

CHALMERS



GÖTEBORGS UNIVERSITET

## Hata Slumpen

–När bör miljön undersökas på grund av en sjukdoms-  
anhopning?

*Kandidatarbete inom civilingenjörsutbildningen vid Chalmers*

Malin Boström

Sara Emanuelsson

Linus Eriksson

Johan Wiebe



# Hata Slumpen

–När bör miljön undersökas på grund av en sjukdomsanhopning?

*Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik  
vid Chalmers*

Malin Boström   Sara Emanuelsson  
Linus Eriksson   Johan Wiebe

Handledare:   Staffan Nilsson  
Examinator:   Carl-Henrik Fant

Institutionen för matematiska vetenskaper  
Chalmers tekniska högskola  
Göteborgs universitet  
Göteborg 2012



## Sammanfattning

Egendomliga sammanträffanden kan vara viktiga att notera och uppmärksamma om man vill upptäcka miljöproblem. En ovanligt stor ansamling av sjukdomar inom ett litet område eller inom en liten grupp kan ge anledning till oro i omgivningen. Målet med det här projektet var att med statistiska metoder undersöka hur osannolika sådana ansamlingar är och om det går att avfärda dem som slumpmässiga händelser eller inte, samt att konstruera modeller vars syfte är att identifiera hälsofarliga miljöer.

Endast icke smittsamma sjukdomar har studerats i detta projekt och de formler som har konstruerats bygger framför allt på Bayes sats. I de beräkningar som gjorts har det antagits att antalet sjukdomsfall är binomialfördelade.

För att kunna beräkna risken för en skadlig miljöfaktor givet ett antal sjuka behövs en a priori skattning för sannolikheten att denna miljöfaktor finns och kunskap om hur mycket den höjer risken för sjukdom. Dessa värden är såklart osäkra, särskilt om det är okänt vilken miljöfaktor det skulle kunna röra sig om. Det kan få till följd att den beräknade risken innehåller allt för stor osäkerhet för att användas i praktiken. Det är alltså inte alls säkert att det är ekonomiskt lönsamt att utföra sådana beräkningar, vilket leder till slutsatsen att bästa sättet att upptäcka riskfyllda miljöer är att genomföra en undersökning om det finns en oro bland allmänheten.

## Abstract

Strange coincidences may be important to notice and pay attention to if you want to identify hazardous environments. An unusually large cluster of illnesses in a small area or within a small group can give rise to concern in the community. The aim of this project was to use statistical methods to examine how unlikely such clusters are and if it is, or is not, possible to reject them as random events and to construct models whose purpose is to identify hazardous environments.

Only non-contagious diseases have been studied in this project and the equations which has been designed is primarily based on Bayes' theorem. In the calculations performed it has been assumed that the cases of illnesses are binomially distributed.

In order to calculate the risk of a harmful environmental factor, given a number of ill persons, a prior estimate for the probability that this factor exists and knowledge about how much it increases the risk of disease is needed. These values are of course uncertain, especially if it is unknown which environmental factor it could be. As a result the calculated risk can include too much uncertainty to be used in practice. It is thus not certain that it is economically viable to do these calculations which lead to the conclusion that the best way to detect environmental hazards is to perform an investigation if there is concern among the public.

# Innehåll

<b>1</b>	<b>Inledning</b>	<b>1</b>
1.1	Bakgrund . . . . .	1
1.2	Syfte . . . . .	1
1.3	Problemformulering . . . . .	1
1.4	Metod . . . . .	2
1.5	Avgränsningar . . . . .	2
1.6	Rapportens uppbyggnad . . . . .	2
1.6.1	Teori . . . . .	2
1.6.2	Konstiga händelser – <i>Företeelser som verkar osannolika</i> . . . . .	2
1.6.3	Okända miljöfaktorer . . . . .	2
1.6.4	Kända miljöfaktorer – <i>När en specifik miljöfaktor misstänks</i> . . . . .	3
1.6.5	Exempel med radon som skadlig miljöfaktor . . . . .	3
1.6.6	Tillvägagångssätt i praktiken . . . . .	3
1.6.7	Diskussion och Slutsatser . . . . .	3
<b>2</b>	<b>Teori</b>	<b>4</b>
2.1	Betingad sannolikhet och Bayes sats . . . . .	4
2.2	Sannolikhetsuppskattningar . . . . .	4
2.3	P-värde och signifikansnivå . . . . .	4
2.4	Multipel inferens . . . . .	5
2.5	Binomialfördelning . . . . .	5
2.6	Relativ risk . . . . .	5
<b>3</b>	<b>Konstiga händelser</b> – <i>Företeelser som verkar osannolika</i>	<b>6</b>
<b>4</b>	<b>Okända miljöfaktorer</b>	<b>9</b>
4.1	Modell . . . . .	9
4.2	Test av modell . . . . .	10
<b>5</b>	<b>Kända miljöfaktorer</b> – <i>När en specifik miljöfaktor misstänks</i>	<b>14</b>
5.1	Modell . . . . .	14
5.2	Antagande om binomialfördelning . . . . .	16
5.3	Olika sannolikhet för olika grupper . . . . .	17
5.4	Olika sannolikheter för olika grupper med antagande om binomialfördelning . . . . .	17
5.5	Flera miljöfaktorer . . . . .	18
<b>6</b>	<b>Exempel med radon som skadlig miljöfaktor</b>	<b>19</b>
6.1	Introduktion till problemet . . . . .	19
6.2	Den enklaste modellen . . . . .	19
6.3	Lite mer komplicerad modell . . . . .	21
6.4	Flera möjliga radonhalter . . . . .	21
<b>7</b>	<b>Tillvägagångssätt i praktiken</b>	<b>24</b>
7.1	Hur gör de på Chalmers? . . . . .	24
7.2	Kommunal nivå . . . . .	24
7.3	VMC . . . . .	25
<b>8</b>	<b>Diskussion och Slutsatser</b>	<b>27</b>
8.1	Konstiga händelser . . . . .	27
8.2	Miljöfaktorer . . . . .	27
8.3	Tillvägagångssätt i praktiken . . . . .	29
8.4	Slutsatser . . . . .	29

<b>Referenser</b>	<b>30</b>
<b>A Radon</b>	<b>33</b>
A.1 Skada . . . . .	33
A.2 Riskökning . . . . .	33
A.3 Förekomst . . . . .	33
A.4 Mätning och åtgärder . . . . .	34
<b>B Kod</b>	<b>35</b>
B.1 Exempel med polishuset i Hagfors . . . . .	35
B.2 Okända miljöfaktorer . . . . .	37
B.3 Exempel med radon som miljöfaktor . . . . .	38

## Förord

Detta projekt har inte haft någon tydlig uppdelning utan vi har till stora delar arbetat tillsammans. Alla har bidragit lika mycket till planering, metodval, genomförande, modellkonstruerande, problemlösning med mera. I princip går det att säga att alla har skrivit, eller på annat sätt hjälpt till, med alla avsnitt. Det innebär att även om en person står som huvudförfattare till ett avsnitt betyder det inte att övriga gruppmedlemmar inte har skrivit på det. I korthet kan dock uppdelningen läsas i tabell 1.

Tabell 1: Huvudförfattare till de olika avsnitten. Notera att detta är en mycket grov uppskattning och att alla har hjälpt till med alla avsnitt.

Avsnitt	Huvudförfattare
Sammanfattning/Abstract	Johan
1 Inledning	Alla
2 Teori	Malin
3 Konstiga händelser – <i>Företeelser som verkar osannolika</i>	Malin och Johan
4 Okända miljöfaktorer	Johan
5 Kända miljöfaktorer – <i>När en specifik miljöfaktor misstänks</i>	Sara
6 Exempel med radon som skadlig miljöfaktor	Sara
7 Tillvägagångssätt i praktiken	Linus
8 Diskussion och Slutsats	Alla
A Radon	Sara
B Kod	Malin, Sara och Johan

Under projektets gång har vi skrivit en dagbok och dessutom fört varsin tidslogg över hur mycket tid vi har lagt ner på projektet.

Till sist skulle vi vilja tacka vår handledare Staffan Nilsson för stort stöd under hela arbetets gång.



# 1 Inledning

## 1.1 Bakgrund

Det finns många områden där det är viktigt att kunna avgöra om en händelse kan betraktas som slumpmässig eller om den kräver andra förklaringar. Till exempel kan risken för att drabbas av en given sjukdom påverkas av den miljö man vistas i. Därför är det viktigt att upptäcka riskfyllda miljöfaktorer, men att mäta allt överallt är uppenbart omöjligt. Således kan det finnas ett behov av att identifiera de platser där sannolikheten för en viss riskfylld faktor är hög för att kunna genomföra relevant mätning och eventuella åtgärder.

Ett exempel är radon i byggnader som ökar risken för lungcancer[1]. För att undvika fler fall är det viktigt att finna hus där radonhalten är över gränsvärdet  $200 \text{ Bq/m}^3$ [1]. Om till exempel två personer på samma arbetsplats drabbas är det konstigt, men hur konstigt? Hur osannolikt är det att två personer på någon arbetsplats, eller i någon grupp av en viss storlek drabbas? När går det inte längre att avfärda som slump? När bör ansvariga myndigheter göra en mätning?

Det kan också vara att ett större geografiskt område är mer drabbat än sin omgivning. Om det till exempel är mycket högre andel som blir sjuka i stadsdelen Majorna än i resten av Göteborg kan det vara något som bör uppmärksammas. Även här kan det bero på miljöfaktorer som kan vara viktiga att åtgärda.

## 1.2 Syfte

Syftet med kandidatarbetet är att utifrån statistiska metoder skapa modeller för att avgöra om det är möjligt att avfärda en anhopning av sjukdomsfall som slumpmässig eller om det kan antas finnas andra orsaker samt avgöra om en åtgärd är att rekommendera. Modellerna skall sedan jämföras med hur liknande frågor hanteras i praktiken.

## 1.3 Problemformulering

Det går att dela in problemet i två delproblem. Det första delproblemet är de tillfällen då det är okänt vad, eller om det ens finns någonting, som höjer risken för nya sjukdomsfall, till exempel om ovanligt många har insjuknat på en arbetsplats. Problemet här består i att med hjälp utav statistiska test avgöra om det går att förkasta att en ansamling sjukdomsfall beror på slumpen samt att konstruera en modell för att beräkna sannolikheten att det existerar en hälsofarlig miljö i en omgivning givet ett kluster av sjukdomsfall. Det kan även behöva göras skattningar av värden hos problemet innan beräkningarna börjar.

Den andra delen handlar om när det är känt vilken miljöpåverkan det skulle kunna vara. Det finns då två olika scenarion, att det är känt att det finns en viss miljöpåverkan och då är frågan om ett sjukdomsfall har orsakats av miljön eller om det fortfarande kan bero av slumpen eller någon annan påverkan. Det andra scenariot är när det finns misstankar om en viss miljöpåverkan men det är osäkert om den verkligen finns där och då är frågan hur sannolikt det är att en sådan miljö existerar. Ett exempel som kommer betraktas är förhållandet mellan lungcancer och radonexponering i byggnader.

Då en del värden i beräkningarna kommer att vara okända behöver det göras skattningar av dem. När de görs bygger de på vad som anses passande och det finns ingen ordentlig grund för hur de ska sättas. I arbetet kommer även skillnaden i resultat givet olika skattningar att jämföras.

Även om en viss händelse är osannolik, är det eventuellt inte lika osannolikt att det sker en lika extrem händelse någonstans i landet. För att reglera detta kommer något som kallas för multipel inferens att användas. Här uppkommer problemet om i vilket perspektiv det inträffade ska betraktas.

## 1.4 Metod

Statistiska metoder och sannolikhetslära, framförallt bayesiansk sannolikhetslära, kommer att utgöra den matematiska grunden för det här projektet. Eftersom utgångspunkten inte är färdiga modeller kommer en stor del av arbetet med att ta fram modeller bestå av gemensamma diskussioner kring rimlighet i olika metoder utifrån kända fakta. Det konstrueras först enklare modeller för att sedan utvecklas med fler faktorer som till exempel att somliga personer har större risk att utveckla en given sjukdom etc. När modellerna utvecklas och blir mer komplexa är det rimligt att studera den litteratur som finns på området genom sökningar på vetenskapliga artiklar i tillgängliga databaser.

Vid konstruerande av grafer och då mer avancerade beräkningar behöver genomföras kommer MATLAB och Mathematica att användas.

Fakta om hur det går till i praktiken kommer att fås genom intervjuer med arbetsmiljöexpertis och informationssökning. Det här görs först efter att ett par modeller tagits fram då syftet främst är att konstruera en egen modell och inte bygga vidare på en redan tillgänglig sådan.

## 1.5 Avgränsningar

Vid skapandet av modellerna kommer en del förenklingar att göras eftersom informationen som ges inte alltid stämmer helt överens. Vid multipel inferens krävs relevant gruppindelning vilket alltid kommer att vara en avgränsning eftersom alla grupper inte ser exakt likadana ut. Gruppindelning och anpassning till den information som finns till hands är en viktig del då modellerna anpassas till konkreta exempel. Ingen gruppering kommer att täcka mer än Sverige och det på grund av att de som kontaktas verkar i Sverige samt att aktuella siffror erhålles enklare.

I alla beräkningar kommer förenklingar göras då indata väljs. Till exempel antas män och kvinnor löpa lika stor risk att drabbas av cancer. Alla dessa förenklingar specificeras där de görs.

Endast icke smittsamma sjukdomar kommer att betraktas. Det betyder att sannolikheten för att en person blir sjuk inte påverkas av vilka andra personer som drabbas. Dessutom antas att risken att bli sjuk av andra orsaker än en viss skadlig miljöfaktor inte påverkas av huruvida denna miljöfaktor finns eller ej. Det innebär att den skadliga miljöfaktorn inte är korrelerad med andra riskfaktorer.

## 1.6 Rapportens uppbyggnad

Nedan följer en beskrivning av rapportens uppbyggnad och de olika avsnittens funktion för rapporten.

### 1.6.1 Teori

I teoridelen redovisas den teori som används i rapporten. Bland annat finns det en genomgång av Bayes sats och av hur en lämplig signifikansnivå väljs vid multipel inferens.

