

SCB



Statistikmyndigheten och Socialstyrelsen ansvarar för många nationella register som kan användas för forskning.

Registerbaserad – en svensk paradgren

En månad in på Sveriges ordförandeskap i EU 2023 ordnade regeringen en konferens utanför Stockholm för att belysa tre viktiga områden inom cancersjukvården. Ett av dem var “Health data – a key to modern and equitable cancer care”. Övergången till en digitaliserad sjukvård skapar väldiga datamängder som kan användas för både sjukvård och forskning. I Sverige har vi tack vare våra personnummer och våra många nationella register med hög kvalitet och täckningsgrad sällsynt goda förutsättningar för att bedriva registerbaserad forskning i världsklass – inte minst cancerforskning.

CANCERREGISTRET OCH KVALITETSREGISTER

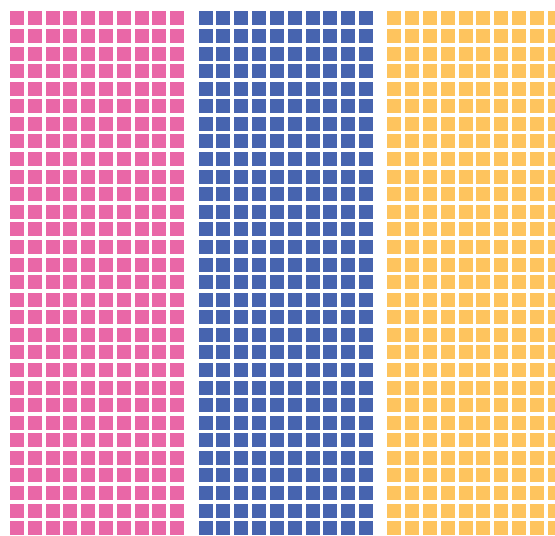
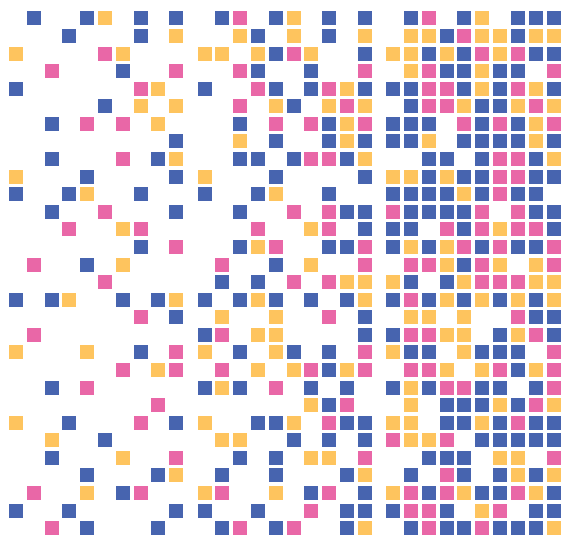
Sedan 1958 är läkare skyldiga att anmäla alla nydiagnostiserade cancerfall till *Cancerregistret*, som sköts av Socialstyrelsen. Uppgifterna i Cancerregistret omfattar personnummer, kön, hemort vid diagnos, anmälande sjukhus och klinik, diagnosdatum, klinisk och morfologisk diagnos, samt tumörutbredning (TNM) vid diagnostillfället. Eftersom dessa uppgifter är alltför knapphändiga för att utveckla sjukvården började nationella *kvalitetsregister för cancersjukdomar* byggas upp några decennier senare. Idag finns ett trettioital sådana, som var och ett administreras och stöds av något av de sex regionala cancercentrumen.

Omfattningen av vad som registreras i kvalitetsregistren varierar betydligt. I nästan alla finns detaljerade sjukdomskarakteristika vid tiden för diagnos och uppgifter om vilka undersökningar som ligger till grund för dem, samt primär be-

handling. Några innehåller uppgifter om uppföljning och återfall, andra inte. Allt fler registrerar någon form av patientrapporterade mått (PROM). Även täckningsgraden varierar, men de flesta har mycket hög täckningsgrad, åtminstone för de uppgifter som rapporteras in vid tiden för och månaderna efter diagnostillfället.

