för. I denna artikel ger jag en kort introduktion till DAG:s och hur de kan användas inom epidemiologisk forskning.



Anton Nilsson under det föredrag om DAG:s som han höll under SVEP:s arrangemang tillsammans med NEON i juni

Att bestämma orsakssamband utgör ofta en avsevärd utmaning inom både medicinsk och samhällsvetenskaplig forskning, särskilt i studier som baserar sig på observationsdata. Ett statistiskt samband mellan två variabler innebär inte nödvändigtvis att den ena av dessa variabler utövar ett inflytande på den andra, eller att ett eventuellt orsakssamband mellan de båda nödvändigtvis måste vara av samma magnitud som det statistiska sambandet. En omtalad artikel i *New England Journal of Medicine* från 2012 visade exempelvis att länder där invånarna äter mer choklad kan stoltsera med fler Nobelpris per capita¹. Att detta samband i någon meningsfull

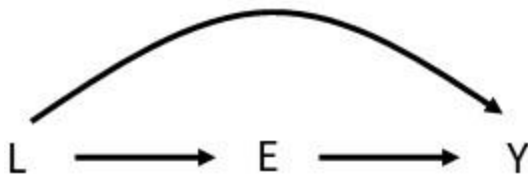
utsträckning skulle representera en effekt av chokladintag på förmågan att producera banbrytande forskning (eller att skriva romaner som faller Svenska Akademien i smaken) ska man nog dock inte tro på – ej heller, för den delen, att fler Nobelpriser skulle driva upp intaget av choklad. Mer troligt är att sambandet drivs av skillnader i länders ekonomiska utveckling, geografi, kultur och andra så kallade *förväxlingsfaktorer* (eng: confounders).

Riktade acykliska grafer²⁻³ (DAG:s) är ett verktyg som kan användas för att illustrera och hantera fenomenet *förväxling*, liksom andra komplikationer som kan förekomma i analyser där syftet är att bestämma ett orsakssamband. Att DAG:s blivit så populära hänger sannolikt ihop med att de både är intuitiva och vilar på en solid matematisk grund. Genom att tillämpa några relativt enkla regler kan forskaren utifrån sin graf dra slutsatser om vilka variabler som bör kontrolleras för – liksom vilka som inte bör kontrolleras för – givet en uppsättning kvalitativa antaganden om hur olika variabler hänger samman, som forskaren har valt att utgå från.

En DAG består av två sorters komponenter: noder och pilar. Varje nod representerar en variabel. Typiskt sett innehåller en DAG en nod som representerar den exponering som forskaren är intresserad av att bestämma effekten av, och en som representerar det utfall forskaren vill bestämma effekten av exponeringen på. Dessutom innehåller figuren noder som representerar andra variabler, som på ett eller annat sätt hänger ihop med (eller misstänks hänga ihop med) exponeringen och utfallet.

Genom att rita in pilar i sin DAG illustrerar forskaren de orsakssamband hen utifrån tidigare kunskaper tror eller misstänker existerar. Endast enkelriktade pilar är tillåtna i en DAG, och det får inte lov att finnas några loopar, så om nod A pekar på nod B och nod B pekar på nod C , så får inte nod C i sin tur peka på nod A . Detta är rimligt om man betänker att orsakssamband är något som alltid inträffar i tidens riktning; något som sker i framtiden kan inte påverka något som hänt i det förflutna. För att tydliggöra tidsaspekten är det vanligt att rita en DAG på så sätt att noder längre till höger alltid representerar händelser längre fram i tiden. Alla pilar i DAG:en kommer då att peka åt höger.

Figur 1 är ett exempel på en DAG, ritad i linje med denna princip. Figuren illustrerar ett enkelt scenario med förväxling. Vi antar att en forskare är intresserad av att bestämma effekten av en exponering (E) på ett utfall (Y); denna effekt representeras av pilen från E till Y . Men samtidigt som exponeringen tros kunna påverka utfallet, finns det en bakomliggande faktor (L) som utgör en gemensam orsak till exponeringen och utfallet. Variabeln L är en förväxlingsfaktor, ty den ger upphov till ett statistiskt samband mellan E och Y som inte är ett orsakssamband. Även om E kanske har en effekt på Y , kommer våra skattningar av denna effekt därför att tendera att bli snedvridna.



