

# Andragenerationsregister ny resurs för cancerforskningen

**Statistiska centralbyrån har utarbetat ett »andragenerationsregister» på över 6 miljoner individer organiserade i familjer. Detta har använts i genetisk epidemiologi av cancer genom att etablera Familjecancerdatabasen med hjälp av uppgifter från cancerregistret. Andragenerationsregistret kommer att vara en unik resurs för medicinskt genetiska studier.**

I Sverige finns unika möjligheter att genomföra studier på nedärvda sjukdomar: befolkningen är relativt homogen och samarbetsvillig, hälsovården är väl täckande och avancerad, hälsostatistiken är mångsidig och tillförlitlig. Nyligen har ytterligare en källa uppstått som ökar möjligheterna för genetisk forskning inom medicin: familjedatabas, »andragenerationsregister» som upprättats hos Statistiska centralbyrån (SCB). Den innehåller uppgifter om alla svenskar födda efter 1940 och deras biologiska föräldrar. Detta ger stora forskningsmöjligheter, till exempel inom cancerforskningen. Registret passar väl även för andra sjukdomar för vilka tillförlitliga diagnosdata finns tillgängliga från några decennier, helst täckande mer än en generation.

Kunskaper om genetiska faktorer i

## Författare

KARI HEMMINKI

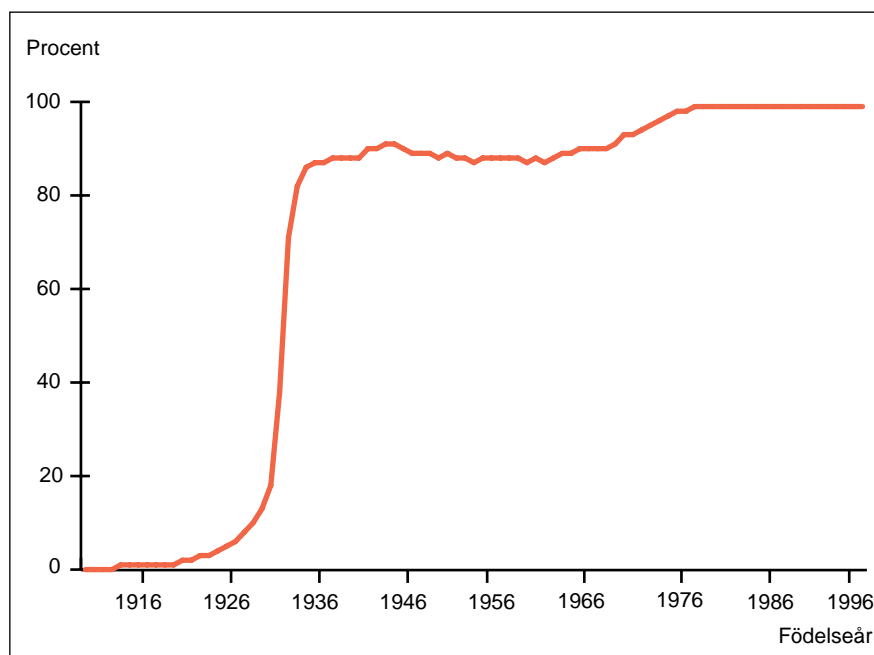
med dr, professor, institutionen för bioteknik vid Novum, Karolinska institutet, Huddinge

ÅKE NILSSON

fil kand, avdelningsdirektör, Statistiska centralbyrån, Örebro

PAULI VAITTINEN

fil kand, statistiker, institutionen för bioteknik vid Novum, Karolinska institutet, Huddinge, samt vid Epidemiologiskt centrum, Socialstyrelsen, Stockholm.



**Figur 1.** Andel personer efter födelseår för vilka det finns uppgifter om moderns personnummer. Födda 1935 eller senare har god täckningsgrad vad gäller föräldrarnas personnummer i andragenerationsregistret. Framtida förbättringar kommer att förskjuta kurvan åt vänster. För personer födda på 1940-, 50- och 60-talen torde täckningsgraden kunna förbättras endast marginellt. Det är nämligen för utrikes födda som föräldrarnas personnummer saknas. Flertalet av de utrikes födda har sina föräldrar boende i hemlandet, varför det inte finns något personnummer på föräldrarna.

cancer och andra sjukdomar har stor betydelse för diagnostik, rådgivning, förebyggande, hälsouppfostran och allmän uppfattning om risker. Familjedata och DNA-prov från drabbade familjer är även värdefulla för identifiering av cancerrelaterade gener och kvantifiering av risker som dessa gener medför. Gener för vissa nedärvda former av bröst- och tjocktarmscancer har identifierats och lett till kliniska framgångar i diagnostik och förebyggande [1].