Kvalitetsregistrens primära syfte är att utvärdera och förbättra sjukvårdens kvalitet och utfallen för patienterna, men de kan också användas för forskning. Nästan all svensk registerforskning om cancer baseras på ett kvalitetsregister, oftast kompletterad med information från andra nationella register (se nedan). Antalet publicerade studier baserade på de svenska kvalitetsregistren för cancersjukvård ökar snabbt (Figur 1).

Rutinerna för datauttag för forskning varierar något mellan registren, beroende på vilket regionalt cancercentrum som administrerar det. Det går lätt att hitta uppgifter om dataut-



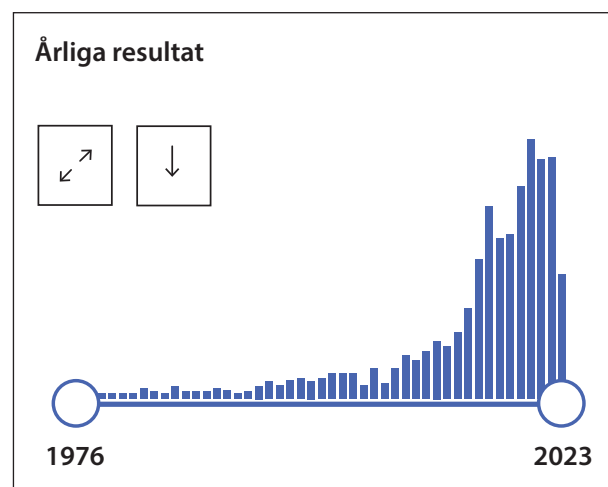
cancerforskning

tag på nätet (cancercentrum.se). I princip kan alla forskare med ett av Etikprövningsmyndighetens godkända forskningsprojekt få tillgång till specificerade uppgifter från ett kvalitetsregister efter en men- och sekretessprövning, men om forskarna själva inte aktivt arbetar med det aktuella kvalitetsregistret är det klokt att diskutera med registerhållaren eller styrgruppen och helst knyta någon av dem till forskargruppen. Utan kunskap om det enskilda registrets begränsningar är det risk för att resultaten och slutsatserna blir felaktiga. Många kvalitetsregister har gjort valideringsstudier som ger en uppfattning om registrets täckningsgrad och uppgifternas korrekthet. Om en sådan studie finns bör den refereras i de vetenskapliga publikationer som baseras på registret.

Utöver de nationella kvalitetsregistren för cancersjukdomar finns ett hundratal *nationella kvalitetsregister för benigna sjukdomar och ingrepp*. Vissa specifika forskningsprojekt kan fordra samkörning med ett sådant, till exempel om man vill studera cancerutveckling efter organtransplantation.

ANDRA NATIONELLA REGISTER

Statistiska centralbyrån (SCB, *Statistics Sweden*) har i *Registret över totalbefolkningen* uppgifter om en individ är levande, död eller har emigrerat (och förstås datum för död och emigration). Socialstyrelsen håller i *Dödsorsaksregistret*. Ett komplett dödsorsaksintyg fylls i efter 99 procent av alla dödsfall i Sverige, men särskilt för äldre och för personer med flera kroniska sjukdomar måste man vara medveten om att dödsorsaken ofta är osäker, eftersom endast en tiondel av de döda obduceras. Många gånger är det ju dessutom svårt att avgöra när en cancer verkligen är den bakomliggande dödsorsaken, till exempel när någon med spridd cancer dör av en lunginflammation.



Figur 1: En sökning på PubMed med "register AND cancer AND national AND Sweden" juli 2023 gav 1412 träffar. Sökningen gav utan tvekan många både falskt positiva och falskt negativa fynd, men den ger ändå en uppfattning om utvecklingen.

Det är upp till den intygsskrivande läkaren att avgöra vilken som är den bakomliggande dödsorsaken och denne har kanske mycket begränsad information om patienten och cancersjukdomen. Inte sällan anges en tidigare diagnostiserad cancer som dödsorsak för äldre individer som tynar bort av oklar anledning, även om canceren var kurativt behandlad eller av en beskedlig typ som lokaliserad, högt differentierad prostatacancer.