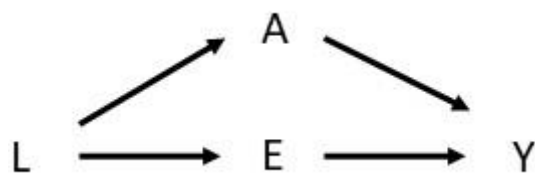
Figur 1: Ett enkelt exempel på förväxling illustrerat med en DAG. Förväxlingsfaktorn L (säg ekonomisk utveckling) är en gemensam orsak till exponeringen E (chokladintag per capita) och utfallet Y (Nobelpris per capita).

Med DAG-terminologi brukar man säga att en gemensam orsak till en exponering och ett utfall, såsom L i figuren ovan, ger upphov till en *bakväg* (eng: backdoor path) mellan exponeringen och utfallet. Bakvägen i Figur 1 utgörs av pilen mellan L och E , tillsammans med pilen mellan L och Y . För att hantera den förväxling som uppstår på grund av bakvägen kan forskaren kontrollera för L i sin analys. Med DAG-terminologi säger man då att man *blockerar* bakvägen, och bakvägen blir då *stängd*. Ibland markerar man att man

kontrollerat för en viss variabel genom att rita en kvadrat runt denna. Om vi kontrollerar för L kan vi i ovanstående figur således rita en kvadrat runt L . Bakvägen mellan E och Y blir nu stängd – det ”flöde” mellan E och Y som löper via L är brutet.

Det finns flera olika sätt att kontrollera för en variabel, såsom L . Ett sätt är att *stratifiera* med avseende på variabeln. Stratifiering innebär att man studerar sambandet mellan exponering och utfall separat inom olika grupper av observationer, inom vilka variabeln L har ett och samma värde. Ett specialfall av detta är *restriktion*, där man bara studerar sambandet bland observationer med ett visst värde denna variabel. En annan metod är *matchning*, där varje exponerad individ kopplas ihop med en icke-exponerad med samma värde på L . Det vanligaste sättet att kontrollera för en variabel är dock att skatta en *multipl regressionsmodell*, där variabeln L , utöver exponeringen E , inkluderas som en kovariat.

I vissa fall kan bakvägar mellan exponering och utfall blockeras genom att man kontrollerar för något annat än just den gemensamma orsaken. Betrakta till exempel DAG:en i Figur 2, där vi kan tänka oss att exponeringen E är rökning och utfallet Y är kardiovaskulär sjukdom. En underliggande variabel L , säg hälsomedvetenhet, påverkar huruvida man röker, men påverkar också kardiovaskulär sjukdom via alkoholintag, A . Precis som i Figur 1 är L en förväxlingsfaktor; den är en gemensam orsak till exponering och utfall. Bakvägen mellan exponering och utfall utgörs av pilen från L till E , tillsammans med pilen från L till A och pilen från A till Y . På något sätt måste denna bakväg blockeras. Detta kan potentiellt göras genom att man kontrollerar för L , liksom vi tidigare har diskuterat. I praktiken är detta dock kanske inte möjligt, då forskaren inte har tillgång till några bra mått på hälsomedvetenhet. Ett alternativ är därför att kontrollera för A . Även detta kommer att göra att bakvägen mellan E och Y blockeras, och förväxlingsproblemet undviks.



Figur 2: Ett något mera komplicerat exempel på förväxling illustrerat med en DAG. Förväxlingsfaktorn L (hälsomedvetenhet) är en gemensam orsak till

exponeringen E (rökning) och utfallet Y (kardiovaskulär sjukdom), det senare via A (alkoholintag).

I praktiken finns det förstås många hälsorelaterade beteenden som kan bero på hälsomedvetenhet, och som i sin tur kan tänkas påverka risken för kardiovaskulär sjukdom. Vi kan därför tänka oss en mer realistisk DAG, som inte bara innehåller en variabel A , utan flera stycken; A_1 , A_2 , A_3 och så vidare, var och en med en inkommande pil från L och en utgående till Y . Var och en av dessa ger upphov till en bakväg som behöver kontrolleras för. Givet att vi inte har möjlighet att kontrollera för L behöver samtliga dessa "A-variabler" därmed kontrolleras för. Om vi lyckas med detta, blockerar vi samtliga bakvägar mellan exponering och utfall, och förväxlingsproblemet elimineras.