### 1.6.2 Konstiga händelser

*–Företeelser som verkar osannolika*

I avsnitt 3 ges ett exempel på när det inom en grupp inträffar "ovanligt" många sjukdomsfall. Exemplet ses på ur ett grundläggande perspektiv och den matematik som används är hypotestest och även multipel inferens.

### 1.6.3 Okända miljöfaktorer

I avsnitt 4 börjar modellen ta form. Här behandlas fallen då det är osäkert om det finns någon påverkande miljöfaktor och, om en sådan finns, okänt vilken miljöfaktor det handlar om. Exemplet som användes i föregående avsnitt kan nu appliceras på modellen och en djupare analys kan sedan göras. Här kommer det att behövas göras olika skattningar av sannolikheter.

#### **1.6.4 Kända miljöfaktorer**

##### ***-När en specifik miljöfaktor misstänks***

När det är känt vilken miljöfaktor det handlar om går det att räkna fram mer från den data som finns tillgänglig. En modell anpassad till detta kommer att skapas i avsnitt 5.

#### **1.6.5 Exempel med radon som skadlig miljöfaktor**

När det finns vetskap eller misstankar om radonexponering är det ett tydligt exempel på en känd miljöfaktor. För att få några värden att testa på har därför ett radonexempel konstruerats där modellen i avsnittet om kända miljöfaktorer används för att göra olika beräkningar baserade på data för radonexponering

#### **1.6.6 Tillvägagångssätt i praktiken**

I avsnitt 7 jämförs modellen med hur liknande frågor hanteras på riktigt. Var och hur tas besluten om eventuella mätningar och åtgärder är några frågor som behandlas i avsnittet. Kunskapen bygger på intervjuer samt information som går att finna hos olika myndigheter.

#### **1.6.7 Diskussion och Slutsatser**

I det här avsnittet diskuteras de olika resultaten som framkommit i de tidigare delarna av rapporten och slutsatser dras.

## 2 Teori

Nedan följer den grundläggande teori som används i arbetet. Förutom Bayes sats finns det även en beskrivning av a priori skattningar, vad p-värdet är för något och dess koppling med signifikansnivån  $\alpha$ . Dessutom finns det en förklaring till hur multipel inferens går till och när det behöver användas. Sist kommer teori som är mer anpassad till sjukdomsrelaterade problem vilket kommer att vara användbart i arbetets modeller.

### 2.1 Betingad sannolikhet och Bayes sats

Den matematiska grunden för projektet består av bayesiansk sannolikhetsteori som kretsar kring Bayes sats vilken visas i ekvation (1).

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{P(B)} \quad (1)$$

där  $A_1, A_2, \dots, A_k$  är en partition och  $B$  är en händelse med  $P(B) > 0$

$P(B)$  bestäms med hjälp av lagen för total sannolikhet som ges av ekvation (2). Bayes sats ger sannolikheten för att en händelse  $A_i$  inträffar då man vet  $B$ .

$$P(B) = \sum_j P(B | A_j)P(A_j) \quad (2)$$

Det kan även vara värt att påminna om att för betingad sannolikhet gäller att:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3)$$

### 2.2 Sannolikhetsuppskattningar

Ibland är vissa värden okända av olika anledningar. För att då kunna göra beräkningar krävs det att en a priori skattning av dessa värden görs. I de fall där det inte finns någon information alls ska sannolikheten delas lika på de olika utfallen som kan ske. Om det däremot finns kunskap kan fördelningen bli annorlunda, eftersom vissa utfall kan ses som mer troliga än andra. Ett exempel är om någon person läspar i en grupp av  $n$  personer. Om det saknas kunskap om gruppen kommer en a priori skattning innebära att alla individer har en sannolikhet på  $1/n$  per person att den personen läspar. Men om det däremot är känt att den läspande personen talar engelska och att halva gruppen är engelsktalande kan den skattningen vara en sannolikhet på  $2/n$  för de engelsktalande individerna att den personen läspar[2].

I vissa fall kan det ge olika resultat beroende på vilken skattning som väljs, men dessvärre måste det göras någon skattning. Det kan vara svårt att avgöra hur mycket ett värde ska viktas och det kan vara svårt att veta vilka konsekvenser som blir till följd av de olika valen. Då kan det underlätta att med grafer eller tabeller försöka se om det blir en märkbar skillnad i resultatet beroende på vilken skattning som väljs.

### 2.3 P-värde och signifikansnivå

Vid statistiska tester är det så kallade p-värdet användbart. P-värdet är ett mått på hur sannolikt det skulle vara att få det observerade värdet eller ett ännu mer extremt värde givet att nollhypotesen  $H_0$  gäller[3]. Vid hypotestestning är signifikansnivån  $\alpha$  gränsen för vad p-värdet måste underskrida för att nollhypotesen ska kunna förkastas. Ett lågt  $\alpha$  medför att om nollhypotesen ska kunna förkastas krävs det att testet ger ett resultat som är tillräckligt osannolikt enligt nollhypotesen. Ofta är  $\alpha = 0.05$  vilket betyder att i 5% av testerna kommer en sann nollhypotes förkastas.

## 2.4 Multipel inferens

När många hypoteser  $H_{01}, \dots, H_{0n}$  testas med en viss signifikansnivå  $\alpha$  ökar sannolikheten för att någon sann hypotes förkastas. Om  $H_{01}, \dots, H_{0n}$  är oberoende av varandra kommer sannolikheten att öka markant. Om till exempel  $\alpha = 0.05$  och tio hypoteser testas, det vill säga  $n = 10$ , och samtliga tio nollhypoteser är sanna blir sannolikheten att inga nollhypoteser förkastas  $1 - (1 - \alpha)^n = 1 - 0.95^{10} = 0.40$  vilket är ett orimligt högt tal. Det förväntade antalet sanna hypoteser som förkastas blir  $\alpha n = 0.05 \cdot 10 = 0.5$ [4]. För att lösa problemet införs istället något som kallas för en global signifikansnivå  $\alpha'$  vilken definieras genom  $P(\text{någon sann } H_{0i} \text{ förkastas} \leq \alpha')$  som är den önskade signifikansnivån[5]. Enligt ekvation (4) fås ett  $\alpha$  som kan användas för att  $\alpha'$  ska uppnås[6].

$$\begin{aligned}\alpha' &= 1 - (1 - \alpha)^n \\ \Rightarrow \alpha &= 1 - (1 - \alpha')^{\frac{1}{n}}\end{aligned}\quad (4)$$

Om  $H_{0i}$ ,  $i = 1, \dots, n$  däremot är beroende av varandra används istället Booles olikhet (se ekvation (5)) för att få fram  $\alpha'$ [7]. Ekvation (6) visar hur beräkningarna går till.

$$P(A \cup B) \leq P(A) + P(B) \quad (5)$$

$$\begin{aligned}&P(\text{någon sann } H_{0i} \text{ förkastas}) \\ &= P\left(\bigcup_{i=1}^n H_{0i} \text{ förkastas}\right) \\ &\leq \sum_{i=1}^n P(H_{0i} \text{ förkastas}) \\ &= n\alpha\end{aligned}\quad (6)$$

Väljs  $\alpha = \frac{\alpha'}{n}$  är det därmed säkert att olikheten ovan gäller. Det brukar kallas för Bonferroni-korrektion[5].

## 2.5 Binomialfördelning

Binomialfördelning är en diskret sannolikhetsfördelning associerad med något experiment i vilket  $n$  oberoende försök som endast har två utfall utförs och där de två utfallen är ömsesidigt uteslutande. Varje försök har en sannolikhet,  $p$ , att lyckas. Låt antalet lyckade försök betecknas  $K$ , där  $K \in \text{Bin}(n, p)$ . Sannolikheten att lyckas med  $k$  försök av totalt  $n$  stycken kan då skrivas som:

$$P(K = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (7)$$

## 2.6 Relativ risk

Relativ risk (RR) är kvoten mellan sannolikheten att en person som utsätts för en miljöfaktor blir sjuk och sannolikheten att en ej utsatt person blir det.

$$r = \frac{(\text{sjuk}|\text{miljöfaktor finns})}{(\text{sjuk}|\text{miljöfaktor finns ej})} \quad (8)$$

RR ger ett mått på hur mycket en miljöfaktor förändrar risken att drabbas av en sjukdom[8].

### 3 Konstiga händelser

#### –Företeelser som verkar osannolika

Anställda vid polishuset i Hagfors kommun i Värmland slog år 2011 larm då det framkommit att 17 personer som arbetat i huset insjuknat i cancer sedan år 1992[9]. Händelsen sågs som mycket ovanlig och det fanns misstankar om en skadlig miljö som bakomliggande faktor och det beslutades därför att diverse undersökningar på huset skulle genomföras. Undersökningarna som gjordes visade att det inte fanns något fel på huset som kan förklara de många cancerfallen[10]. Går det då att, utan att göra en undersökning på huset, påstå att det kunde vara en slumpmässig händelse och att det därför inte behöver vidtas några åtgärder samt att det inte finns skäl till vidare oro bland de anställda?

Med statistiska metoder är det möjligt att avgöra hur konstig händelsen med cancerfallen på polishuset i Hagfors är. En metod är att beräkna p-värdet (se avsnitt 2.3), det vill säga undersöka hur sannolikt det är att 17 eller fler personer drabbas av cancer givet att nollhypotesen,  $H_0 = \text{Händelsen beror på slumpen}$ , gäller.

Det kan antas att antalet sjuka under en given tidsperiod är binomialfördelade, vilket kan göras då det rör sig om en grupp av ändlig storlek[11]. Notera att binomialfördelning endast kan antas om antalet sjukdomsfall kan anses oberoende av varandra, det fungerar därför inte om en smittsam sjukdom betraktas. Således kan antalet cancerfall,  $K$ , under en given tidsperiod antas vara binomialfördelade, det vill säga  $K \in \text{Bin}(n, p)$  där  $p$  betecknar risken att drabbas av cancer under den givna tidsperioden och  $n$  det totala antalet personer i gruppen. Sannolikheten att  $k$  eller fler personer insjuknar i cancer under den givna tidsperioden blir då:

$$P(K \geq k) = \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i} \quad (9)$$

I det här fallet är antalet insjuknade,  $k = 17$ , och antalet personer som har jobbat på polisstationen sedan det öppnade år 1980 är ungefär 100 personer, vilket ger  $n = 100$ [12]. Sverige har cirka 9 490 683 invånare och det inträffar ungefär 55 000 cancerfall per år[13][14]. Sannolikheten att en person drabbas av cancer under ett år är således 0,0058. Notera att detta är en grov förenkling av risken att drabbas av cancer under ett år då de flesta som drabbas är äldre människor och här antas det att det är lika risk att drabbas oavsett ålder. För att finna värdet på  $p$  i binomialfördelningen multipliceras värdet för att drabbas av cancer under ett år med det antal år som motsvarar det önskade tidsintervallet.

I fallet med Hagfors finns det två olika tidsintervall som kan vara intressanta. Det första intervallet är på 20 år och motsvarar tiden från första cancerfallet till idag som ger värdet  $p = 0,1160$ . Det andra tidsintervallet är från det år då huset började användas vilket ger ett intervall på 32 år och en sannolikhet på 0,1856. Beräknas  $P(K \geq k)$  med de ovan givna siffrorna fås p-värden på 0,0684 respektive 0,6949 (se bilaga B för beräkningar).

Utgående från att nollhypotesen,  $H_0 = \text{Händelsen beror på slumpen}$ , gäller och en signifikansnivå på  $\alpha = 0,05$  ges att  $p > \alpha$  för båda tidsintervallen. Det medför att  $H_0$  ej kan förkastas med 95% säkerhet (se avsnitt 2.3 och tabell 2 nedanför). Oavsett ifall det ses till tidsperioden från första cancerfallet till idag, 20 år, eller från den tidpunkt då byggnaden togs i bruk blir resultatet alltså att det inte går att förkasta att händelsen var en slump även fast det är stor skillnad på resultaten i de båda fallen. De här resultaten stämmer överens med de tekniska undersökningar som gjordes på huset där inga fel hittades som kunde förklara cancerfallen.

Hur konstigt är det då om händelsen i Hagfors ställs i relation till att det i någon grupp om 100 personer i Sverige insjuknar 17 personer i cancer under en given tidsperiod? Det går även att ställa det i relation till att det på någon arbetsplats i Sverige eller Värmland, med samma antal anställda, inträffar 17 fall av cancer under en given tidsperiod.

För att testa det första påståendet, hur osannolikt det är att det i någon grupp om 100 personer i Sverige insjuknar 17 personer i cancer, antas det att de som jobbade/jobbar på polisstationen är minst 30 år gamla samt att cancer utvecklades innan de fyllt 76 år. Det

finns idag 5 323 900 personer mellan 30 och 75 år i Sverige[15]. Det ger att det finns 53 239 grupper om 100 personer i det åldersintervallet. Eftersom varje person endast ingår i en grupp är antalet sjuka i en grupp oberoende av antalet sjuka i alla andra grupper. För varje grupp är sannolikheten för ett givet antal sjuka lika stor och därför kan antalet grupper med 17 eller fler cancerfall antas vara binomialfördelade. Med en sannolikhet,  $p$ , att det i en grupp om 100 personer inträffar 17 eller fler cancerfall kan sannolikheten att det i minst  $g$  grupper i Sverige insjuknar 17 eller fler personer i cancer beräknas som:

$$P(G \geq g) = 1 - P(G < g) = 1 - \sum_{i=0}^{g-1} \binom{n}{i} p^i (1-p)^{n-i} \quad (10)$$

Genom att sätta antalet grupper  $n = 53\,239$  och det gamla p-värdet (se beräkning av ekvation (9) ovan) till

$$p = \begin{cases} 0,0684 & \text{för tidsintervallet 20 år} \\ 0,6949 & \text{för tidsintervallet 32 år,} \end{cases}$$

om  $g = 1$  fås ett p-värde på cirka 1. Då det nu finns många grupper finns det även många hypoteser,  $H_{01}, \dots, H_{0n} = \text{Händelsen beror på slumpen}$ . För att testa multipla oberoende hypoteser mot en signifikansnivå  $\alpha$  används en global signifikansnivå  $\alpha'$  som beskrivs i avsnitt 2.4. Med en global signifikansnivå på  $\alpha' = 0,05$  och enligt ekvation (4) fås  $\alpha = 9,63 \cdot 10^{-7}$  oberoende av vilket p-värde som används. Det här medför att hypoteserna  $H_{01}, \dots, H_{0n}$  inte kan förkastas ty  $p > \alpha$ .