## Andragenerationsregistret omfattar flera generationer

Inom den svenska folkbokföringen har sedan länge registrerats uppgifter om familjeförhållanden och släktskap. Tyvärr har släktskapsbanden mellan barn och föräldrar blivit tillgängliga på datamedia först under 1990-talet. Tidigare fanns uppgifterna endast i de manuellt förda personakterna. De uppgifter som nu finns på datamedia om föräldrarnas personnummer täcker dock inte hela befolkningen. I de fall då föräldrarna hade avlidit före den 1 juli 1991 (när folkbokföringen överfördes från kyrkan till staten) saknas föräldrarnas personnummer i folkbokföringens register. Vidare saknar invandrare som flyttat till

Sverige i vuxen ålder (18 år och äldre) uppgifter om föräldrarna.

Med uppgifter från folkbokföringen har SCB etablerat det så kallade andragenerationsregistret. Det innehåller uppgifter om eget personnummer, föräldrarnas personnummer och föräldrarnas födelseland och en del annan information.

Undan för undan har registret förbättrats med hjälp av annan registerinformation (Figur 1). Föddaregistret för åren 1961–1997 har använts för att komplettera med uppgifter om föräldrar. Likaså har folkräkningarna 1960 och 1970 använts i samma syfte.

En brist i dagens andragenerationsregister är bl a att personer som är adopterade har koppling till sina adoptivför-

**ANNONS**

**ANNONS**

**Tabell I.** Familjära risker (hazard ratios, FHR) för cancer hos söner och döttrar justerade för ålder och födelseår. CI = konfidensintervall. Källa: Hemminki et al, 1998.

Lokalisation	Söner			Döttrar			Tillsammans		
	N	FHR	CI 95 procent	N	FHR	CI 95 procent	N	FHR	CI 95 procent
Mage	11	1,6	(0,8–3,2)	11	2,3	(1,1–5,0)	22	1,9	(1,1–1,3)
Tjocktarm	43	2,5	(1,7–3,6)	51	2,1	(1,5–3,0)	94	2,3	(1,8–2,9)
Ändtarm	13	2,0	(1,0–4,0)	16	2,1	(1,1–3,9)	29	2,1	(1,3–3,3)
Lunga	44	1,7	(1,2–2,5)	38	2,6	(1,7–3,9)	82	2,0	(1,5–2,7)
Bröst	2	5,8	(0,6–1,4)	820	2,1	(1,9–2,3)	822	2,1	(1,9–2,3)
Cervix uteri				67	2,0	(1,5–2,6)			
Corpus uteri				27	4,1	(2,3–7,5)			
Äggstock				53	2,7	(1,9–3,7)			
Prostata	14	2,3	(1,1–4,9)						
Testikel	11	5,4	(2,6–11,3)						
Njure	3	1,6	(0,8–3,0)	10	2,6	(1,1–6,1)	24	1,9	(1,1–3,1)
Urinblåsa	14	1,0	(0,6–1,8)	11	2,2	(1,0–4,6)	25	1,3	(0,8–2,1)
Melanom	59	3,1	(2,3–4,3)	69	2,6	(2,0–3,5)	128	2,8	(2,3–3,5)
Hud	13	3,3	(1,6–6,8)	4	1,4	(0,5–4,4)	17	2,5	(1,4–4,6)
Nervsystem	48	2,0	(1,4–2,7)	42	1,8	(1,3–2,6)	90	1,9	(1,5–2,4)
Sköldkörtel	20	18,2	(8,5–39,1)	27	8,2	(4,7–14,1)	47	10,7	(6,9–16,6)
Endokrina körtlar	17	3,8	(2,0–7,0)	15	1,9	(1,0–3,4)	32	2,6	(1,7–3,9)
Lymfom	42	1,7	(1,2–2,4)	18	1,1	(0,7–1,9)	60	1,5	(1,1–2,0)
Leukemi	18	1,5	(0,9–2,5)	16	2,2	(1,3–4,0)	34	1,8	(1,2–2,6)

äldrar. Folkbokföringen har dock länkar till både biologiska föräldrar och adoptivföräldrar, och under 1999 kommer SCB också att införa länkar både till biologiska och adoptivföräldrar i andragenerationsregistret.