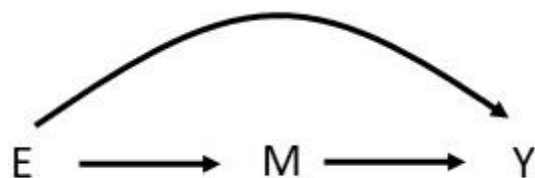
Även scenarier liknande de ovanstående, men med variabler mellan L och E kan förstås förekomma. I dessa fall har vi ett ytterligare ett sätt att hantera förväxlingsproblemet på, nämligen att kontrollera för samtliga variabler mellan L och E , vilket också skulle blockera samtliga bakvägar mellan E och Y . Generellt måste alla bakvägar mellan en exponering och ett utfall i en DAG på något sätt blockeras för att en skattning ska avspegla ett orsakssamband. En sådan blockering kan ske genom att samtliga gemensamma orsaker till exponeringen och utfallet kontrolleras för, eller mer allmänt genom att minst en variabel längs varje bakväg kontrolleras för. Notera att begreppet förväxlingsfaktor används på lite olika sätt – vissa använder det enbart för att syfta på gemensamma orsaker till exponering och utfall, medan andra använder det för samtliga variabler längs bakvägar mellan exponering och utfall, dvs. samtliga de variabler som kan utnyttjas för att eliminera förväxlingsproblemet.

Som vi har sett kan ett orsakssamband verka genom ett eller flera mellanliggande steg, där en variabel påverkar en annan, som i sin tur påverkar en tredje. I själva verket är detta alltid fallet. En "fullständig" DAG med alla mellanliggande steg (till exempel med alla biologiska mekanismer på molekylnivå) utritade skulle bli helt ohanterlig. Det är i allmänhet dock inte nödvändigt att rita ut några mellanliggande steg i en DAG, såvida man inte överväger att kontrollera för någon av de mellanliggande variablerna i sin analys.

Om en variabel utgör ett mellanliggande steg mellan exponering och utfall så kallas den för en

mediator. En mediator är ingen förväxlingsfaktor, men är ibland relevant att kontrollera för, då detta möjliggör en så kallad *mediationsanalys*, som syftar till att utreda på vilket sätt exponeringen påverkar utfallet. Betrakta till exempel Figur 3, där vi kan anta att E representerar rökning, M ansamling av tjära i lungorna och Y lungcancer. Rökning leder till ansamling av tjära i lungorna, som i sin tur leder till ökad risk för lungcancer. Rökning påverkar alltså lungcancer genom en mekanism som sammantaget kan beskrivas av en pil från E till M , och en pil från M till Y . Variabeln M är en mediator; den ligger mellan E och Y .

Men kanske påverkar rökning även risken för lungcancer via andra mekanismer än genom ansamling av tjära i lungorna, till exempel via ansamling av andra ämnen i kroppen? I figuren har jag illustrerat denna möjlighet genom att också lägga in en direkt pil från E till Y . Dessa mekanismer kan även göras explicita med egna noder, som skulle kunna ritas in mellan E och Y , men detta är inte nödvändigt. Om forskningsfrågan är huruvida (eller i vilken mån) rökning påverkar lungcancer via andra mekanismer än ansamling av tjära i lungorna, spelar det ingen roll exakt vilka dessa mekanismer är, utan de kan sammanfattas med en enda pil.



Figur 3: Exponeringen E (rökning) antas kunna påverka M (mängden tjära i lungorna), som i sin tur antas kunna påverka Y (lungcancer). Dessutom antas E direkt kunna påverka Y genom andra mekanismer (inandning av andra substanser än tjära).

För att identifiera de effekter av rökning på lungcancer som inte verkar via M kan vi kontrollera för denna. Den effekt av exponeringen på utfallet som verkar via mediatorsn kommer nu att blockeras, och det som återstår är den *direkta effekten*, dvs. det samband som ritats ut med en direkt pil i figuren. Om vi trots att vi kontrollerat för M finner en effekt av rökning på lungcancer, talar detta för att rökning påverkar lungcancer via andra mekanismer än via mediatorsn, tjära i lungorna. Den *indirekta effekt* som verkar via mediatorsn kan i sin tur också skattas. I en enkel linjär modell är denna lika med differensen mellan

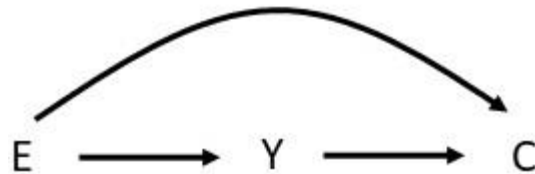
totaleffekten (dvs. effekten av E på Y ej kontrollerad för M) och den direkta effekten. Andra modeller kräver mer avancerade ansatser.⁴