Vid test av hur konstig händelsen är ställt i relation till antal arbetsplatser i Sverige och Värmland görs ett antal antaganden vid beräkningarna. Först antas det att alla arbetsplatser med samma antal anställda som polishuset i Hagfors (mellan 10 och 19 anställda) har samma intensitet på omsättningen av personal. Det antas alltså att alla arbetsplatser som har mellan 10 och 19 anställda totalt har haft 100 personer anställda under den givna tidsperioden (sedan 1980)[12][9]. Vidare antas det att fördelningen på antal arbetsplatser med 10 – 19 anställda är samma i Värmland som i övriga Sverige. Med de här och med tidigare givna antaganden går det att beräkna sannolikheten för 17 eller fler cancerfall på någon arbetsplats, av samma storlek, i Sverige och Värmland.

I Sverige finns det 37 535 arbetsplatser med 10 – 19 anställda och i Värmland 1 026[16][17]. Följaktligen är  $n = 37\,535$  respektive  $n = 1\,026$  i ekvation (10). Med samma sannolikheter som ovan, fås p-värden på cirka 1 för båda tidsintervallen för Sverige och Värmland. Då samma sak gäller för nollhypoteserna i de här två fallen som tidigare och med en global signifikansnivå som ovan fås enligt ekvation (4) att  $\alpha = 1,36654 \cdot 10^{-6}$  respektive  $\alpha = 4,99922 \cdot 10^{-5}$  för båda p-värdena. Även här är alltså  $p > \alpha$  vilket återigen ger att nollhypotesen att händelsen beror på slumpen inte kan förkastas.

Resultaten från beräkningarna ovan kan ses i tabell 2 (se bilaga B för beräkningar). Från de resultaten går det att utläsa att då de anställda på polishuset ses som en enskild grupp är sannolikheten för det inträffade liten och deras oro kan utifrån deras perspektiv förstås. Om händelsen undersöks noggrannare, och ställs i relation till fler grupper, fås att händelsen inte är så ovanlig som det kan tyckas vid en första anblick. Det går inte att utesluta att händelsen berodde på slumpen. Det är dock inte sagt att det inte berodde på någon annan faktor ty vid statistiska test där en nollhypotes  $H_0$  testas går det endast att avgöra om den kan förkastas eller ej. Kan inte nollhypotesen förkastas betyder det endast att det inte går att utesluta att nollhypotesen kan vara giltig.

Det går även att vända på problemet och ställa frågan om hur många som skulle behöva bli sjuka för att det statistiska testet skulle ge att slumpen kan förkastas. Med multipel inferens kommer det krävas ännu fler sjukdomsfall för att nollhypotesen ska kunna förkastas. För att få en översikt finns det i tabell 3 siffror på hur många sjukdomsfall som krävs för varje tidsintervall och gruppering för att händelsen ska anses som konstig, det vill säga att hypotesen förkastas.

Tabell 2: Tabell över sannolikhet för minst 17 fall av cancer i minst en grupp om 100 personer med nollhypotesen/nollhypoteserna att händelsen beror på slumpen.

Betraktade grupper	Tidsintervall	p-värde	Signifikansnivå $\alpha$	Förkasta hypotes
Polishuset	20 år	0,0684	0,05	Nej
Polishuset	32 år	0,6949	0,05	Nej
Sveriges befolkning (30-75 år) indelat i gr. om 100	20 samt 32 år	0,9999	$9,63 \cdot 10^{-7}$	Nej
Arbetsplatser i Sverige med 10-19 anställda	20 samt 32 år	0,9999	$1,36654 \cdot 10^{-6}$	Nej
Arbetsplatser i Värmland med 10-19 anställda	20 samt 32 år	0,9999	$4,99922 \cdot 10^{-5}$	Nej

Tabell 3: Minsta antalet sjuka som behövs för att nollhypotesen ska förkastas.

Betraktade grupper	Tidsintervall	Signifikansnivå $\alpha$	Antal sjuka som krävs för att förkasta hypotes
Polishuset	20 år	0,05	18
Polishuset	32 år	0,05	26
Sveriges befolkning (30-75 år) indelat i gr. om 100	20 år	$9,63 \cdot 10^{-7}$	30
Sveriges befolkning (30-75 år) indelat i gr. om 100	32 år	$9,63 \cdot 10^{-7}$	40
Arbetsplatser i Sverige med 10-19 anställda	20 år	$1,36654 \cdot 10^{-6}$	30
Arbetsplatser i Sverige med 10-19 anställda	32 år	$1,36654 \cdot 10^{-6}$	40
Arbetsplatser i Värmland med 10-19 anställda	20 år	$4,99922 \cdot 10^{-5}$	27
Arbetsplatser i Värmland med 10-19 anställda	32 år	$4,99922 \cdot 10^{-5}$	36

Exemplet ovan syftade till att visa på hur det går att göra en första enkel statistisk beräkning för att testa en hypotes. Beroende på resultaten av den här typen av test är det rimligt att antingen gå vidare i undersökningen (det vill säga anta att det inte är en slumpmässig händelse) eller att besluta att inte göra något (det vill säga tro på att det var slumpen). Nästa avsnitt bygger vidare på hur det är möjligt att gå vidare och göra fler beräkningar.



## 4 Okända miljöfaktorer

När det uppstår situationer med ett till synes ovanligt/osannolikt stort antal sjukdomsfall inom en specifik gruppering (som i exemplet i avsnitt 3) är det troligt att det misstänks vara något skadligt i den gemensamma miljön. Är det då okänt vad det kan vara för möjlig riskhöjande miljöfaktor kan det vara problematiskt att avgöra om sjukdomsfallen kan vara orsakade av den. Det går att, som i avsnitt 3, avgöra om det går att förkasta antagandet om att det är en slumpmässig händelse, men det kan vara rimligt att gå vidare i analysen och beräkna sannolikheten att det finns en miljöfaktor som höjer risken för sjukdom.

### 4.1 Modell

För att beräkna sannolikheten att det finns en miljöfaktor som höjer risken för sjukdom givet ett antal sjuka kan Bayes sats (se avsnitt 2.1) användas. Med  $A = \text{miljöfaktor som höjer risken}$  och  $B = \text{data (antalet sjuka i en grupp)}$  fås sannolikheten för en riskhöjande miljöfaktor, enligt Bayes sats, till:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) \quad (11)$$

För att kunna beräkna ekvation (11) behövs sannolikheten för ett visst antal sjuka i en grupp,  $P(B)$ , sannolikheten för samma antal sjuka i en grupp givet en skadlig miljöfaktor,  $P(B|A)$ , samt a priori sannolikheten att det finns en miljöfaktor som höjer risken,  $P(A)$ .

Som i avsnitt 3 antas antalet sjuka,  $K$ , i en grupp vara binomialfördelade med parametrarna  $n$  och  $p$ , där  $n$  betecknar det totala antalet personer i gruppen och  $p$  betecknar risken att drabbas av en sjukdom, det vill säga  $K \in \text{Bin}(n, p)$ . Alltså kan sannolikheten för att  $k$  slumpvis valda personer blir sjuka under en given tidsperiod,  $P(K = k)$ , skrivas som ekvation (12).

$$P(K = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (12)$$

Om  $K_A$  betecknar antalet sjuka under samma tidsperiod givet att personerna är utsatta för en skadlig miljöfaktor  $A$  och  $r$  den relativa risken (se avsnitt 2.6) associerad med  $A$  så gäller att  $K_A \in \text{Bin}(n, r \cdot p_{\bar{A}})$  där  $p_{\bar{A}}$  är sannolikheten att en person som inte är utsatt för miljöfaktorn  $A$  blir sjuk. Enligt lagen om total sannolikhet gäller att risken,  $p$ , att en person blir sjuk kan skrivas som:

$$p = p_{\bar{A}} \cdot P(\bar{A}) + r \cdot p_{\bar{A}} \cdot P(A). \quad (13)$$

Vilket i sin tur ger att

$$p_{\bar{A}} = \frac{p}{P(\bar{A}) + r \cdot P(A)}. \quad (14)$$

Sätts  $P(\bar{A}) + r \cdot P(A) = \lambda$  fås att  $K_A \in \text{Bin}(n, r \cdot p/\lambda)$  och  $K_{\bar{A}} \in \text{Bin}(n, p/\lambda)$ . Sannolikheten att  $k$  personer som är, respektive inte är, utsatta för miljöfaktorn  $A$  blir sjuka under en given tidsperiod kan då beräknas som:

$$P(K = k|A) = P(K_A = k) = \binom{n}{k} \left(\frac{r \cdot p}{\lambda}\right)^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k} \quad (15)$$

$$P(K = k|\bar{A}) = P(K_{\bar{A}} = k) = \binom{n}{k} \left(\frac{p}{\lambda}\right)^k \left(1 - \frac{p}{\lambda}\right)^{n-k} \quad (16)$$

Nu gäller att antingen är alla  $n$  personerna vi betraktar utsatta för miljöfaktorn  $A$  eller så är de det inte. Med hjälp av lagen om total sannolikhet fås då att sannolikheten att  $k$  personer i den här gruppen blir sjuka kan beräknas som:

$$P(K = k) = P(K_A = k) \cdot P(A) + P(K_{\bar{A}} = k) \cdot P(\bar{A}) \quad (17)$$

Om den data,  $B$ , som är tillgänglig ger att  $k$  personer av totalt  $n$  stycken blivit sjuka under en given tidsperiod går det att, med hjälp av Bayes sats (11) och ekvation (15) och (16), beräkna sannolikheten att det finns en miljöfaktor som höjer risken för sjukdom givet ett antal sjuka som:

$$P(A|B) = P(A|K = k) = \frac{\left(\frac{r \cdot p}{\lambda}\right)^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k}}{\left(\frac{r \cdot p}{\lambda}\right)^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k} \cdot P(A) + \left(\frac{p}{\lambda}\right)^k \left(1 - \frac{p}{\lambda}\right)^{n-k} \cdot P(\bar{A})} P(A) \quad (18)$$

Uttrycket i ekvation (18) beror endast av sannolikheten för en skadlig miljöfaktor, den relativa risken och sannolikheten att drabbas av en sjukdom, det vill säga  $P(A)$ ,  $r$  och  $p$ . Sannolikheten att drabbas av en sjukdom,  $p$ , går oftast att bestämma då det är känt vilken sjukdom/sjukdomar det handlar om. Det är däremot svårt att ta fram användbara siffror på sannolikheten för en riskhöjande miljöfaktor och den relativa risken om det inte är känt vilken typ av miljöfaktor det skulle kunna vara. Exemplet med polishuset i Hagfors kan återigen användas för att visa på svårigheten i att beräkna ekvation (18).

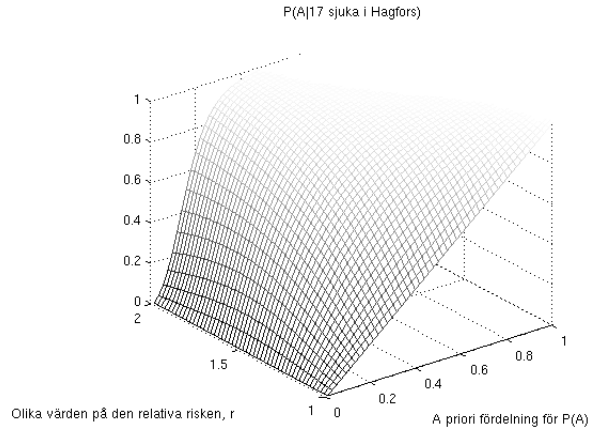
## 4.2 Test av modell

Som det beskrevs i avsnitt 3 misstänktes det att det kunde vara en skadlig miljöfaktor som låg bakom de många cancerfallen. Ekvation (18) kan användas för att beräkna sannolikheten att det finns en skadlig miljö givet ett antal cancerfall. Problemet ligger då i att skatta  $P(A)$  och  $r$ . Att skatta en a priori sannolikhet (se avsnitt 2.2) för  $P(A)$  är svårt då det, som sagt, är okänt vilken typ av miljöfaktor det handlar om. Det är möjligt att göra vissa antaganden och utifrån erfarenhet skatta en a priori sannolikhet. Samma sak gäller för den relativa risken,  $r$ , då den beror på miljöfaktorn  $A$ .

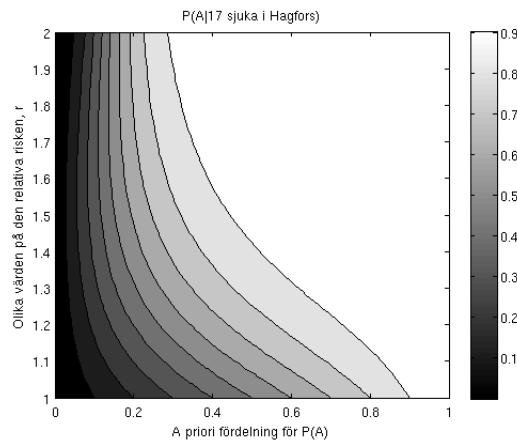
I fallet med polishuset i Hagfors är det inte nödvändigtvis möjligt att, trots stor tidigare erfarenhet, skatta en rimlig a priori för  $P(A)$  då det handlade om många olika typer av cancer som kan ha orsakats av många skilda faktorer. Det är såklart möjligt att anta en a priori på 0,5 då det endast finns två utfall för  $P(A)$  i det här fallet (antingen existerar det en miljöfaktor eller ej) och om det är ett rimligt antagande eller inte kan sedan diskuteras.

Vilken a priori som väljs kan ha stor betydelse för resultatet, vilket också bidrar till osäkerheten i att skatta en sådan. För att illustrera detta har `Matlab` använts (se bilaga B för `Matlab`-program) för att beräkna och plotta hur  $P(A|B)$  (sannolikheten för att det finns en miljöfaktor som höjer risken givet data) påverkas då  $r$  och  $P(A)$  varieras. Det är återigen exemplet med cancerfallen på polishuset i Hagfors som använts och ekvation (18) användes för att beräkna  $P(A|B)$ .