### Flera användningsområden

Det primära användningsområdet av registret var att avgränsa andragenerationsinvandrarna, dvs personer födda i Sverige med en eller båda föräldrarna födda utomlands. Andra användningsområden är:

- att sammanföra sambor med gemensamma barn till en familj i befolkningsstatistiken,
- att koppla samman vuxna hemmaboende barn med sina föräldrar i befolkningsstatistiken,
- att länka samman barn och föräldrar i olika undersökningar som förekommer av olika cancerformer, missbildningar, barns studiegång och betyg i förhållande till föräldrarnas utbildning, risker för olycksfall, bland barn beroende på föräldrarnas utbildning/etniska härkomst,
- att länka syskon för studier av cancerformer och missbildningar,
- att länka kusiner (kan än så länge bara göras för unga populationer).

### Vissa populationer saknas

I andragenerationsregistret saknas uppgifter om följande populationer:

- samtliga som avlidit/utvandrat före 1 november 1960,
- samtliga som avlidit/utvandrat under perioden 1 november 1960–hösten 1997 födda 1941 eller tidigare,
- personer som invandrat fram till hö-

sten 1992 och som avlidit/utvandrat före hösten 1992 (med undantag av personer födda 1941–1960 och som fanns i landet vid folkräkningen den 1 november 1960),

- personer vars båda föräldrar avlidit före 1 juli 1991,
- personer som invandrat till Sverige i vuxen ålder.

### Möjligheter till utveckling

Idéer om hur registret kan utvecklas finns. Något beslut har inte fattats ännu. Möjliga utvecklingslinjer är:

1. Att tillföra registret personer som utvandrat eller avlidit från och med 1 juli 1991. SCB har vänt sig till Riksskatteverket för att få kompletteringen genomförd men med negativt resultat. SCB har därefter vänt sig till regeringen.
2. I de fall som personnummer finns för en förälder kan personnumret för den andre föräldern hämtas från t ex en gammal folkräkning med hjälp av maka/makes personnummer villkorat med att vigseldatum är lägre än eget personnummer (föräldrarna har gift sig innan de fick barnet).
3. En mer krävande arbetsinsats men effektiv metod är att registrera uppgifter från personakter över avlidna och utvandrade. Flera landsarkiv har redan samlat in betydande material från pastorsämbeten. Kontakter har också tagits med Demografiska databasen i Umeå/Haparanda. Punkterna 1 och 2 torde kunna genomföras utan alltför krävande insatser. Efter länkningen av t ex ett patientregister är det möjligt att komplettera båda generationerna med inte ba-

ra utbildning och födelseland utan även inkomster, transfereringar, sysselsättning, boende, fertilitet, migration, etc.

### Utlämnande av uppgifter följer statistiksekretessen

Statistiksekretessen enligt 9 kap 4 § sekretesslagen (1980:100) är enligt huvudregeln absolut till sin karaktär med några undantag. Det finns således möjlighet att lämna ut uppgifter för forskningsändamål om det står klart att uppgifterna kan röjas utan att den som uppgiften rör eller någon honom närstående lider skada eller men.

När en förfrågan ställts till SCB om utlämnande av mikrodata för forskningsändamål gör SCB dels en bedömning av huruvida det är fråga om forskning, dels en skadeprövning. Med skada avser lagen utslutande ekonomisk skada och med men i första hand integritetskränkningar. Leder prövningen till en positiv bedömning kan uppgifterna lämnas ut. Huvudregeln är att SCB lämnar ut avidentifierade uppgifter. Utlämnande av personidentifierade data förutsätter alltid en speciell prövning.

Om utlämnandet av uppgifter för forskningsändamål sker till myndighet (t ex universitetsinstitutioner) följer statistiksekretessen med till mottagaren av materialet, som måste göra en sekretessprövning om materialet ska lämnas vidare.