Ett sista DAG-koncept som är bra att känna till är *kollideringsbias* (eng: collider bias), vilket även kallas selektionsbias. För att förstå detta måste vi först introducera begreppet *kolliderare* (eng: collider). En kolliderare är en variabel som påverkas av minst två andra. Att en DAG innehåller kolliderare är inget problem i sig, men problem kan uppstå om man väljer att kontrollera för en kolliderare i sin analys. Om man kontrollerar för en kolliderare uppstår nämligen statistiska samband mellan de variabler som påverkar kollideraren, utan att det behöver finnas något orsakssamband mellan dessa. Det mest uppenbara exemplet på när detta blir ett problem är scenariot där man kontrollerar för en variabel som beror på både exponeringen och utfallet. I detta fall kommer ett statistiskt samband mellan exponeringen och utfallet att uppstå, trots att något orsakssamband mellan dem kanske inte existerar. Och även om ett orsakssamband mellan exponering och utfall skulle existera, så kommer det statistiska sambandet mellan dessa två nu att tendera att avvika från orsakssambandet.

Berksons bias⁵ är ett klassiskt exempel på kollideringsbias. I sin artikel från 1946 diskuterade Joseph Berkson sambandet mellan diabetes och gallblåseinflammation, och visade med ett räkneexempel att dessa sjukdomar skulle kunna vara negativt korrelerade i en population av sjukhuspatienter även om inget samband mellan de båda existerade i totalpopulationen. Låt oss nu med hjälp av en DAG visa hur detta kan gå till.

Antag att E i Figur 4 representerar diabetes, som är den exponering vi är intresserade av. Utfallet gallblåseinflammation representeras av Y . Det finns ingen effekt av E på Y , så pilen mellan E och Y hade egentligen kunnat tas bort, men vi ritar ändå ut den eftersom det är denna effekt vi tänker oss att forskaren vill bestämma, och vi antar att forskaren inte känner till att det i själva verket inte finns någon effekt av E på Y . Noden C representerar huruvida en person har blivit inlagd på sjukhus ($C=1$ för sjukhuspatienter; $C=0$ för personer som inte blivit inlagda på sjukhus). En person kan hamna på sjukhus av en mängd olika orsaker, bland annat som en följd av diabetes eller gallblåseinflammation. Både E och Y påverkar alltså C , som därmed är en kolliderare. Detta innebär att det är problematiskt att kontrollera för

C , eftersom ett statistiskt samband mellan E och Y då kommer att genereras. Som tidigare nämnts finns det olika sätt att kontrollera för en variabel, men restriktion är ett. Om vi begränsar oss till populationen $C=1$ kommer ett samband mellan E och Y alltså att uppstå även om ingen av dessa båda variabler påverkar den andra.



Figur 4: Exponeringen E (diabetes) antas kunna påverka Y (gallblåseinflammation). Båda dessa påverkar i sin tur C (huruvida man har blivit inlagd på sjukhus).

Intuitivt är ovanstående problem ganska lätt att förstå om vi – för enkelhets skull – tänker oss att det bara finns två anledningar till att en person kan hamna på sjukhus: diabetes och gallblåseinflammation. Bland sjukhuspatienter kommer en person som inte har diabetes nu garanterat att ha gallblåseinflammation, och en person som inte har gallblåseinflammation kommer garanterat att ha diabetes. Ett negativt samband mellan de båda tillstånden har uppstått.

Jag har i denna artikel introducerat DAG:s och visat hur dessa kan användas i analyser av orsakssamband. Jag har med några relativt enkla exempel visat att vissa variabler (förväxlingsfaktorer) är önskvärda att kontrollera för, medan andra (kolliderare) kan vara högst problematiska att kontrollera för, och ytterligare andra (mediatorer) endast bör kontrolleras för om man vill få fram den del av en totaleffekt som verkar via vissa specifika mekanismer. Man bör dock notera att en DAG alltid bygger på en mängd antaganden kring hur olika variabler hänger ihop. I praktiken är det sällan självklart hur olika variabler påverkar varandra, och vilka variabler som är förväxlingsfaktorer, mediatorer och kolliderare. Det är därför ofta klokt att rita upp flera alternativa DAG:s, och göra känslighetsanalyser utifrån dessa.