I figur 1 ses hur  $P(A|17$  sjuka) påverkas då  $P(A) \in [0, 1]$  och  $r \in [1, 2]$ . Övriga parametrars värden i beräkningen var  $n = 100$ ,  $k = 17$  och  $p = 0,1160$ . Med  $p = 0,1160$  antogs alltså tidsperioden vara 20 år och var det värde på  $p$  som i avsnitt 3 gav ett p-värde på 0,0684 (se tabell 2).



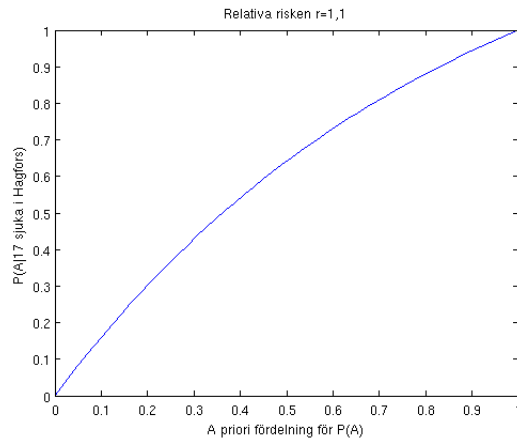
(a) 3D plot för  $P(A|17$  sjuka)



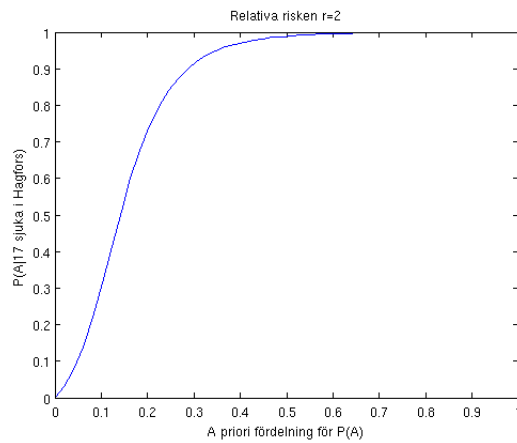
(b) Nivåkurvor för  $P(A|17$  sjuka)

Figur 1: Sannolikheten att det existerar en miljöfaktor  $A$  givet 17 sjuka där  $P(A) \in [0, 1]$  och den relativa risken  $r \in [1, 2]$  och där tidsperioden antas vara 20 år vilket ger en sannolikhet på  $p = 0,1160$  att drabbas av cancer.

För att ytterligare illustrera hur  $P(A|B)$  förändras för olika a priori fördelningar testas hur en priori för  $P(A)$  påverkar resultatet då den relativa risken hålls fix vid ett visst värde och även tvärtom då  $P(A)$  hålls fix och den relativa risken varieras. I figur 2 ses hur  $P(A|17$  sjuka i Hagfors) påverkas då  $P(A) \in [0, 1]$  och den relativa risken hålls fix vid  $r = 1, 1$  respektive  $r = 2$ .



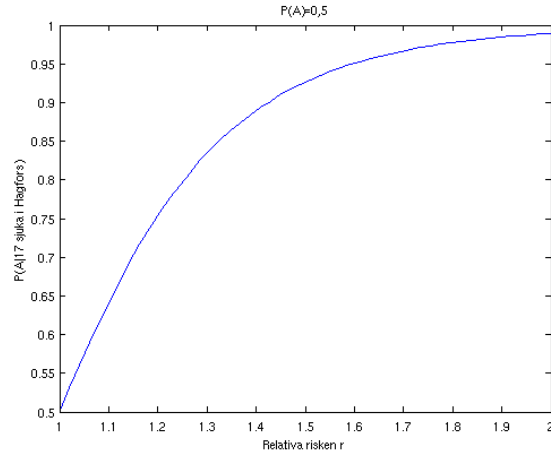
(a) Plot för  $P(A|17 \text{ sjuka})$  då  $r = 1,1$



(b) Plot för  $P(A|17 \text{ sjuka})$  då  $r = 2$

Figur 2: Sannolikheten att det existerar en miljöfaktor  $A$  givet 17 sjuka där  $P(A) \in [0, 1]$  och den relativa risken  $r = 1,1$  respektive  $r = 2$  och där tidsperioden antas vara 20 år vilket ger en sannolikhet på  $p = 0,1160$  att drabbas av cancer.

I figur 3 visas hur  $P(A|B)$  varierar då  $P(A)$  hålls fix vid 0,5 och den relativa risken  $r \in [1, 2]$ .



Figur 3: Sannolikheten att det existerar en miljöfaktor  $A$  givet 17 sjuka där  $P(A) = 0,5$  och den relativa risken  $r \in [1, 2]$  och där tidsperioden antas vara 20 år vilket ger en sannolikhet på  $p = 0,1160$  att drabbas av cancer.

Slutligen kan det sägas att ju mindre data och information som är tillgänglig desto fler a priori skattningar behöver göras vilket påverkar säkerheten i resultatet på det som ska beräknas. Detta medför att det i sådana situationer kan vara farligt att göra beräkningar och sedan helt förlita sig på resultatet. Det finns dock tillfällen där fler värden är kända vilket ökar säkerheten i de uträknade värdena. Nästa avsnitt syftar till att bygga en modell för sådana fall.

## 5 Kända miljöfaktorer

### *-När en specifik miljöfaktor misstänks*

I avsnitt 3 och 4 undersöktes dels huruvida till synes osannolika företeelser kan sägas bero på slumpen eller ej och dels möjligheter och svårigheter i att beräkna sannolikheten att det existerar en riskhöjande miljöfaktor givet ett antal sjukdomsfall då miljöfaktorn i fråga är okänd. Det finns även situationer då det är känt vilken typ av miljöfaktor/miljöfaktorer det skulle kunna handla om. I sådana fall finns det ofta kunskap till att göra en bättre a priori skattning av sannolikheten för en skadlig miljöfaktor än i avsnitt 4.2. Det är även möjligt att undersöka fler saker än i fallet då det är en okänd miljöfaktor, som till exempel sannolikheten att det är en miljöfaktor som ligger bakom sjukdomsfallen, sannolikheten för nya sjukdomsfall orsakade av miljön, den förväntade kostnaden av att åtgärda respektive inte åtgärda problemet.

### 5.1 Modell

För att bygga en modell där ovan nämnda risker och sannolikheter kan undersökas används följande definitioner:

$A$  = miljöfaktor som höjer risken

$B$  = data (antalet sjuka i en grupp)

$C$  = nytt sjukdomsfall

$I_C$  = antalet nya sjukdomsfall

$D$  = nytt sjukdomsfall orsakat av miljö ( $A$ )

$I_D$  = antalet nya sjukdomsfall orsakade av miljö ( $A$ )

$N$  = antalet friska personer i gruppen

$H$  = kostnad för handling/åtgärd ( $\bar{H}$  är därmed kostnaden för att inte göra något)

Grunden i modellen utgörs, som i avsnittet om okända miljöfaktorer, av Bayes sats (se avsnitt 2.1) vilken ger att sannolikheten för att det finns en miljöfaktor som höjer risken,  $A$ , givet ett antal sjuka i en grupp,  $B$ , ges av:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) \quad (19)$$

Förutom att beräkna sannolikheten att det existerar en hälsofarlig miljöfaktor givet ett antal sjuka personer skall modellen även ha fler användningsområden. Den ska kunna användas till att beräkna risken för ett nytt sjukdomsfall orsakat av miljön givet en skadlig miljöfaktor,  $P(D|A)$ , samt sannolikheten att det är en miljöfaktor som ligger bakom ett givet sjukdomsfall då miljöfaktorn existerar, det vill säga  $P(D|C \cap A)$ . De här två sannolikheterna kan härledas enligt följande beräkningar:

Med hjälp av lagen för total sannolikhet kan risken för ett nytt sjukdomsfall givet en skadlig miljöfaktor skrivas som:

$$P(C|A) = P(C \cap D|A) + P(C \cap \bar{D}|A) \quad (20)$$

För händelsen  $C \cap \bar{D}$ , det vill säga ett nytt sjukdomsfall som inte är orsakat av  $A$ , gäller för de sjukdomar som betraktas (se avsnitt 1.5), att  $P(C \cap \bar{D})$  är oberoende av  $A$ . Följaktligen gäller att:

$$P(C \cap \bar{D}|A) = P(C \cap \bar{D}) = P(C \cap \bar{D}|\bar{A}) \quad (21)$$

Det gäller att  $P(\bar{D}|\bar{A}) = 1$  eftersom det är omöjligt att ha ett sjukdomsfall orsakat av  $A$  om  $A$  inte finns. Därför är  $P(C \cap \bar{D}|\bar{A}) = P(C|\bar{A})$ . Notera att  $D$  är en delmängd av  $C$ , det vill säga  $D \subset C$ , och därför är  $C \cap D = D$ . Om det här och ekvation (21) sätts in i ekvation (20) fås att:

$$P(C|A) = P(D|A) + P(C|\bar{A}). \quad (22)$$

Utifrån detta fås ett uttryck för risken för nytt sjukdomsfall orsakat av en miljöfaktor,  $A$ , givet existensen av en riskhöjande sådan som:

$$P(D|A) = P(C|A) - P(C|\bar{A}) \quad (23)$$

För  $P(D|C \cap A)$  görs, med hjälp av regler för betingad sannolikhet (se ekvation (3) i avsnitt 2.1), omskrivningen:

$$P(D|C \cap A) = \frac{P(D \cap C \cap A)}{P(C \cap A)} = \frac{P(D \cap C|A) \cdot P(A)}{P(C|A) \cdot P(A)} \quad (24)$$

Eftersom  $D \subset C$  gäller att  $P(D \cap C|A) = P(D|A)$ . Med hjälp av ekvation (23) kan ekvation (24) skrivas som:

$$P(D|C \cap A) = \frac{P(D|A)}{P(C|A)} = 1 - \frac{P(C|\bar{A})}{P(C|A)} \quad (25)$$

Vidare ger lagen om total sannolikhet att:

$$P(D|B) = P(D|A \cap B) \cdot P(A|B) + P(D|\bar{A} \cap B) \cdot P(\bar{A}|B) \quad (26)$$

Eftersom  $P(D|\bar{A}) = 0$  (vilket ger att  $P(D|\bar{A} \cap B) = 0$ ) blir den andra termen i ekvation (26) noll. Dessutom gäller att om det är känt huruvida det finns en skadlig miljöfaktor,  $A$ , eller ej så påverkas inte sannolikheten för att en person blir sjuk på grund av  $A$  av hur många som tidigare blivit sjuka. Följaktligen gäller att  $P(D|A \cap B) = P(D|A)$ . Ekvation (26) kan därför skrivas som:

$$P(D|B) = P(D|A) \cdot P(A|B) \quad (27)$$

Ekvation (25) ger sannolikheten för att ett givet sjukdomsfall, i en situation där  $A$  existerar, är orsakat av  $A$  och ekvation (27) ger sannolikheten för ett nytt sjukdomsfall orsakat av en miljöfaktor,  $A$ , givet data,  $B$ .

Ekvation (23) och (27) gäller även då det handlar om  $i$  stycken nya sjukdomsfall orsakade av  $A$ , vilket kan visas analogt. Det vill säga följande samband gäller:

$$P(I_D = i|A) = P(I_C = i|A) - P(I_C = i|\bar{A}) \quad (28)$$

$$P(I_D = i|B) = P(I_D = i|A) \cdot P(A|B) \quad (29)$$

För att utvidga modellen till att kunna beräkna kostnader får nedanstående beteckningar symbolisera de olika kostnader som kan uppstå. Det är en situation där det kanske kan finnas en miljöfaktor som har orsakat ett antal sjukdomsfall och beslut om eventuella åtgärder ska tas.

$K_u$  = kostnad för att genomföra en undersökning

$K_a$  = kostnad för att åtgärda miljöproblemet så att det inte längre orsakar någon skada

$K_s$  = den kostnad som tillskrivs ett sjukdomsfall.

Med ovanstående definitioner går det nu att beräkna den förväntade kostnaden av olika beslut som:

$$E[H] = K_u + K_a \cdot P(A|B) \quad (30)$$

$$E[\bar{H}] = K_s \sum_{i=1}^N iP(I_D = i|B) = K_s \cdot E[(I_D|B)] \quad (31)$$

$E[H]$  är den förväntade kostnaden att genomföra en undersökning och, om det finns en skadlig miljöfaktor, åtgärda problemet.  $E[\bar{H}]$  är den förväntade kostnaden som uppstår då ingen åtgärd görs.

Väntevärdena i ekvation (30) och (31) tar inte hänsyn till kostnader för de personer som kan förväntas bli sjuka oavsett om miljöfaktorn  $A$  är närvarande eller ej eftersom den blir samma oberoende av vilket beslut som tas.

## 5.2 Antagande om binomialfördelning

För att beräkna sannolikheterna och väntevärdena i föregående stycke antas det återigen, precis som i avsnitt 4, gälla att antalet sjuka i en grupp är binomialfördelade och då gäller, enligt ekvation (18) att

$$P(A|B) = \frac{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k}}{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k} \cdot P(A) + \left(1 - \frac{p}{\lambda}\right)^{n-k} \cdot P(\bar{A})} P(A) \quad (32)$$

där  $n$  är det totala antalet personer i gruppen,  $k$  antalet sjukdomsfall i gruppen,  $p$  risken att drabbas av sjukdomen,  $r$  den relativa risken och  $\lambda = P(\bar{A}) + r \cdot P(A)$ . Låt sedan  $N$  beteckna antalet friska personer som finns i gruppen (notera att det ej behöver gälla att  $N = n - k$  eftersom det kan ha tillkommit eller försvunnit personer av andra anledningar). Sätts antagandet om binomialfördelning in i ekvation (28) fås ett uttryck för att beräkna sannolikheten för  $i$  stycken sjukdomsfall orsakade av  $A$  givet förekomst av  $A$ , det vill säga

$$P(I_D = i|A) = \binom{N}{i} \left( \left( \frac{r \cdot p}{\lambda} \right)^i \left( 1 - \frac{r \cdot p}{\lambda} \right)^{N-i} - \left( \frac{p}{\lambda} \right)^i \left( 1 - \frac{p}{\lambda} \right)^{N-i} \right) \quad (33)$$

Genom att sätta in ekvation (32) i ekvation (30) och genom att kombinera ekvation (30), (29), (32) och (33) fås uttryck för att beräkna  $E[H]$  och  $E[\bar{H}]$ . som:

$$E[H] = K_u + K_a \cdot \frac{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k}}{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k} \cdot P(A) + \left(1 - \frac{p}{\lambda}\right)^{n-k} \cdot P(\bar{A})} P(A) \quad (34)$$

$$E[\bar{H}] = K_s \cdot N p \frac{(r-1)}{\lambda} \cdot \frac{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k}}{r^k \left(1 - \frac{r \cdot p}{\lambda}\right)^{n-k} \cdot P(A) + \left(1 - \frac{p}{\lambda}\right)^{n-k} \cdot P(\bar{A})} P(A) \quad (35)$$

Talet  $(r-1)$  som förekommer här kallas även ERR (*Excess relative risk*) [8].