Utöver Cancerregistret och Dödsorsaksregistret finns hos Socialstyrelsen även Läkemedelsregistret, Patientregistret och

Medicinska födelseregistret. *Läkemedelsregistret* omfattar alla läkemedel som hämtats ut på recept på apotek från 2005, inklusive dosering. *Patientregistret* omfattar diagnoskoder satta vid läkarbesök i sluten och specialiserad öppenvård (inte primärvård), samt kirurgiska åtgärder. Kirurgiska åtgärder som gjorts på en operationsavdelning har för det allra mesta registrerats, medan ingrepp som utförts på en vanlig mottagning kan ha dålig täckning, särskilt diagnostiska sådana.

Flergenerationsregistret hos SCB innehåller samtliga individer ("indexpersoner") som är födda 1932 eller senare och har varit folkbokförda i Sverige någon gång sedan 1961. För indexpersonerna finns uppgifter om personnummer och födelseland för dem själva och deras föräldrar, inklusive om de är biologiska föräldrar eller adoptivföräldrar. Utifrån dessa primäruppgifter kan man få ut uppgifter om indexpersonens syskon och barn. Flergenerationsregistret ger fantastiska möjligheter att studera ärftlighetens betydelse för cancersjukdomar. I mitt avhandlingsarbete på 1990-talet försåg jag via behandlande urologläkare 400 män med nydiagnostiserad prostatacancer med ett frågeformulär om deras släktingars cancersjukdomar. Det var många brev som skulle skickas, inte minst påminnelsebrev eftersom läkaren ofta missade att överlämna formuläret. Ibland kunde patienten uppge personnummer för släktingar med en cancersjukdom så att vi kunde verifiera diagnosen via Cancerregistret, men oftast hade vi bara patientens uppfattning om diagnosen att gå på. I en studie med likande frågeställning 20 år senare använde vi Flergenerationsregistret. Den studien omfattade 52 000 bröder till 33 000 män med prostatacancer, vars cancerdiagnoser hämtades från Cancerregistret – en betydande förbättring jämfört med frågeformulären!

För forskning om arv kontra miljö som orsak till cancer kan man använda Svenska *Tvillingregistret*. Det finns vid Karolinska institutet och omfattar 87 000 tvillingpar med känd zygositet (en- eller tvåäggstvillingar).

Socioekonomiska faktorer påverkar hälsa, inklusive risk för cancersjukdomar och hur dessa diagnostiseras och behandlas. Man kan få socioekonomiska uppgifter från SCB på två olika sätt. Det ena är individuella uppgifter från Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (*LISA*). Det andra är uppgifter om det bostadsområde som individen bor i. SCB har delat in Sverige i 5 984 "demografiska statistikområden" (*DeSO*) med 700 till 2 700 invånare. Varje DeSO avgränsas i görligaste mån efter geografiska socioekonomiska skillnader. En fördel med DeSO är att uppgifterna inte är sekretessbelagda; de kan hämtas hem direkt från SCB på nätet. Detta innebär att man kan spara både tid och pengar jämfört med att begära ut data från LISA. En förutsättning är naturligtvis att forskargruppen har en statistiker som kan hantera DeSO-data. Socioekonomiska skillnader på områdesnivå (DeSO) har visat sig vara utslagsgivande särskilt i studier av screening, diagnostiska åtgärder och sjukdomskarakteristika vid diagnos.

De socioekonomiska variabler som oftast används vid cancerforskning är utbildning, inkomst och födelseland. Utbildning kategoriseras oftast som grundskola, gymnasium eller postgymnasial utbildning. För att kategorisera inkomst och födelseland på bästa sätt i förhållande till den aktuella stu-

diens frågeställningar krävs mer tankearbete; det finns väldigt många olika möjligheter.