Framställningen har i denna artikel av utrymmes-skal varit ganska översiktlig, men kan – om jag har lyckats med min föresats – förmodligen vara tillräcklig för den genomsnittlige läsaren. Den som önskar fler exempel eller en mer

fullständig framställning av teorin bakom DAG:s kan exempelvis vända sig till någon av referenserna.²⁻³

Referenser

1. Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. *N Engl J Med* 2012;367:1562-4.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiol* 1999;10:37-48.
3. Hernán MA, Robins JM. Graphical representation of causal effects. I *Causal Inference: What If*, Boca Raton, 2020. Chapman & Hall/CRC.
4. VanderWeele T. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17-32.
5. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics* 1946;2:47-53.

Får vi äntligen ett nationellt primärvårdsregister?

Björn Wettermark, professor i Läkemedelsepidemiologi vid Uppsala universitet
E-post: bjorn.wettermark@farmaci.uu.se

Behovet av ett nationellt hälsodataregister för primärvård har diskuterats i många år. Innan sommaren gav regeringen den enskilda utredaren Ingela Alverfors i uppdrag att se över hälsoregistrerna. Kommer det göra det möjligt att få bättre data om vad som sker i primärvården?



Björn Wettermark

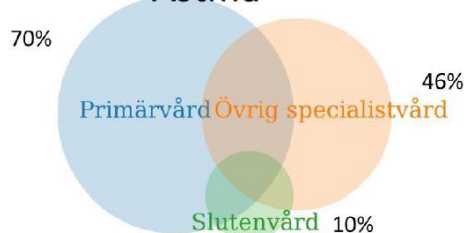
Nationella hälsodataregistren och fler än 100 kvalitetsregister har bidragit till Sveriges unika förutsättningar för epidemiologisk forskning om hälsa och sjukvård. Men det finns flera luckor i de nationella registren, en av dem är avsaknaden av ett nationellt primärvårdsregister med de diagnoser och åtgärder som sker i primärvården.

Mer än hälften av befolkningen har minst ett läkarbesök i primärvård varje år och för många är det första vårdkontakten. I rapporten "Primärvårdens roll i hälso- och sjukvårdssystemet" beskrev vi var patienter i Region Stockholm får sin vård och behandling. Ungefär hälften av alla patienter med vanliga folksjukdomar som hypertoni, ångest och astma får under en femårsperiod bara sin diagnos i primärvården. Deras diagnoser och vårdåtgärder saknas helt i patientregistret.

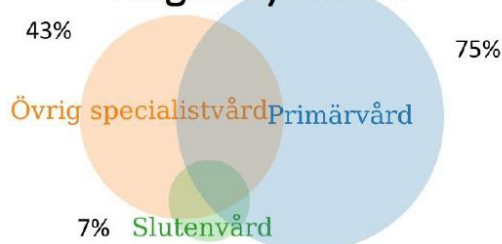
Hypertoni



Astma



Ångestsyndrom



I direktiven till den nya utredningen lyfts även andra uppgifter där det saknas en bra uppföljning i nationella register. Det gäller bl.a. ordination av sjukskrivning och utfärdande av sjukintyg samt väntetider i vården.

Ett annat område som lyfts är läkemedel. Sedan 2005 har vi ett nationellt register över alla uthämtade recept, men data saknas om läkemedel som administreras på sjukhus. Bättre uppföljning av dem är angeläget för att kunna följa upp alla nya biologiska läkemedel som ges på sjukhus.

Det återstår att se hur förslaget till utveckling av de svenska hälsodataregistren kommer att se ut när utredaren är klar. Såväl primärvårdsdiagnoser som rekvisitionsläkemedel på sjukhus har varit föremål för flera tidigare utredningar. Det som talar för en större framgång nu är att det hänt mycket med utvecklingen av elektroniska journalsystem, samt det ökade behovet att följa upp den omställning som sker i alla regioner med ökat fokus på primärvården.

Mängden klinisk information kommer sannolikt att bli begränsad om det blir ett nytt register, men sedan flera år finns redan Primärvårdskvalitet i drift där vården själva kan följa sina resultat och jämföra sig med andra liknande verksamheter. Primärvårdskvalitet kan inte generera nationella individdata för forskning, men omfattar ett stort antal kvalitetsindikatorer för akuta och kroniska tillstånd, psykisk ohälsa, rehabilitering samt kärnområden såsom kontinuitet, samsjuklighet och levnadsvanor. På registrets hemsida uppges att 89% av Sveriges vårdcentraler är med.