Notera att  $P(C|\bar{A}) = p_{\bar{A}}$  och  $P(C|A) = r \cdot p_{\bar{A}}$ . Ekvation (25) kan då beräknas som

$$P(D|C \cap A) = 1 - \frac{p_{\bar{A}}}{r \cdot p_{\bar{A}}} = 1 - \frac{1}{r} \quad (36)$$

Det har nu tagits fram uttryck för att beräkna sannolikheter och förväntade värden som kan användas i situationer där det är känt vilken riskhöjande miljöfaktor det handlar om. I följande avsnitt utvidgas modellen och görs mer komplex.



### 5.3 Olika sannolikhet för olika grupper

I föregående stycken antogs alla personer ha samma sannolikhet att bli sjuk. I detta avsnitt ska modellen utvidgas till att hantera olika sjukdomsrisker och påverkan av miljöfaktorer för olika personer.

Antag att olika sjukdomsfall är oberoende av varandra (det vill säga att, som angivet i avsnitt 1.5 en icke smittsam sjukdom betraktas). Antag vidare att det går att dela in gruppen i  $M$  stycken olika undergrupper där  $P(C_j)$  är sannolikheten att någon person i undergrupp  $j$  blir sjuk och  $P(D_j)$  är sannolikheten att någon person i grupp  $j$  blir sjuk på grund av miljöfaktorn  $A$ . Notera att om varje person skulle ha ett eget värde (vilket är fallet i verkligheten) skulle det innebära lika många grupper som personer. Eftersom sannolikheten att en person i grupp  $j$  blir sjuk inte påverkas om en person i grupp  $k$  blir sjuk ( $j \neq k$ ) gäller att:

$$P(I_{C_j} = i \cap I_{C_k} = l) = P(I_{C_j} = i)P(I_{C_k} = l)$$

och

$$P(I_{D_j} = i \cap I_{D_k} = l) = P(I_{D_j} = i)P(I_{D_k} = l)$$

Om  $B_j$  är data över hur många som blivit sjuka i grupp  $j$  (och  $B$  betecknar all data) gäller att:

$$P(A|B) = P(A | \bigcap_{j=1}^M B_j) = \left( \prod_{j=1}^M \frac{P(B_j|A)}{P(B_j)} \right) P(A) \quad (37)$$

Låt  $N_j$  vara antalet friska personer i grupp  $j$ . Eftersom sjukdomsfallen i de olika grupperna är oberoende av varandra fås, med hjälp av ekvation (28) och (29) att:

$$E[I_{D_j}|B] = \sum_{i=1}^{N_j} i (P(I_{C_j} = i|A) - P(I_{C_j} = i|\bar{A})) \cdot P(A|B) \quad (38)$$

$$E[(I_D|B)] = \sum_{j=1}^M E[(I_{D_j}|B)] \quad (39)$$

där  $I_{D_j}$  betecknar antalet sjukdomsfall orsakade av miljöfaktorn  $A$  i grupp  $j$ . Om detta sätts in i ekvation (30) och (31) fås:

$$E[H] = K_u + K_a \cdot P(A|B) \quad (40)$$

$$E[\bar{H}] = K_s \sum_{j=1}^M E[(I_{D_j}|B)] = K_s \cdot E[(I_D|B)] \quad (41)$$

### 5.4 Olika sannolikheter för olika grupper med antagande om binomialfördelning

Om, precis som i föregående avsnitt 5.3, gruppen ska delas in i  $M$  olika undergrupper går det fortfarande att använda binomialfördelning. Antag att  $p_j$  är risken att en person i undergrupp  $j$  blir sjuk under tidsperioden  $T$ ,  $n_j$  är antalet personer i undergrupp  $j$  och  $r_j$  den relativa risken associerad med miljöfaktor  $A$  för undergrupp  $j$ . Låt  $K_j$  beteckna antalet sjukdomsfall i undergrupp  $j$  under  $T$  och  $K_{A,j}$  antalet sjukdomsfall i undergrupp  $j$  under  $T$  givet miljöfaktorn  $A$ . Sätt dessutom  $\lambda_j = P(\bar{A}) + r_j \cdot P(A)$ . Antagandet om binomialfördelning ger då att  $K_j \in \text{Bin}(n_j, p_j)$ ,  $K_{\bar{A},j} \in \text{Bin}(n_j, p_j/\lambda_j)$  och  $K_{A,j} \in \text{Bin}(n_j, r_j \cdot p_j/\lambda_j)$ . (Notera att  $K = \sum_{j=1}^M K_j$  och att  $n = \sum_{j=1}^M n_j$ .) Eftersom det redan i avsnitt 5.3 antagits att antalet sjukdomsfall i de olika undergrupperna är oberoende av varandra går det att skriva:

$$P\left(\bigcap_{j=1}^M (K_{A,j} = k_j)\right) = \prod_{j=1}^M \binom{n_j}{k_j} \left(\frac{r_j \cdot p_j}{\lambda_j}\right)^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j} \quad (42)$$

$$P\left(\bigcap_{j=1}^M (K_{\bar{A},j} = k_j)\right) = \prod_{j=1}^M \binom{n_j}{k_j} \left(\frac{p_j}{\lambda_j}\right)^{k_j} \left(1 - \frac{p_j}{\lambda_j}\right)^{n_j - k_j} \quad (43)$$

Om datan,  $B$ , är just att  $K_j = k_j$ ,  $j = 1, \dots, M$  ger Bayes sats (19) att:

$$P(A|B) = \frac{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j}}{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(A) + \prod_{j=1}^M \left(1 - \frac{p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(\bar{A})} \cdot P(A) \quad (44)$$

Låt nu  $N_j$  beteckna antalet friska personer i grupp  $j$ . För varje grupp gäller att sannolikheten för att  $i$  personer blir sjuka på grund av miljöfaktorn  $A$  kan beräknas på samma sätt som i ekvation (33). Sätts ekvation (39) in i (40) och ekvation (41) kombineras med (44) fås:

$$E[H] = K_u + K_a P(A|B) = K_u + K_a \cdot \frac{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j}}{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(A) + \prod_{j=1}^M \left(1 - \frac{p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(\bar{A})} \cdot P(A) \quad (45)$$

$$E[\bar{H}] = K_s \sum_{j=1}^M N_j p_j \frac{(r_j - 1)}{\lambda_j} \cdot \left( \frac{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j}}{\prod_{j=1}^M r_j^{k_j} \left(1 - \frac{r_j \cdot p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(A) + \prod_{j=1}^M \left(1 - \frac{p_j}{\lambda_j}\right)^{n_j - k_j} \cdot P(\bar{A})} \cdot P(A) \right) \quad (46)$$

## 5.5 Flera miljöfaktorer

Hur ändras sambanden i 5.1 om det kan finnas flera olika skadliga miljöfaktorer? Låt faktorerna betecknas  $A_1, A_2, \dots, A_m$  ( $A_0$  står för normalfallet med ingen riskhöjande miljöfaktor). Antag att dessa miljöfaktorer är ömsesidigt uteslutande (det vill säga  $A_0, A_1, \dots, A_m$  utgör en partition). Då kan Bayes sats (19) skrivas som:

$$P(A_l|B) = \frac{P(B|A_l)}{P(B)} P(A_l) = \frac{P(B|A_l)}{\sum_{j=0}^m (P(B|A_j) \cdot P(A_j))} \cdot P(A_l) \quad (47)$$

Med den nya beteckningen  $D_l$ =nytt sjukdomsfall orsakat av miljö  $A_l$  blir ekvationerna (23), (25), (27), (28) och (29) analoga (med tillägg av index på  $A$  och  $D$ ). Låt nu  $K_u$  vara kostnaden för att undersöka om någon av miljöfaktorerna  $A_1, A_2, \dots, A_m$  är närvarande och  $K_{a,l}$  kostnaden för att åtgärda miljöproblem  $l$ . Då kan den förväntade kostnaden för att genomföra en undersökning och eventuellt åtgärda problem respektive den förväntade kostnaden för att inte vidta åtgärd beräknas som:

$$E[H] = K_u + \sum_{l=1}^m K_{a,l} \cdot P(A_l|B) \quad (48)$$

$$E[\bar{H}] = K_s \sum_{l=1}^m \sum_{j=1}^N j P(I_{D_l} = i|B) = K_s \sum_{l=1}^m E[(I_{D_l}|B)] \quad (49)$$

Självklart kan även dessa ekvationer för flera miljöfaktorer anpassas efter antagande om binomialfördelning.

## 6 Exempel med radon som skadlig miljöfaktor

Tanken med detta avsnitt är att exemplifiera hur de formler som presenterades i avsnitt 5 kan användas. Den miljöfaktor som betraktas är radon (som det går att läsa mer om i bilaga A). Givetvis kommer flera förenklingar att göras för att problemet ska bli hanterbart.

### 6.1 Introduktion till problemet

Antag att i en grupp på  $n$  personer som har bott i samma hus har  $k$  personer drabbats av lungcancer innan de nått 75 års ålder. Vad är då sannolikheten att radonhalten i huset överstiger 200 Bq/m<sup>3</sup>? Vad är den mest sannolika radonhalten? De  $k$  personer av de  $n$  stycken som bott i huset och som blivit sjuka är den "konstiga händelser" som ska betraktas och som betecknas  $B$  i formler i detta avsnitt. Om det i framtiden kommer att bo  $N$  stycken personer i huset, är det då rimligt att vidta åtgärder för att de inte ska bli sjuka av den eventuellt förhöjda radonhalten?

I hela detta avsnitt antas att antalet sjuka vid en given radonhalt är bionomialfördelade. Kostnaden för att undersöka om ett hus har en förhöjd radonhalt antas vara 500 kronor och kostnaden för att sänka den om så är fallet sätts till 50 000 kronor. (Den verkliga kostnaden för att mäta ligger på några hundra kronor och kostnaden för att sänka radonhalten kan ligga på allt från 100 till 100 000 kronor, se bilaga A.4.) I en rapport gjord av Boverket 2010 uppskattas den samhällsekonomiska kostnaden för att en person får lungcancer (och troligtvis dör till följd av detta) till 22 miljoner kronor [18] och samma värde kommer användas i kostnadsuppskattningar i detta avsnitt.

Eftersom det uppskattas att cirka 10% av svenskarna bor i ett hus där gränsvärdet överstiger riktvärdet går det, om ingen annan information finns tillgänglig, att sätta  $P(A) = 0,1$  där  $A$  är en förhöjd radonhalt. Det genomsnittliga radonvärdet i svenska hem ligger på 83 Bq/m<sup>3</sup>. (se bilaga A.3).

### 6.2 Den enklaste modellen

I den allra enklaste modellen antas alla löpa lika stor risk att drabbas av lungcancer. Sannolikheten att drabbas av lungcancer före 75 års ålder är 2,5% för män och 2,2% för kvinnor[19]. Under antagandet att det finns ungefär lika många män som kvinnor i Sverige blir då sannolikheten att en slumpmässig person drabbas av lungcancer före 75 års ålder 2,35%. För att göra det så enkelt som möjligt antas dessutom att antingen ligger radonhalten på medelvärdet 83 Bq/m<sup>3</sup> eller så är den 200 Bq/m<sup>3</sup>. Den skadliga miljöfaktorn,  $A$ , som kan förekomma är alltså en radonhalt på 200 Bq/m<sup>3</sup>. Den relativa risken är då  $(1 + 0,16 \cdot 200/100)/(1 + 0,16 \cdot 83/100) = 1,16525$  (se bilaga A.2) och det tal som betecknas  $\lambda$  i avsnitt 5.2 kan då beräknas till  $\lambda = 0,9 + 1,16525 \cdot 0,1 = 1,016525$ . Med hjälp av ekvation (32) kan då sannolikheten för en förhöjd radonhalt beräknas som:

$$P(A|B) = \frac{0,1 \cdot 0,0269382^k \cdot 0,973062^{-k+n}}{0,1 \cdot 0,0269382^k \cdot 0,973062^{-k+n} + 0,9 \cdot 0,023118^k \cdot 0,976882^{-k+n}} \cdot 0,1 \quad (50)$$

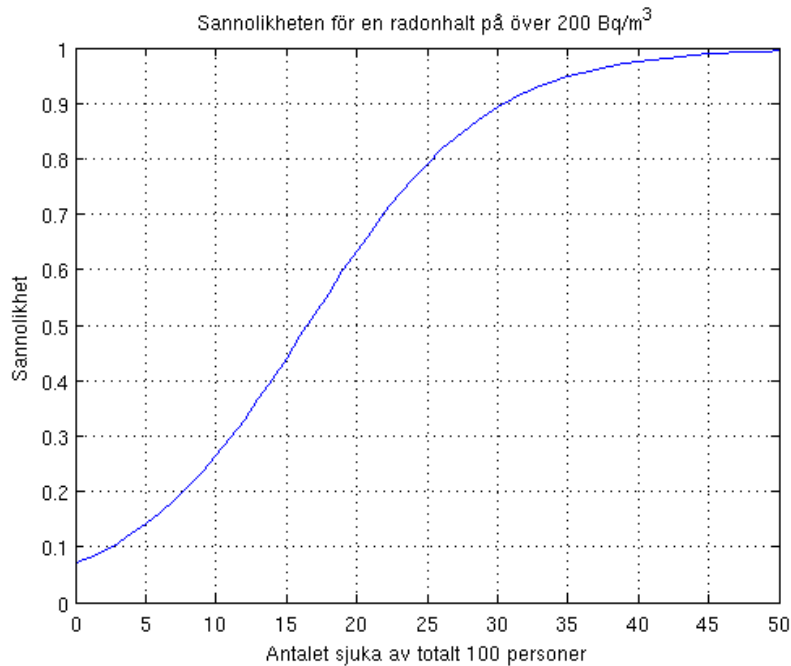
Antag att 100 personer bott i huset, det vill säga  $n = 100$ . I figur 4 visas sannolikheten för en förhöjd radonhalt beroende på hur många, av totalt 100 personer, som drabbats av lungcancer.