ATT LÄNKA OLIKA REGISTER

Såväl Socialstyrelsen som SCB står till tjänst med att mot en avgift länka datafiler med personnummer till sina olika register. Avgiften står i förhållande till arbetsinsatsen och är oftast mellan några tiotusen och 100 000 kronor. Givetvis krävs först ett godkännande från Etikprövningsmyndigheten, där det tydligt ska framgå att länkningen planeras. Processen tar ofta flera månader, ibland upp till ett år. Det finns naturligtvis också möjlighet att länka till andra kvalitetsregister. För överföring av datafiler och för analyserna bör man anlita en statistiker med vana av sådant arbete.

FORSKNINGSDATABASER UTGÅENDE FRÅN KVALITETSREGISTER

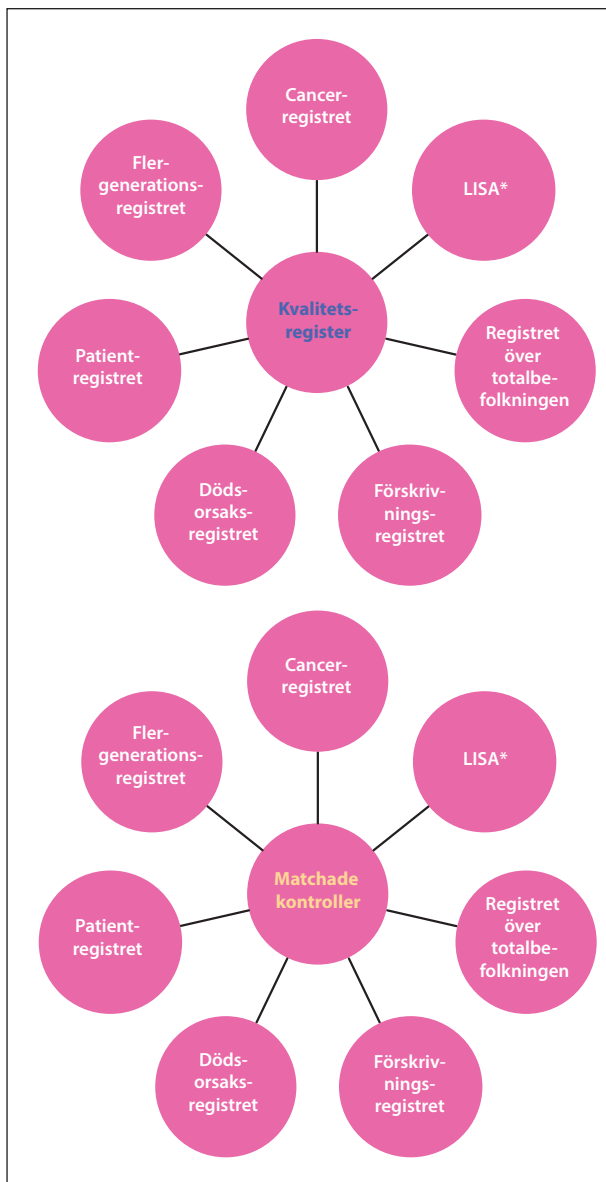
År 2008 skapade professor Pär Stattin (då Umeå, nu Uppsala universitet) tillsammans med statistiker, epidemiologer och kliniker PCBaSe, en forskningsdatabas som omfattar samtliga data i det nationella prostatacancerregistret (NPCR) och data från flera andra nationella register. Förutom data för alla individer med en prostatacancerdiagnos som registrerats i NPCR omfattar PCBaSe data för fem manliga kontrollindivider per prostatacancerpatient, matchade för ålder och läningsregion¹. PCBaSe är en låst databas med personnummer ersatta med löpnummer. Databasen uppdateras med några års mellanrum². Den nuvarande versionen PCBaSe 5.0 omfattar data för över 200 000 män med prostatacancer och över en miljon kontroller. Externa forskare har möjlighet att analysera selekterade data via en server som inte medger kopiering av datafilen. Sammanlagt har över 200 vetenskapliga artiklar publicerats från PCBaSe; dessa har citerats över 6 000 gånger.

PCBaSe har inspirerat andra forskare att bygga upp liknande databaser utgående från svenska kvalitetsregister för cancersjukvård. För närvarande finns sådana även för bröstcancer, kronisk myeloid leukemi, lungcancer, lymfom, myelodysplastiskt syndrom, njurcancer, peniscancer, tjocktarms- och ändtarmscancer och urinblåsecancer. Den principiella uppbyggnaden illustreras i Figur 2.