Referens

<https://skr.se/primarvardskvalitet.32919.html>
Forsslund T, Wettermark B. Primärvårdens roll i hälso- och sjukvårdssystemet. Hälso- och sjukvårdsförvaltningen, SLL, 2019

Hannah Brooke är ny i SVEP:s styrelse

Hannah Brooke, Inst för kirurgiska vetenskaper, Medicinsk epidemiologi, Uppsala universitet
E-post: hannah.brooke@uu.se

Hannah Brooke är ny ledamot i SVEP:s styrelse. Hon är dessutom ny redaktör för Svepet efter Eva Andersson som nu lämnar uppdraget. Här presenterar Hannah sig själv:

My name is Hannah Brooke and I am a new member of the SVEP committee. I completed my Bachelor's degree in Sports and Exercise Science at the University of Birmingham, UK. The program offered a broad education including everything from exercise physiology, basic biochemistry, anatomy, biomechanics, sports psychology, to nutrition for sports performance. I realised rather early on in the program that I was more interested in physical activity for its impact on population level health, than in trying to make elite athletes run a little faster or jump a little higher, and this led me towards epidemiology.

I got my first taste of conducting real research during the thesis project of my bachelors program and I thought it was really interesting. I enjoyed reading to understand what was known and what questions remained, the problem solving required to successfully conduct a study, and the analysis of data to get results that answered a question that no one else knew the answer to! That experience inspired me to pursue a career in research.

I completed a Masters and PhD in Epidemiology at the University of Cambridge, UK. The aim of my PhD was to examine the characteristics and changes in physical activity in the transition from childhood to adolescence, in order to improve interventions that aim to prevent the decline in physical activity observed during adolescence. The results of my work during this period suggested that targeting physical activity interventions at weekends and out of school may be valuable to promote physical activity maintenance into adolescence.

During my PhD I read about the amazing resources for epidemiological research in Sweden. So I pursued a post-doc at Karolinska Institutet, with Prof. Rickard Ljung as my mentor. The cohort data I was working with during my PhD included ~2000 school children with data at baseline and 4 years later, which was considered large, compared with previous studies in the field. So you can imagine my astonishment when I started working on my first paper using Swedish nationwide register data, and I realised the data



contained 100,682,794 person-years of follow-up accumulated by 7,821,897 individuals!

I have been largely working with Swedish national register data since 2015, mainly focusing on lifestyle and socioeconomic factors in relation to cancer and diabetes. I recently established my own subgroup, within Prof. Karl Michaëlsson's research group (Medical epidemiology) at the Department of Surgical Sciences, Uppsala University. We are based at Epihubben at Uppsala University, which provides a great opportunity for epidemiological discussions and exchange of ideas. I have great interest in epidemiological methods, and just now we are working a lot with longitudinal data analysis, including group-based trajectory modelling and joint latent class modelling. I have also been thinking recently about the application of machine learning methods within epidemiological research...watch this space!

I am really excited to join the SVEP board, and hope I can make a good contribution to the ongoing work.

Två spännande konferenser om läkemedel

Björn Wettermark, professor i Läkemedelsepidemiologi vid Uppsala universitet
E-post: bjorn.wettermark@farmaci.uu.se

Det finns många vetenskapliga konferenser av relevans för epidemiologer. Arbetar man med läkemedel så hör ICPE och EuroDURG till de konferenser man bör besöka. I somras fanns det möjlighet att delta i båda. I denna artikel presenteras några intryck från dem. Mer information och konferensprogrammen finns på hemsidan www.pharmacoepi.org

Läkemedelsepidemiologi kallas den del av epidemiologin som omfattar studier av användning och effekter av läkemedel i befolkningen. Läkemedelsepidemiologiska studier kan fokusera på att beskriva eller förstå hur läkemedel förskrivs eller används i olika grupper, eller analysera de positiva eller negativa effekterna av läkemedel i klinisk användning. Forskning om läkemedelsanvändningen brukar på engelska kallas för "Drug Utilization Research", medan forskning kring läkemedlens effekter traditionellt ofta kallas för farmakoepidemiologi. Det finns dock ingen tydlig gräns och allt kan betraktas som en del av samma forskning om läkemedel i användning. I ett temanummer av SVEPet kan man läsa mer om läkemedelsepidemiologi – vad det är, dess historia och framtid (SVEPet 2019; 37 (4)).