Med de här siffrorna kan väntevärdet av kostnaden för olika ageranden beräknas. Med hjälp av ekvation (34) och (35) fås att:

$$E[H] = K_u + K_a \cdot P(A|B) \quad (51)$$

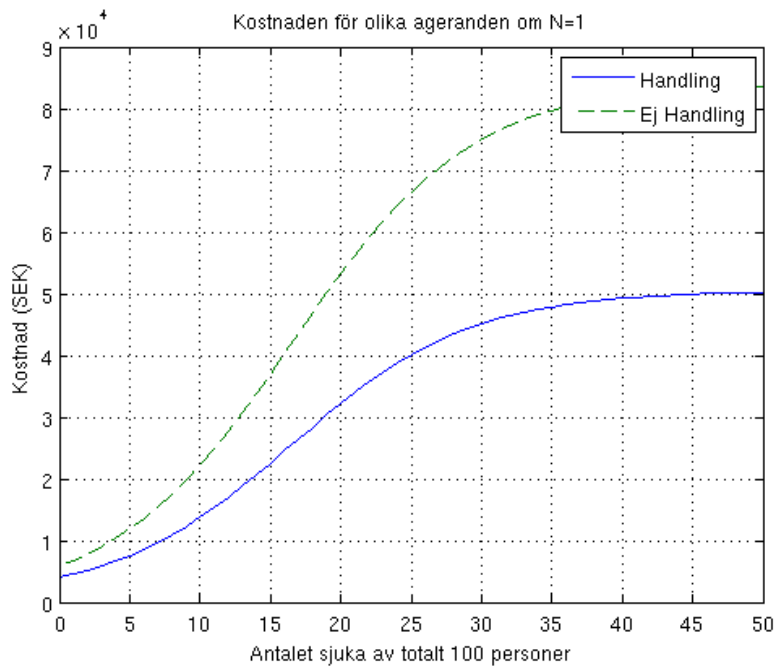
$$E[\bar{H}] = K_s \cdot (E[I_A] - E[I_{\bar{A}}])P(A|B) = K_s \cdot 0,0235 \cdot N \cdot \frac{1,16525 - 1}{1,016525} P(A|B) \quad (52)$$

där  $I_A$  och  $I_{\bar{A}}$  är det förväntade antalet nya fall av lungcancer om radonhalten är 200 Bq/m<sup>3</sup> respektive 83 Bq/m<sup>3</sup>. Antag att endast en frisk person kommer att bo i huset från och med



Figur 4: Sannolikheten för en förhöjd radonhalt i den enklaste modellen.

nu tills det rivs, det vill säga antag att  $N = 1$ . Med tidigare angivna värden på  $K_u$ ,  $K_a$  och  $K_s$  kan väntevärdet av kostnaderna för olika agerande beräknas och resultatet visas i figur 5. Som syns i diagrammet är det alltid lönt att vidta åtgärder trots att det bara är en enda person som kommer att bo i huset.



Figur 5: Kostnaden för olika ageranden om en person kommer att bo i huset.

### 6.3 Lite mer komplicerad modell

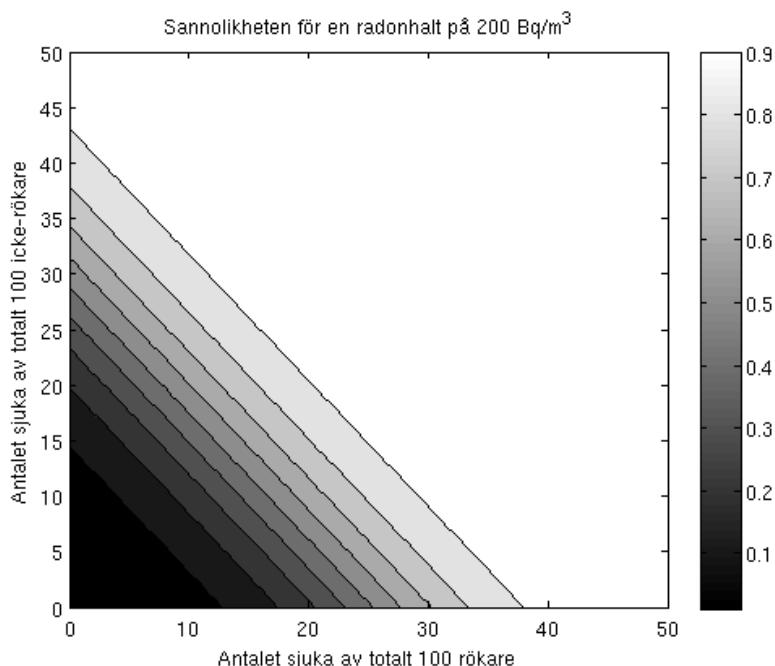
Alla personer löper inte samma risk att drabbas av lungcancer. En rökare löper mycket större risk att drabbas av lungcancer. För att få ett mer korrekt värde på sannolikheten att huset har en förhöjd radonhalt bör därför rökare och icke-rökare studeras var för sig på det sätt som anges i avsnitt 5.4. Antag att i gruppen är  $n_1$  personer rökare och  $n_2$  personer icke-rökare. Av dessa har  $k_1$  respektive  $k_2$  personer drabbats av cancer. Sannolikheten att en icke radonutsatt person drabbas av lungcancer är 0,41% respektive 10,1% för icke-rökare respektive rökare (se bilaga A.2).

Precis som i avsnitt 6.2 antas, för att göra det så enkelt som möjligt, att antingen ligger radonhalten på medelvärdet 83 Bq/m<sup>3</sup> eller så är den 200 Bq/m<sup>3</sup>. (Även i detta avsnitt är alltså den skadliga miljöfaktorn,  $A$ , en radonhalt på 200 Bq/m<sup>3</sup>.) Den relativa risken jämfört med radonhalten 0 är då, för såväl rökare som icke-rökare, 1,32 om personen är utsatt för en radonhalt på 200 Bq/m<sup>3</sup> och 1,1328 om personen är utsatt för en radonhalt på 83 Bq/m<sup>3</sup> (se bilaga A.2)

Sannolikheten för en förhöjd radonhalt kan nu beräknas med hjälp av ekvation (44) i avsnitt 5.4 till (Notera att eftersom både  $p_A$  och  $p_{\bar{A}}$  är kända för såväl rökare som icke-rökare behöver inget  $\lambda$  beräknas):

$$P(A|B) = \frac{((1,32 \cdot 0,101)^{k_1} (1 - 1,32 \cdot 0,101)^{n_1 - k_1} \cdot (1,32 \cdot 0,0041)^{k_2} (1 - 1,32 \cdot 0,0041)^{n_2 - k_2} \cdot 0,1)}{((1,32 \cdot 0,101)^{k_1} (1 - 1,32 \cdot 0,101)^{n_1 - k_1} \cdot (1,32 \cdot 0,0041)^{k_2} (1 - 1,32 \cdot 0,0041)^{n_2 - k_2} \cdot 0,1 + (1,1328 \cdot 0,101)^{k_1} (1 - 1,1328 \cdot 0,101)^{n_1 - k_1} \cdot (1,1328 \cdot 0,0041)^{k_2} (1 - 1,1328 \cdot 0,0041)^{n_2 - k_2} \cdot 0,9)} \quad (53)$$

Låt nu  $n_1 = n_2 = 100$ . I figur 6 visas en konturplot för hur sannolikheten för en förhöjd radonhalt på 200 Bq/m<sup>3</sup> beror av  $k_1$  och  $k_2$ .



Figur 6: Sannolikheten för en förhöjd radonhalt i en modell som tar hänsyn till att rökare och icke-rökare löper olika stor risk att drabbas.

### 6.4 Flera möjliga radonhalter

I avsnitt 6.2 och 6.3 antogs att radonhalten var antingen 83 Bq/m<sup>3</sup> eller 200 Bq/m<sup>3</sup>. Givetvis kan radonhalten i ett hus anta fler värden än dessa två. Olika radonhalter kan ses som olika

miljöfaktorer som är ömsesidigt uteslutande och formlerna i avsnitt 5.5 kan då användas. Antag nu att det är ett flerbostadshus som studeras. (Tillvägagångssättet för ett småhus är analogt.) Fördelningen av flerbostadshus med olika radonhalt kan ses i tabell 4.

Tabell 4: Andelen svenska hem med olika radonhalt[20].

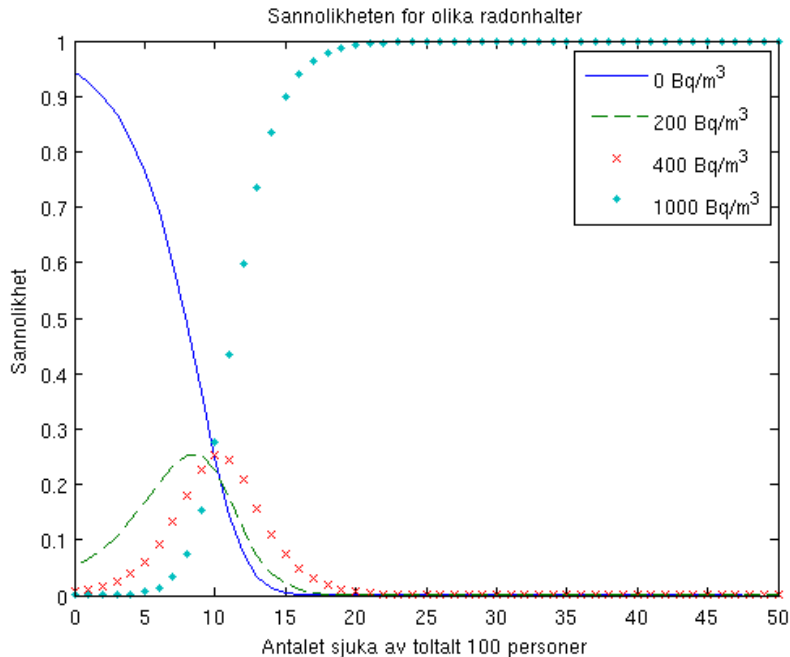
Radonhalt	Lägenheter	Småhus
0 – 200 Bq/m <sup>3</sup>	87,7 %	71,7 %
200 – 400 Bq/m <sup>3</sup>	9,7 %	19,0 %
400 – 1000 Bq/m <sup>3</sup>	2,4 %	8,1 %
> 1000 Bq/m <sup>3</sup>	0,2 %	1,2 %

Antag, för enkelhetens skull, att radonhalten i alla hus i en grupp ligger på lägsta möjliga radonvärde (det vill säga att alla hem i gruppen 200 – 400 Bq/m<sup>3</sup> antas ha en radonhalt på 200 Bq/m<sup>3</sup>). Beteckna de olika radonhalterna med  $A_1, \dots, A_4$ .

Till att börja med antas, precis som i avsnitt 6.2, att alla löper lika stor risk att drabbas av lungcancer och risken att en slumpmässig person drabbas är  $p = 2,35\%$ . Eftersom den relativa risken ökar med 0,16 per 100 Bq/m<sup>3</sup> (se bilaga A.2) gäller det att:

$$0,0235 = p = \sum_{l=1}^4 p_l P(A_l) = p_1 \cdot (1 \cdot 0,877 + 1,32 \cdot 0,097 + 1,64 \cdot 0,024 + 2,6 \cdot 0,002) \quad (54)$$

Där  $p_l$  är risken att drabbas av lungcancer vid radonhalten  $A_l$ . Med hjälp av ovanstående och ekvation (47) kan sannolikheten för olika radonhalter beräknas. Om, precis som i avsnitt 6.2,  $n = 100$  fås resultatet som visas i figur 7.



Figur 7: Sannolikheten för olika radonhalter givet ett visst antal sjuka.

Låt nu, som i avsnitt 6.3, de som bott i huset vara indelade i rökare och icke-rökare. Sannolikheten att en person drabbas av lungcancer ses i tabell 5 (beräknas med hjälp av siffror från bilaga A.2)

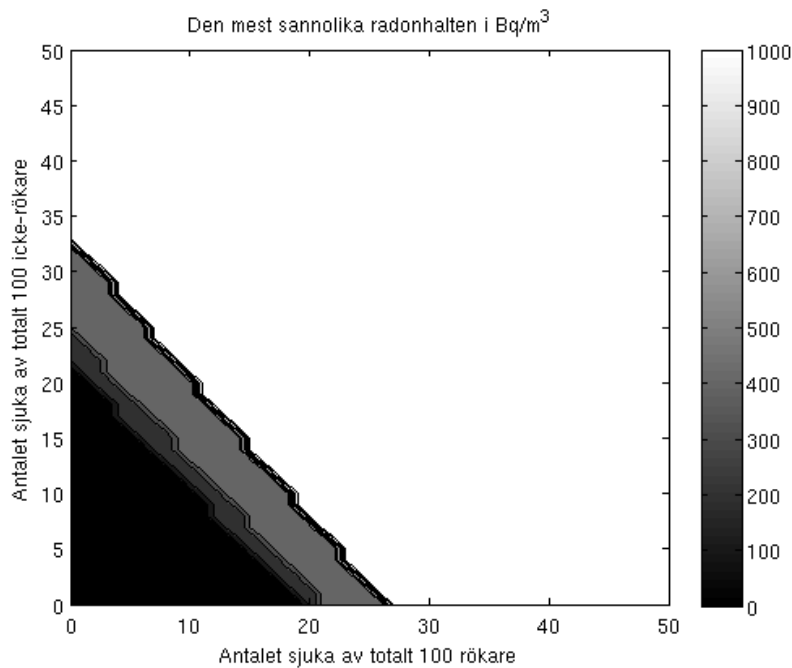
Tabell 5: Risken att en person drabbas av lungcancer vid olika radonhalter.