OLIKA TYPER AV STUDIEDESIGN

Det finns enormt mycket värdefull kunskap att vaska fram ur de ovan nämnda registren. Även enkla *deskriptiva studier* kan vara av stort värde, både inom och utanför Sveriges gränser. Exempel är rapporter om incidens, sjukdomskaraktäristika, utfall av diagnostiska undersökningar, behandlingar, prognos och mortalitet, samt geografiska och temporala skillnader för dessa variabler. Mer avancerade analyser inom ramen för *fallkontrollstudier* och *kohortstudier* kan blottlägga nya och kvantifiera kända riskfaktorer för olika cancerformer. Man kan identifiera *prognostiska* och *behandlingsprediktiva faktorer* och kanske skapa ett nomogram eller en algoritm för kliniskt bruk. För de senare är det nödvändigt att validera sin modell i en ny population, annars kommer den sannolikt att ge felaktigt underlag för kliniska beslut.

Att *jämföra olika behandlingars effekt* är ofta viktigt i sjukvården. Grundprincipen är att sådana jämförelser ska göras i en randomiserad klinisk studie, en RCT. Att planera, genom-



Figur 2: Principiell uppbyggnad av forskningsdatabaser som utgår från ett nationellt kvalitetsregister för en cancersjukdom. Det som förbinder de olika registren är personnumren. *LISA = Statistiska centralbyråns register över socioekonomiska faktorer (Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier).

föra och följa upp en RCT tar emellertid lång tid och kräver stora resurser, både ekonomiska och personella. Behandlingar för sällsynta tillstånd kan vara närmast omöjliga att utvärdera med en RCT. Inte sällan är en registerstudie den mest attraktiva möjligheten att utvärdera en ny behandling.

Med detta sagt finns det uppenbara risker för att resultaten påverkas av olika selektionsmekanismer och störfaktorer (confounders) så att slutsatserna blir felaktiga. När det är slumpen som avgör om patienten får behandling A eller behandling B kan vi med statistiska metoder beräkna graden av osäkerhet i resultaten; när det är läkaren och patienten som väljer behandling behöver vi även göra statistiska justeringar för selektionsbias, men för många sådana kan man aldrig justera fullt ut (se nedan) och vi kan naturligtvis inte alls justera för icke registrerade eller okända variabler.

Om man ändå till slut bestämmer sig för att genomföra en RCT bör man överväga möjligheten att använda *INCA* som studieplattform. Det finns flera fördelar med detta:

- Kostnaden är lägre.
- Personalen är van att använda mjukvaran.
- Uppgifter om sjukdomskaraktäristika kan hämtas direkt från kvalitetsregistret.

Ett aktuellt exempel på en RCT som baseras på INCA är Hypo-M1 (NCT0461290), som jämför olika fraktioneringar av strålbehandling mot primärtumören vid metastaserad hormonkänslig prostatacancer.