Är forskaren knuten till en enskild person, ett företag eller en organisation löses sekretessen på annat sätt. Enligt 14 kap 9 § sekretesslagen kan man (SCB) uppställa förbehåll som gör att mottagaren bara får använda uppgif-

terna på det sätt som SCB föreskriver.

### **Cancerepidemiologi samverkan mellan arv och miljö**

Cancer orsakas av samverkande ärftliga faktorer och miljöfaktorer. Genetiska riskfaktorer påverkar förekomsten av vissa sällsynta former men också en andel av allmänna former av cancer. Bland de vanligaste cancerformerna där den ärftliga komponenten förekommer är bröst-, prostata- och tjocktarmscancer, ca 10 procent av dessa förklaras med ärftlighet. Segregationsanalyser för dessa cancerformer tyder på att ärftligheten sker genom autosomal dominant nedärvning baserad på sällsynta gener. Recessivt nedärvda tillstånd är svåra att identifiera i familjerna beroende på att cancer är en vanlig sjukdom, och familjedata kräver många generationer. Det typiska familjemönstret för recessiv nedärvning är att det finns sjuka individer i kärnfamiljen och inte hos andra släktingar. I de fall komplex (polygen eller multifaktoriell) nedärvning förekommer är den sannolikt ännu svårare att identifiera. Inom den molekylärgenetiska forskningen har ett växande antal kandidatgener identifierats för olika cancerformer [2]. Viktigt är också att många cancergener är defekta (muterade) både i ärftliga och sporadiska cancersjukdomar. Ärftlig cancer har ansetts som vetenskaplig inkörsport till mekanismer för all cancer.

Anhopning av cancer i vissa familjer har observerats för många cancerlokaliseringar, till exempel bröst, kolon, prostata, lunga och mage [3]. Tvillingar är en informativ undersökningspopulation för genetisk epidemiologi men har endast använts i begränsad utsträckning inom cancerepidemiologi [4, 5].

Huvuddelen av resultat på nedärvda risker i cancer är baserade på data från selekterade familjer med hög förekomst av cancer eller på intervjuer med cancerpatienter. Även om sådana studier kan ha systematiska fel, råder det ingen tvekan om att genetiska faktorer har betydelse för uppkomsten av vissa cancersjukdomar, även befolkningsbaserade undersökningar har kommit till motsvarande resultat [6].

Den enda databas som motsvarar den svenska Familjecancerdatabasen finns i Utah, USA, där delstatens populationsdatabas på mormoner, cancerregister och dödsorsaksregister använts för genetisk epidemiologi [6] och framgångsrikt för genidentifiering [2].

Det finns stora förväntningar på att den molekylärgenetiska forskningen ska göra det möjligt att finna genotyper som gör vissa individer särskilt känsliga för exponering för miljömässiga riskfaktorer för cancer, dvs man räknar

med att kunna kartlägga interaktion mellan genetiska faktorer och miljömässiga riskfaktorer. Identifiering av nya gener involverade i cancer eller andra sjukdomar beror i mycket hög grad på tillgänglighet av DNA från individer som tillhör informativa familjer [7, 8]. Tillgång till familjer har t ex haft en stor betydelse för kartläggning av tumörsuppressorgener i bröstcancer, melanom, prostata- och hudcancer [2].

### **Resultat av samkörda register**

Den avidentifierade Familjecancerdatabasen bildades efter samkörning av andragenerationsregistret och det svenska cancerregistret [9]. I sin nuvarande form täcker den individer födda i Sverige efter 1941 med sina biologiska föräldrar, totalt ca 6 miljoner personer länkade till cancerdata från diagnosåren 1958–1994, över en halv miljon fall. Jämfört med Utahdatabasen har Familjecancerdatabasen ett mycket större (cirka tio gånger större) befolkningsunderlag och cirka tio gånger flera cancerfall [6, 9–11]. Familjecancerdatabasen har omfattat två generationer men ju längre uppföljningstiden är, desto flera cancerfall förekommer det i den äldrande tredje generationen. Detta kommer att öka styrkan i alla genetiska analyser, eftersom tre generationer och kusiner blir tillgängliga. Utveckling av andragenerationsregistret till äldre födelsekohorter pågår hos SCB.