Faktorer som påverkar läkemedelsanvändningen	Läkemedelsanvändning	Positiva och negativa effekter av läkemedel
T.ex., Egenskaper hos patienter & läkare, sjukdomsförlopp, marknadsföring, media lagstiftning and förmånsbeslut	Hur läkemedel förskrivs, och används i olika befolkningsgrupper, t.ex. barn, äldre, gravida. Samt studier av kvaliteten i läkemedelsanvändningen	Positiva effekter som medicinska resultat och ökad livskvalitet, men också biverkningar som dödlighet, sjuklighet och sjukhusinläggningar
DRUG UTILIZATION RESEARCH	PHARMACOEPIDEMIOLOGY	

Den vetenskapliga förening som samlar flest läkemedelsepidemiologer globalt idag är International Society for Pharmacoepidemiology (ISPE). Föreningen har årliga konferenser – vartannat år i Nordamerika och vartannat år i Europa. Arbetet organiseras i olika s.k. Special Interest Groups (SIG) och Regional Interest groups (RIG). Antalet SIG uppgår idag till drygt 20 med olika fokus som cancer, biologiska läkemedel, äldre och läkemedel, sällsynta

sjukdomar, medicinteknik och vacciner. RIG är regionala nätverk för forskare inom ämnet. Eftersom organisationen är global så är de regionala nätverken breda och täcker världsdelar. Det finns även regionala konferenser där den asiatiska idag är den största. ISPEs årliga konferens – International Conference for Pharmacoepidemiology (ICPE) brukar samla kring 2000 deltagare och vara en mötesplats för personer som är verksamma inom akademien, regulatoriska myndigheter, läkemedelsindustrin och hälso- och sjukvården.



För forskare med huvudfokus på rationell läkemedelsanvändning fanns det sedan flera år innan ISPEs tillkomst flera andra vetenskapliga grupper kopplade till Världshälsoorganisationen WHO. Den mest kända av dem är WHO Drug Utilization Research group (WHO DURG) som grundades 1976, och som 1996 bytte namn till European Drug Utilization Research group – EuroDURG. Under många år arrangerade EuroDURG egna konferenser, men sedan 2006 gick EuroDURG ihop med ISPE och utgör idag den europeiska delen av ISPEs SIG för "Drug Utilization and Health Services Research". EuroDURG har dock fortfarande behållit sin tradition av att arrangera årliga konferenser.

Under sommaren 2023 arrangerade både EuroDURG och ISPE konferenser – i Bologna och Halifax. I denna artikel presenteras några "highlights" från de båda konferenserna. Mer information om de båda konferenserna kan man finna på ISPEs hemsida www.pharmacoepi.org.

EuroDURG 2023 i Bologna

I slutet av juni ägde EuroDURGs konferens rum i Bologna, Italien. Lite speciellt att komma till det som anses vara världens äldsta universitet. Bologna är fortfarande en stor lärdomsstad med mer än 90.000 studenter. Konferensen ägde rum i universitetets lokaler i gamla stan och samlade kring 250 deltagare.



Temat för konferensen var hållbar läkemedelsanvändning – “Sustainability of drug use: equity and innovation”. Temat utgick från FNs definition av hållbarhet – en utveckling som tillgodoser dagens behov utan att äventyra kommande generationers möjligheter att tillgodose sina behov. Hållbarhet är aktuellt inom alla samhällssektorer – men inte minst inom läkemedelsområdet där dagens läkemedelsanvändning är långt ifrån hållbar, vare sig ur ett ekonomiskt, socialt eller ekologiskt perspektiv. Dessa tre perspektiv genomsyrade den inledande plenarsessionen där Giampiero Mazzaglia (Italien), Aukje Mantel-Teeuwisse (Nederländerna) och Marmar Nekoro (Sverige) gav sin bild över vilka utmaningar vi står inför.