Radonhalt (Bq/m <sup>3</sup> )	Risken att drabbas av lungcancer (%)	
	Ej rökare	Rökare
0	0,41	10,1
200	0,54	13,33
400	0,67	16,56
1000	1,07	26,26

Sannolikheten för en given fördelning av antalet sjuka vid en given radonhalt blir:

$$P(B|A_l) = \prod_{j=1}^2 \left( \binom{n_j}{k_j} p_{j,A_l}^{k_j} (1 - p_{j,A_l})^{n_j - k_j} \right) \quad (55)$$

Låt, precis som i avsnitt 6.3 antalet rökare och antalet icke-rökare vara 100. Sannolikheten för en given radonhalt kan då beräknas med hjälp av ekvation (47). I figur 8 visas en konturplot för den mest sannolika radonhalten för olika värden på antalet sjuka.



Figur 8: Den mest sannolika radonhalten beroende på antalet sjuka rökare och icke-rökare.

För varje kombination av antalet sjuka rökare och icke-rökare kan såklart kostnaden för olika handlingssätt beräknas med hjälp av ekvation (48) och (49).

## 7 Tillvägagångssätt i praktiken

I ett samhälle är det alltså viktigt att effektivt kunna finna och åtgärda olika miljöfaktorer. Hur görs detta i praktiken?

Det finns en ganska bred och lång styrkedja inom miljöarbetet som stort i Sverige. Dels finns internationella direktiv och lagar från EU. Riksdag och regering är, som bekant, de högsta svenska beslutande organen och, förutom lagstiftning, beslutar de vilka miljömål landet ska ha (just nu tillämpas ett system där det finns 16 olika miljömål som skall dirigera hur miljöarbetet kommer att se ut i Sverige de närmaste åren) samt utser vilka myndigheter som skall ansvara över vad. Naturvårdsverket har en viktig roll i det nationella arbetet och ansvarar för 7 av de 16 målen, då det dock främst rör sig om ett flertal miljöområden som ej direkt är kopplat till mänsklig hälsa undersöks inte de närmare. En rad andra myndigheter ansvarar över de resterande miljömålen, däribland Socialstyrelsen, som bland annat producerat nationella miljöhälsorapporter grundat på stora enkäter med frågor kring miljöfaktorer, hälsotillstånd och olika besvär[22][23]. På en mer regional nivå har Länsstyrelsen ofta en stor roll, exempelvis har många utgått från Socialstyrelsens enkät men använt sig av regional data och skapat material som är mer relevant i området[24].

Hittills har väldigt övergripande organisationer med breda arbetsfält tagits upp och deras uppgifter är snarare mer av det samordnande slaget än faktiskt utförande. Så nu finns en känsla för var underlagen till beslut kommer ifrån, och det är intressant att se hur dessa beslut fattas på en lokal nivå.

### 7.1 Hur gör de på Chalmers?

Vid en intervju med Jennica Kjällstrand vid miljöenheten på Chalmers<sup>1</sup>, framgick en del hur det kan gå till i en mindre skala. Det arbete som Chalmers själva ansvarar för är att följa de förordningar och lagar som finns, och givetvis bedriva ett så gott säkerhetsarbete som möjligt[25]. Det finns till exempel ett register över alla de som arbetat med cancerframkallande ämnen på Chalmers, som kan tillämpas om orsaken behöver spåras hos någon som får cancer flera år senare.

De som jobbar med något hälso- eller miljöfarligt måste även rapportera in detta till miljöförvaltningen. Detta är särskilt vanligt för diverse projekt vid kemiinstitutionen. Det är i första hand de respektive institutionerna som själva ansvarar för att reda ut miljö- och hälso-problem, med hjälp av Chalmers miljöenhet vid behov. Om miljöenheten får in en rapport om sjukdomar gör de själva en undersökning, och i för komplicerade fall tar de kontakt med konsulter.

I allmänhet undersöks inte givna sjukdomsfall, utan fokus ligger på verksamhetsområden där miljön anses vara riskfylld (som vid användning av kemiska ämnen) där både förebyggande arbete och uppföljning av symptom förekommer, förklarar Kjällstrand.

### 7.2 Kommunal nivå

På Göteborgs Stads hemsida står det att finna:

I Göteborg är det miljöförvaltningen som ansvarar för tillsynen över det som klassas som miljöfarliga verksamheter. Förvaltningens inspektörer gör besök på plats och går igenom handlingar för att se om verksamheten följer miljöbalkens regler. De handlägger också anmälningsärenden och skriver yttranden till byggnadsnämnden och Länsstyrelsen. I arbetet ingår också att informera och ge råd till företag i miljöfrågor[26].

Förutom bland annat miljöbalken och strålskyddslagen styrs tillsynen i Göteborg av Miljö- och klimatnämnden (Miljönämnden innan 2011)[27]. Den del av miljöförvaltningens arbete

<sup>1</sup>Jennica Kjällstrand (enhetschef, miljöenheten, Chalmers) intervjuad av författarna den 28 mars 2012.



som är mest relevant för denna rapport är miljötillsyn av företag, offentliga lokaler (såsom skolor, idrottsanläggningar med mera) och bostadsmiljö (lägenheter och dess närmiljö), och specifikt de delar där inriktningen är hälsa och inte så mycket själva miljöpåverkan. Förvaltningen gör även mycket annat som indirekt är relevant, däribland kartläggning av kvävedioxidhalten i luften och annan miljöövervakning som gynnar presumtiva utredningar av sjukdomars orsak[27].

Andreas Severinsson<sup>2</sup> vid miljöförvaltningen i Göteborg, kontaktades och mer detaljer diskuterades. Även hos dessa ligger fokus till större del på riskfyllda verksamheter samt på de anmälningar de får in där misstanke om miljöorsak föreligger. Det går inte till så att en andel sjukdomsfall inom en grupp observeras och därefter undersöks, menar Severinsson. Likt tankesättet på Chalmers krävs det inte ett högt incidenstal för att de ska agera. Exempelvis om en hyresgäst klagar på fuktskador eller symptom på fuktskador där denne bor ingår det i miljöförvaltningens rutin att först och främst kontakta fastighetsägaren och begära fastighetsinformation och fråga om problemet är känt eller ej. Det handlar om huruvida fastighetsägaren redan är medveten om detta och ämnar vidta åtgärder (är det fuktskador måste det givetvis åtgärdas). När problemet är okänt sänds en handläggare till platsen för en första undersökning. När det gäller fuktrelaterade skador mäts luftfuktigheten i fastigheten samt tittar, luktar och känner handläggaren för att göra en kvalificerad bedömning. Oftast reder de själva ut om det förekommer en fuktskada, eller annan relevant miljöorsak. Endast ungefär en gång per år behöver de ta hjälp av expertkompetensen vid VMC, Västra Götalandsregionens Miljömedicinska Centrum, säger Severinsson. Som exemplifieras i en rapport från VMC som kan tittas på vid intresse[28].

Vad gäller tillsyn på företag kan miljöförvaltningen helt enkelt kräva dem att åtgärda sådant som inte uppfyller gällande standarder. Alla tillsynspliktiga företag i landet är skyldiga att lämna in en årlig miljörapport, som helt enkelt syftar till att beskriva hur de arbetar för att motverka diverse miljöproblem (och därmed hålla sig till lagarna). Det rör sig bland annat om utsläpp och hantering av kemikalier, som Kjällstrand tidigare berättade om rörande Chalmers arbete.

Förutom miljöförvaltningen kontaktades ett antal andra personer.

Anna Stjernholm<sup>3</sup> som handhar olika miljö- och hälsofrågor i Partille förklarar att kommunen inte haft anledning att göra en undersökning kring ett högt antal sjukdomsfall och om det beror eller inte beror på någon miljöfaktor. Om de skulle misstänka att ett antal sjukdomsfall berodde på en gemensam faktor, skulle de kontakta VMC.

Gudrun Törnström<sup>4</sup> arbetar på miljöskyddsenheten på Länsstyrelsen i Västra Götaland. Återigen nämns VMC som några specialister de kan kontakta vid behov, och att mycket av arbetet inte sker vid just Länsstyrelsen utan de fungerar mer som en koordinerande enhet.

### 7.3 VMC

Utav all extern kontakt som hafts har alltid VMC kommit på tal, och somliga har hänvisat till dem<sup>5</sup>. VMC är egentligen den miljömedicinska delen av arbets- och miljömedicin (AMM) vid Sahlgrenska Universitetssjukhuset. Som redan framgått är det ett centrum för miljömedicinsk kompetens, och deras uppgifter är bland annat att utreda åt och rådgiva kommuner och länsstyrelse i Västra Götaland. De söker kartlägga hälsorisker och deras orsaker[29].

När det gäller anhopningar av cancersjuka, något som är nära kopplat till tidigare avsnitt, har de i allra högsta grad en rutin hur de skall gå till väga, som man kan finna i ett faktablad på deras hemsida[30]. Redan noterat är att det är många som kan kontakta VMC vid behov, och detta är ett sådant tillfälle. De blir ofta ombudda att göra undersökningar på arbetsplatser eller kommuner, och det första de gör innefattar insamling av data av olika slag, så noggrant som möjligt, exempelvis antal sjuka, ålders- och könsfördelning på personalen, personallistor,

<sup>2</sup>Andreas Severinsson (Göteborgs miljöförvaltning) uppringd den 30 mars 2012.

<sup>3</sup>Anna Stjernholm (Miljöchef Partille kommun) uppringd den 23 mars 2012.

<sup>4</sup>Gudrun Törnström (Miljöskyddsenheten, Länsstyrelsen Västra Götaland) uppringd den 27 mars.

<sup>5</sup>Inger Högstedt (Arbetsmiljöinspektör, Arbetsmiljöverket Göteborg) mailkontakt den 14 mars.

någon misstänkt exponeringsfaktor, etc. Exponeringsfaktorer är något de sätter stor vikt vid att finna. Arbetsgivaren är enligt lag skyldig att göra en riskbedömning och om det inte redan är infört måste det utredas vilka cancerframkallande ämnen olika grupper av anställda utsätts för. VMC är intresserade av att få veta till exempel vilka personalgrupper som exponeras vid vilka moment, om det gjorts exponeringsmätningar och vad deras resultat var, om skyddsutrustning används och om den uppfyller sin funktion och liknande. Det är främst av intresse att undersöka de faktorer som är av störst bidragande orsak till den givna cancerformen. De försöker även beräkna förväntat antal fall, och vid tillfällen då det inträffade antalet är mindre eller lika med detta värde kan lugnande besked oftast ges. De påpekar också svårigheten att avfärda slumpen även vid ett stort antal sjuka, och därför gör de sambandsutredningar och försöker hitta möjliga miljöorsaker. Här undersöks om det finns kända belägg för att givna miljöfaktorer eller ämnen är cancerframkallande, och motsvarande vilka faktorer som är mest bidragande till de givna cancerfallen. De letar efter statistik som svarar huruvida det är högre förekomst av cancerfall i den givna, eller liknande, yrkesgrupper. Sedan diskuteras om den givna exponeringen kan förklara antalet fall och hur farlig fortsatt exponering är och vilka åtgärder som skall tas, om det skall tas några över huvud taget.

Slutligen står det att i de flesta fall av canceranhopningar de undersöker kan ingen gemensam faktor hittas utan slumpen är den enda förklarande orsaken. Eftersom canceranhopningar är skrämmande för de inblandade och deras omgivning är ett viktigt led i arbetet att informera och försöka lugna oro.

## 8 Diskussion och Slutsatser

I det här avsnittet diskuteras de olika resultaten som har framkommit i tidigare avsnitt.

### 8.1 Konstiga händelser

I avsnitt 3 kom vi fram till att det antagligen inte var en sådan underlig händelse med cancerfallen i Hagfors polishus då det inte i ett enda hypotestest gick att, med 95% säkerhet, förkasta nollhypotesen att det var en slumpmässig händelse. Det betyder dock inte att det är helt säkert att cancerfallen var slumpmässiga ansamlingar utan bara att de skulle kunna vara det. Notera även att de siffror som användes i beräkningarna är approximativa och kan således ha påverkat resultatet.

Något som är värt att fundera över är vilket tidsintervall som är det mest relevanta att studera. Det finns två olika tidsintervall med i beräkningarna, det ena är från det år då det nya polishuset började användas och det andra är från det år då det första cancerfallet upptäcktes. Tidsintervallet på 32 år är för stort på grund av att det tar tid att utveckla cancer som följd av en miljöpåverkan. Således är de första åren sedan byggnaden började användas inte representativa då inga cancerfall kan ha uppstått under den tiden så att en riskfylld miljö orsakat dem. Det kan ge som resultat att cancerfallen i Hagfors verkar vanligare än vad de egentligen är. Tidsintervallet från och med det första cancerfallet ger å andra sidan ett något för snävt tidsintervall, men det är ändå antagligen ett bättre tidsintervall på grund av att det troligen ligger närmare det år då cancerfallen skulle ha kunnat utvecklas.

Vid frågan på hur många som hade behövt bli sjuka för att nollhypotesen skulle kunna förkastas går det att utläsa ur tabell 3 att det endast hade krävts ett sjukdomsfall till för att det skulle ha gått att förkasta nollhypotesen vid tidsintervallet 20 år utan multipel inferens. I de andra fallen hade det behövts betydligt fler sjukdomsfall, de närmaste exemplen är under 32 år utan multipel inferens samt under tidsintervallet 20 år i förhållande till lika stora arbetsplatser i Värmland, med 9 respektive 10 fler sjukdomsfall för att nollhypotesen skulle kunna förkastas.