### **Åldersspecifik analys**

De första analyserna av Familjecancerdatabasen utfördes i början av 1997, och i slutet av 1997 gjordes komplettering med information på individer som hade dött före 1992 och med två ytterligare år av cancerfall från cancerregistret. Åldersspecifika familjära risker har räknats för all cancer, primär och sekundär bröstcancer, in situ-cancer och familjer med flera drabbade individer [10–14]. Dessutom har analyser av effekt av föräldrarnas ålder på cancerrisk hos barn utförts på olika cancerformer inklusive barndomscancer.

Tabell I visar familjära risker för avkomma i åldrarna mellan 15 och 53 år, vilkas föräldrar har en viss typ av cancer [11]. Risktalen jämför avkommas risk för samma cancer som föräldern haft med risken för denna cancer hos övrig avkomma, dvs de vilkas föräldrar inte hade cancer i fråga eller var cancerfria. De största familjära riskerna (»familial hazard ratio», FHR>5) konstaterades i sköldkörtel- och testikelcancer. Intermediära risker (FHR 2–5) påträffades för cancer i tjock- och ändtarmen, lunga, bröst, cervix, uterus, äggstock, hud (både melanom och skiv-epitelcancer) och endokrina körtlar. Lägre FHR förekom för cancer i mage,

njure och nervsystem samt för lymfom och leukemi. Både miljömässiga och hereditära faktorer kan bidra till familjära risker. Vissa könsskillnader förekom, t ex familjär bröst- och sköldkörtelcancer visade något större risk hos män än hos kvinnor. Låg diagnosålder hos föräldrar eller barn var ytterligare en riskfaktor för många familjerelaterade cancerformer.

### **Värdefull källa för biomedicinsk forskning**

Det svenska andragenerationsregistret kommer att vara en värdefull källa för framtida biomedicinsk forskning på grund av storlek, systematisk uppläggning och täckningsgrad. Exempel på cancerområden visar registrets möjligheter för genetisk epidemiologi. Viss utveckling av registret genom komplettering av information om avlidna, in- och utvandrade och äldre generationer förstärker forskningspotentialen ytterligare.

### **Referenser**

1. Ponder B. Genetic testing for cancer risk. *Science* 1997; 278: 1050–4.
2. Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. *Science* 1997; 278: 1043–50.
3. Lynch HT, Fusaro RM, Lynch J. Hereditary cancer in adults. *International Society for Preventive Oncology* 1995; 19: 219–33.
4. Pedersen NL, Lichtenstein P, de Faire U, Ahlbom A, Floderus B, Svartengren M. En (inter)nationell resurs. Svenskt tvillingregister ger upplysning om miljöns och arvets betydelse vid sjukdom. *Läkartidningen* 1996; 93: 1127–30.
5. Ahlbom A, Lichtenstein P, Malmström H, Feychting M, Hemminki K, Pedersen N. Genetic and familial risk factors for cancer in twin. *Journal of the National Cancer Institute* 1997; 89: 287–93.
6. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population based assessment of cancer risk in first degree relatives of cancer probands. *Journal of the National Cancer Institute* 1994; 86: 1600–8.
7. Collins F. Positional cloning moves from perditional to traditional. *Nature Genetics* 1995; 9: 347–50.
8. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; 265: 2037–48.
9. Hemminki K, Vaittinen P. National database of familial cancer in Sweden. *Genet Epidemiol* 1998; 15: 225–36.
10. Hemminki K, Vaittinen P. Familial breast cancer in the Family-Cancer Database. *Int J Cancer* 1998; 77: 386–91.
11. Hemminki K, Vaittinen P, Kyyrönen P. Age-specific familial risks in common cancer of the offspring. *Int J Cancer* 1998; 78: 172–5.
12. Hemminki K, Vaittinen P. Familial risks in in situ cancers from the Family-Cancer Database. *Cancer Epidemiol Biomarkers Prev* 1998; 7: 865–8.
13. Hemminki K, Vaittinen P, Kyyrönen P. Modification of cancer risk in offspring by parental cancer. *Cancer Causes Control* 1998; 10: 125–9.
14. Vaittinen P, Hemminki K. Familial cancer risks in offspring from discordant parental cancers. *Int J Cancer* 1999; 81: 12–9.