Därefter vidtog ett intensivt program med en blandning av föredrag, workshops och poster-sessioner på teman som e-hälsa, AI och machine learning, patientperspektiv på läkemedelsanvändningen, lärdomar från Covid-19-pandemin, äldres läkemedel och introduktion av nya behandlingar. En särskild session ägnades åt följsamhet – varför patienter inte tar sina läkemedel (på rätt sätt). Det är ett gammalt problem, men som ofta glöms bort i sjukvården. Det har dock skett en positiv

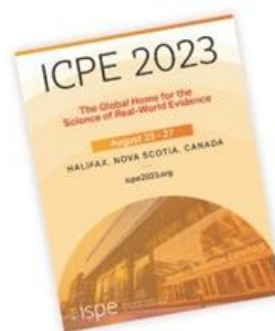
utveckling med ett stort antal digitala tjänster som kan hjälpa patienter att hantera sina läkemedel och följa upp sina behandlingsresultat. Just denna utveckling diskuterades intensivt under konferensen och under en särskild tredagars satellitutbildning som arrangerades av det europeiska COST Enable projektet.

Konferensen avslutades med en session med fokus på gap där det finns en stor sjukdomsburda och vi antingen saknar effektiva läkemedel eller att de av olika skäl inte når ut till patienterna. Vi fick också lyssna till några pionjärer, däribland Ulf Bergman från Sverige som gav sin bild på hur forskningen kring läkemedelsanvändningen började för 50 år sedan. Pionjärerna gav sina bästa råd till en panel yngre forskare som fick spekulera i hur läkemedel kommer att användas om 50 år, något som väckte flera tankar.

ICPE 2023 i Halifax

ISPEs stora konferens ägde rum i Halifax, Kanada i slutet av augusti. Temat var ”The Global Home for the Science of Real-World Evidence”, och konferensen var ganska allmänt hållen kring hur metodutvecklingen inom epidemiologin kan

hjälpa beslutsfattare inom myndigheter och sjukvård att göra bättre bedömningar av läkemedlens nytta och risker.



Det som var tydligt är att begreppen ”Real World Evidence” och ”Real World Data” nu kommit in och används brett inom vetenskapen. Varken RWE eller RWD är ju något nytt i sig utan bara vad epidemiologer ägnat sig åt under många år, men en positiv effekt av orden är att de väcker ökat intresse för epidemiologi. Och det kan behövas för att hantera den nya generationens biologiska läkemedel med cell- och genterapi som nu introduceras i vården.

Kurser och konferenser

Namn	Datum	Plats	Hemsida
EUPHA public health conference	8-11 november	Dublin, Irland	https://ephconference.eu/
Registerforskning och SweReg 2024	13-14 nov	Stockholm	Registerforskning 2023 & The Swedish Register-Based Research Summit: Start (trippus.net)
6th Swedish Meeting for Cohort Studies, in Uppsala 2023	20-21 november	Uppsala	https://www.epihealth.lu.se/en/calendar/6th-swedish-meeting-cohort-studies-uppsala-2023
15th annual NorPEN meeting	22-24 november	Oslo, Norge	http://www.norpen.org/
Swiss Epidemiology Winter School	15-20 januari	Wengen, Schweiz	https://www.epi-winterschool.org/
ICOH (International Conference on Occupational Health)	2-3 december	Sydney, Australien	https://waset.org/occupational-health-conference-in-december-2023-in-sydney
Amsterdam UMC & EpidM Winter Courses in Epidemiology	Jan - april 2024		https://www.epidm.nl/en/
31st Intensive Course in Applied Epidemiology	4-8 mars	Aberdeen, Skotland	https://www.abdn.ac.uk/iahs/research/epidemiology/icae-aberdeen-course-158.php
EuroCIM - causal inference meeting	17-19 april	Köpenhamn Danmark	https://www.eurocim.org
Sommarkurs i moderna metoder i biostatistik och epidemiologi	Juni	Treviso, Italien	http://www.biostatapi.org/
ISPE årlig konferens	24-28 augusti	Berlin, Tyskland	https://www.pharmacoepi.org/meetings/annual-conference/
ISEE årlig konferens	25-28 augusti	Santiago Chile	https://iseeconference.org/

Nästa nummer av Svepet:

”Kohorter och register”

Är du ansvarig för en kohort eller ett register?

Finns det intresse av att dela data från kohorten eller registret?

Ta i så fall kontakt med redaktionen för beskrivning av kohorten eller registret samt process och premisser för att få tillgång till data.

Kontakta oss senast 31 oktober:

hannah.brooke@uu.se eller elisabeth.strandhagen@gu.se

Tack!

SVEPET-redaktionen

c/o Hannah Brooke
Epihubben,
MTC-huset
75185 Uppsala

B

Porto betalt
Sverige