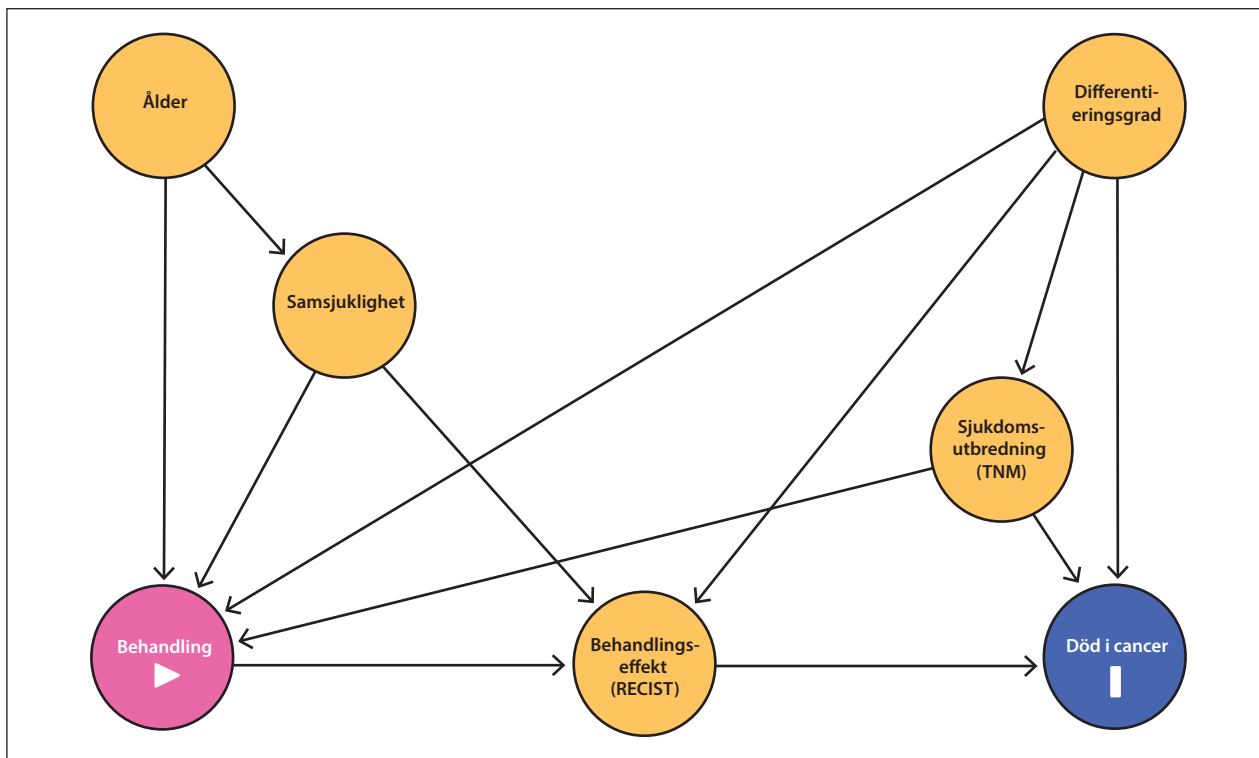
KORRIGERING FÖR STÖRFAKTORER OCH SELEKTIONSBIAS

En störfaktor är en variabel som kan påverka både orsaksvariabeln (till exempel exponeringen för en riskfaktor eller en behandling) och utfallet (till exempel att diagnostiseras med eller dö av cancer). Ett klassiskt exempel är äldre studier av kaffedrickande och cancer risk. Rökare drack oftare kaffe (orsaksvariabeln) än icke-rökare och rökning ökade risken för cancer (utfallet). Av detta kan man felaktigt dra slutsatsen att kaffedrickande ökar risken för cancer, när det i själva verket var rökningen som ökade cancer risken.

Selektionsbias innebär att någonting påverkar sannolikheten för att individer ska inkluderas i studiens olika grupper, till exempel att en patient ska få den behandling som studeras. Om man ska jämföra total överlevnad efter behandling A och behandling B och patienter med hjärtsvikt eller någon annan allvarlig sjukdom inte kan få behandling A, kommer de som får behandling A att i genomsnitt leva längre än de som får behandling B. Man kan också se detta som att hjärtsvikt är en störfaktor, men det kan vara lättare att identifiera relevanta faktorer som kan påverka utfallet om man tänker i termer av vad som styr vår selektion av patienter till den studerade åtgärden.

Poängen med att utvärdera effekten av en viss behandling med en randomiserad studie är att det inte blir någon selektionsbias i valet av behandling och att störfaktorerna med hög sannolikhet fördelas jämnt mellan de olika patientgrupperna. Om vi i stället ska utvärdera effekten av en behandling med en registerstudie måste vi på något sätt korrigera för selektionsbias och ojämn fördelning av störfaktorer. Det finns många olika sätt att göra detta på. Man kan göra stratifierade analyser, till exempel för stadium eller för behandlande sjukhus, men man kan bara stratifiera för några få variabler. Ett alternativ är att matcha, det vill säga att man väljer ut patienter som har fått behandling A som är så lika dem som har fått behandling B som möjligt. Matchning kräver ofta tämligen stora patientpopulationer, eftersom bara en del av patienterna ingår i analysen. Om man vill inkludera samtliga patienter kan man göra en justerad analys, det vill säga låta datorn använda matematisk statistik för att kompensera för skillnader mellan de patienter som fått behandling A och de som fått behandling B. Viktigt att tänka på är att antalet utfallshändelser (till exempel dödsfall) är den faktor som begränsar hur många variabler man kan justera för, inte antalet inkluderade patienter. Helst bör man ha minst tio negativa och positiva utfall i varje undergrupp (till exempel tio som dött och tio som lever).

När man har beräknat hur många störfaktorer man kan justera för är det dags att välja ut vilka man ska justera för. Här



Figur 3: En hypotetisk DAG (directed acyclic graph) ritad med det öppna programmet DAGitty. Programmet anger vilka av de variabler som ritats i grafen som man bör justera för. I exemplet ska man inte justera för behandlingseffekt, eftersom det inte är en störfaktor utan en mediator till effekten.

gäller det att tänka till. Om man matematiskt justerar för en variabel som inte är en störfaktor kan resultatet av analysen och därmed slutsatserna av studien bli felaktiga! För att välja ut vilka variabler man ska och inte ska justera för rekommenderas att rita en "kausal riktad acyklisk graf", en DAG (*directed acyclic graph*), förslagsvis med hjälp av det öppna programmet DAGitty (Figur 3). En introduktion till DAG kan läsas i *Journal of Surgical Research*³.

Det är också viktigt att fundera på vilka möjliga störfaktorer som det saknas data för. I registerstudier saknas till exempel ofta uppgifter om rökning, alkoholintag och övervikt. En annan hake är att man inte kan justera fullt ut för kategoriska variabler som har en gradient inom de olika kategorierna. Exempel: man kan justera för lokalt tumörstadium, T1-T4, men inom varje tumörstadium kan tumörerna vara olika stora. Om små T3-tumörer ofta behandlas med operation medan stora T3-tumörer alltid behandlas med strålning kommer effekten av strålbehandling att felaktigt framstå som sämre än effekten av operation, trots justering för lokalt tumörstadium.

Det finns naturligtvis många fler metodologiska fallgropar och möjligheter att undvika dem än de som nämnts ovan, bland annat hur man ska hantera saknade uppgifter. A och O är alltid att tänka igenom studieupplägget och analyserna noga innan man sätter i gång. Engagera gärna en statistiker redan under planeringsfasen. När man väl har sett studieresultatet får man inte formulera nya hypoteser utifrån de observerade skillnaderna!

SLUTORD

Vi kan använda våra kvalitetsregister och andra nationella register till värdefull forskning som kan förbättra livet för cancerpatienter både i och utanför Sverige. Men det är också lätt att gå vilse i de enorma datamängder som kan sammanställas. Många publikationer inger misstanke om att någon icke vetenskapligt skolad person har släppts lös i ett register (ofta lokalt) med ett "användarvänligt" statistikprogram. Epidemiologisk och statistisk kompetens är lika viktigt för registerforskning som anatomisk kunskap och färdighetsträning är för kirurgi! Slutligen kan betydelsen av väl genomtänkta val av studiedesign, val av studiepopulation och planering av analyser inte nog betonas. Som Tage Danielsson sa: "Säll är den som har till rättesnöre, att man bör tänka efter före."

REFERENSER

- Hagel E, et al. PCBaSe Sweden: a register-based resource for prostate cancer research. *Scand J Urol Nephrol*. 2009;43:342-9.
- Van Hemelrijck M, et al. Cohort Profile: The National Prostate Cancer Register of Sweden and Prostate Cancer data Base Sweden 2.0. *Int J Epidemiol*. 2013;42:956-67.
- Gongola AB & Bradshaw JC. Directed Acyclic Graphs in Surgical Research. *J Surg Res*. 2023;282:285-288.

OLA BRATT, PROFESSOR I KLINISK CANCEREPIDEMIOLOGI, SAHLGRENSKA AKADEMIN, GÖTEBORGS UNIVERSITET, OLA.BRATT@GU.SE