Det kan finnas en fara i att använda sig av multipel inferens när många hypoteser,  $H_{01}, \dots, H_{0n}$ , testas då det vid ett stort  $n$  fås ett mycket litet värde på  $\alpha$  eftersom sannolikheten för att en händelse inträffar ökar med antalet försök. Det kan i sin tur medföra att inga hypoteser förkastas. Hur rimligt det är att alltid avfärda händelser liknande den i Hagfors som slump, enbart för att det med statistiska test med tillräckligt många grupper kommer få det resultatet kan då diskuteras. Det här kan i sin tur leda till att undersökningar och åtgärder som egentligen borde vidtas inte genomförs. På grund av detta är det viktigt att försöka avgöra hur gruppindelningen ska ske vid statistiska test då multipel inferens ska användas. Det går att hitta fler gruppindelningar att testa på, till exempel alla som heter Greger, men de gruppindelningar som har använts är de som vi har ansett mest relevanta. Det gemensamma med de gruppindelningar som valdes är att det intuitivt går att finna en rimlig miljöpåverkan som skulle kunna vara orsaken.

Som det påpekades ovan kan det alltså finnas en fara i att använda multipel inferens vid statistiska test. Det kan därför vara motiverat att istället försöka skatta sannolikheten för en given miljöfaktor och sedan räkna på sannolikheten att det existerar en sådan givet data (antal sjuka).

### 8.2 Miljöfaktorer

Om det inte finns någon misstanke om någon specifik miljöfaktor är det, som sågs i avsnitt 4, svårt att göra säkra beräkningar eftersom det är svårt att både ha en a priori skattning för sannolikheten för miljöfaktorn och att veta hur mycket den påverkar sjukdomsrisken när det inte är känt vilken miljöfaktor det rör sig om. Det kan alltså vara nästan omöjligt att göra en rimlig skattning av risken för en skadlig miljöfaktor givet data. Det är såklart möjligt att, om ingen annan information finns, sätta en a priori sannolikhet på  $1/n$ , där  $n$  är antalet utfall. I figur 3 ses vad detta ger för resultat om det tillämpas på exemplet med polishuset i Hagfors.

(Här är  $n = 2$ , det vill säga, antingen finns miljöfaktorn eller så gör den det inte.) Det som går att utläsa från figuren är att sannolikheten för existens av en skadlig miljöfaktor givet ett antal sjuka ökar med den relativa risken, vilket är ett rimligt resultat. Samma sak går att se i figur 1 där a posteriori fördelningen,  $P(A|B)$ , ökar då a priori fördelningen,  $P(A)$ , och den relativa risken,  $r$ , ökar.

Det vi också kan se i figurerna från avsnitt 4 är att a posteriori fördelningen är känslig för variationer hos dels a priori fördelningen och dels den relativa risken. Slutsatsen som kan dras är att då man har en tillsynes konstig händelse, men ingen mer information att tillgå är det möjligtvis rimligare att bara hypotestesta händelsen, det vill säga testa om det går att förkasta att händelsen beror på slumpen eller ej. Görs detta är det dock viktigt att ta hänsyn till hur man delar in i grupper och att man tittar på rimliga grupperingar för att resultaten ska kunna användas. Annars kan det bli, som det diskuterades ovan, att alltför få nollhypoteser förkastas.

Även om det inte finns någon misstanke om någon miljöfaktor eller om en undersökning redan har utförts utan att någon riskfaktor kunde påvisas är det viktigt att hypotestest genomförs. Ifall händelsen är allt för "konstig" bör någon typ av åtgärder, som till exempel evakuering, genomföras. Det skulle kunna finnas en miljöfaktor vars skadeverknin g är okänd och som därför inte betraktats. Om, till exempel, antalet cancerfall i polishuset i Hagfors (se avsnitt 3) hade varit 70 istället för 17 hade det varit rimligt att utrymma huset även om ingen skadlig miljöfaktor gick att finna vid undersökning.

För de flesta sjukdomar finns det, till skillnad från tidigare i avsnittet, ofta några kända riskfaktorer som skulle kunna misstänkas. För varje enskild faktor är det ofta känt ungefär hur mycket den höjer sjukdomsriskens vilket medför att det är enklare att göra en bra a priori skattning. Ett alternativ skulle därför kunna vara att använda ekvationerna i avsnitt 5.5 för hur flera olika miljöfaktorer ska behandlas. Det här tillvägagångssättet skulle dock inte fungera om det är totalt okänt vilken miljöfaktor det rör sig om, men samtidigt är det svårt att avgöra vad som ska mätas då.

Även då det är en känd miljöfaktor som betraktas finns dock problem med olika skattningar och förenklingar. I avsnitt 6 sattes a priori skattningen av  $P(A)$  till att vara lika stor som andelen hus med en viss radonhalt. Det är så klart bättre än att inte ha någon aning, men är det bra? Det är inte så att alla hus löper samma risk av att vara drabbade, utan faktorer som läge och byggnadsmaterial spelar in. Dessa faktorer är förvisso ofta kända och skulle därför kunna tas med för att få en ännu bättre skattning av  $P(A)$ . Frågan blir då om det är värt besväret att göra en bättre skattning av  $P(A)$  eller om det i vissa fall kanske är bättre att helt enkelt göra en undersökning direkt istället för att räkna. När det sedan är känt huruvida det finns en skadlig miljöfaktor skulle det gå att, med hjälp av bland annat ekvation (28), beräkna väntevärdet av kostnaden för att antingen åtgärda problemet (om det finns) eller inte göra något.

En annan sak som visade sig i exemplet i avsnitt 6 var att om ett sjukdomsfall tilldelas en hög kostnad är det alltid mer lönsamt att utföra en mätning och vid behov vidta åtgärder som sänker radonhalten (se figur 5). Skulle en mindre allvarlig sjukdom betraktas skulle kanske ett annat resultat fås men det är ändå värt att påpeka eftersom det visar att det finns fall, som till exempel när det gäller radon, då det aldrig finns någon anledning att överväga att inte göra något.

I modellen som konstruerades i avsnitt 5 togs ekvationer fram för att beräkna sannolikheten för att ett givet sjukdomsfall, i situationer där en miljöfaktor existerar, är orsakat av just den miljöfaktorn (se ekvation (25) och (36)). Även då sådana beräkningar inte tas upp och används i något exempel är det ändå värt att nämna att det i somliga situationer kan vara viktigt att försöka ta reda på om det är en skadlig miljöfaktor som är orsaken till sjukdomen. Speciellt i försäkringsfrågor kan det vara viktigt att reda ut en sådan fråga. Det kan då handla om ersättning en person har rätt till om sjukdomen i fråga är orsakad av exponering från en riskfylld miljö.

### 8.3 Tillvägagångssätt i praktiken

Vad vi sett finns det ingen som använder en beslutsmodell liknande den vi konstruerat. VMC utför förvisso beräkningar och räknar, liksom vi i vår modell, på sannolikheten att en ansamling sjukdomsfall är slumpmässig. De skiljer sig dock från vår modell på så sätt att VMC försöker ta reda på vilka miljöfaktorer som finns och sedan bedömer huruvida denna faktor kan förhöja risken för sjukdom. För att demonstrera skillnaden mellan vår hantering och VMC:s betraktas återigen problemet med radon och lungcancer i avsnitt 6. Vi räknade på sannolikheten för en given radonhalt givet ett antal lungcancerfall medan VMC i samma situation skulle mätt radonhalten och sedan beräknat sannolikheten att det var radonet som orsakat cancerfallen.

Det återkommande svaret på frågan om hur beslut tas om eventuella åtgärder då det finns misstankar om en hälsofarlig miljöfaktor var att det alltid görs en undersökning. Det resoneras alltså ofta att det är gynnsamt att genomföra en undersökning om inte annat för att lugna de eventuellt utsatta, då det för det mesta är de utsatta själva som slår larm. Om det dessutom är något farligt ämne som hanteras i verksamheten finns det regler som styr och det är således redan en hårdare kontroll. Det är troligtvis därför som beräkningar jämförbara med våra inte utförs innan mätningar genomförs. Skulle det däremot existera någonting med en väldigt hög mätningkostnad skulle vår modell eventuellt kunna vara användbar.

En annan möjlig orsak till att en liknande modell inte används i stor utsträckning är osäkerheten, speciellt om vissa värden behöver skattas. Det blir eventuellt till och med billigare att bara mäta än att först försöka skatta något värde och sedan kanske ändå vara tvungen att mäta.

### 8.4 Slutsatser

Slutsatser som kan dras är att en händelse som kan verka osannolik inte behöver vara så konstig om det betraktas att den skulle kunna inträffa på många ställen. När det gäller att beräkna sannolikheten för en skadlig miljöfaktor är det i princip omöjligt att göra med någorlunda säkerhet när det är okänt vilken miljöfaktor det rör sig om. I sådana fall är det viktigt att undersöka hur "konstig" en händelse är för att veta hur stor oro den bör ge upphov till. Även när det gäller en känd miljöfaktor kan det vara svårt att ha en bra a priori skattning av sannolikheten och det kanske inte är billigare att göra en sådan skattning än att göra en mätning. I praktiken är det alltså bättre att fortsätta göra som det verkar fungera idag, det vill säga helt enkelt mäta om det finns en oro. (Det är klart att det skulle kunna vara ett problem om ingen blir orolig, men då skulle antagligen ingen komma på att räkna på det heller.) Så länge det inte uppstår en våg av oro eller en skadlig miljöfaktor som är dyr att undersöka existensen av är vår modell därmed inte särskilt användbar som hjälp vid beslutsfattande. Ekvation (25) kan dock användas för att ta fram sannolikheten att ett sjukdomsfall är orsakat av en skadlig miljöfaktor.

## Referenser

- [1] Elding L I, Haeger-Aronsen B, Astri Swedjemark G. Radon. 2012 [hämtat: 2012-02-05]. Tillgänglig: <http://www.ne.se>
- [2] Berger J *Statistical Decision Theory, Foundations, Concepts, and Methods*. United States of America: Springer-Verlag New York Publishing; 1980.
- [3] Weissen E. P-Value. 2012 [hämtat: 2012-02-28]. Tillgänglig: <http://mathworld.wolfram.com>
- [4] Altman D, Bland M. *Multiple significance tests: the Bonferroni method*. [www.bmj.com](http://www.bmj.com) (hämtat: 2012-02-22).
- [5] Rice J. *Mathematical Statistics And Data Analysis*. Andra upplagan. USA: International Thomson Publishing; 1995.
- [6] Curtin F, Schulz P. Multiple Correlations and Bonferroni's Correction. *Biological Psychiatry* 1998; 44(8): 775 -7 [http://dx.doi.org/10.1016/S0006-3223\(98\)00043-2](http://dx.doi.org/10.1016/S0006-3223(98)00043-2) [hämtat: 2012-02-22]
- [7] Nationalencyklopedin. Booles olikhet. 2012 [hämtat: 2012-02-23]. Tillgänglig: [www.ne.se](http://www.ne.se)
- [8] Mould R. *Introductory Medical Statistics*. 3:de upplagan. Bristol: Taylor Francis; 1998. Tillgänglig: <http://www.crcnetbase.com/isbn/9781439833681> [hämtat: 2012-02-28].
- [9] Värmlands Folkblad. *17 drabbade av cancer - tio döda*. <http://www.vf.se/nyheter/hagfors/17-drabbade-av-cancer-%E2%80%93-tio-doda> (hämtat: 2012-03-19).
- [10] DN. *Inget fel på polishuset i Hagfors*. <http://www.dn.se/nyheter/sverige/inget-fel-pa-polishuset-i-hagfors> (hämtat: 2012-03-19).
- [11] Lawson A B. *Bayesian Disease Mapping*. Boca Raton:Taylor Francis ; 2008. Tillgänglig: <http://www.crcnetbase.com/isbn/978158488413> [hämtat: 2012-02-18]
- [12] Värmlands Folkblad. *Polishuset ska expertundersökas*. <http://www.vf.se/nyheter/hagfors/polishuset-ska-expertundersokas> (hämtat: 2012-03-19).
- [13] Statistiska Centralbyrån. *Preliminär befolkningsstatistik per månad 2009-2012* [www.scb.se](http://www.scb.se) (hämtat: 2012-04-18).
- [14] Cancerfonden. *Om cancer* <http://www.cancerfonden.se/sv/Om-Cancerfonden/Fragor-och-svar/Vanliga-fragor-om-cancer> (hämtat: 2012-04-18).
- [15] Statistiska Centralbyrån. *Sveriges folkmängd (i ettårsklasser) 1860-2011*. [www.scb.se](http://www.scb.se) (hämtat: 2012-03-22).
- [16] Regionfakta. *Antalet arbetsställen men anställda 1990-2011*. <http://www.regionfakta.com/Varmlands-lan/Naringsliv/Antal-arbetsstallen-med-anstallda/> (hämtat 2012-03-13).
- [17] Statistiska Centralbyrån. *Aktuell statistik ur Företagsregistret*. [www.scb.se](http://www.scb.se) (hämtat: 2012-03-13).

- [18] Boverket. *Radon i inomhusmiljön - en konsekvensanalys av att införa WHO:s nya rekommendationer på radonvärden*  
<http://www.boverket.se/Global/Webbokhandel/Dokument/2011/Radon%20i%20inomhusmilj%C3%B6n.pdf> (hämtat: 2012-04-20)
- [19] Cancerfonden. *Statistik i korthet*.  
<http://www.cancerfonden.se/Global/Dokument/Cancerfonden/Cancerfondsrapporter/Statistik%20s15.pdf> (hämtat 2012-04-20)
- [20] Rönnqvist, T. Gammadata Mätteknik AB. *RADON ett inomhusmiljöproblem*  
[http://webcache.googleusercontent.com/search?q=cache:CrgeCh\\_Ck2oJ:www.miljosamverkanf.se/download/Tidigare%2520kurser/RadonGammaData\\_100204.ppt+](http://webcache.googleusercontent.com/search?q=cache:CrgeCh_Ck2oJ:www.miljosamverkanf.se/download/Tidigare%2520kurser/RadonGammaData_100204.ppt+)  
 (hämtat 2012-04-20)
- [21] Statistiska Centralbyrån. *Lägenheter i det kalkylerade bostadsbeståndet efter region, hustyp och tid* <http://www.ssd.scb.se/databaser/makro/SaveShow.asp> (hämtat 2012-04-20)
- [22] Naturvårdsverket. *Vem gör vad?*  
<http://www.miljomal.se/> (hämtat: 2012-05-04)
- [23] Socialstyrelsen. *Miljöhälsorapporter*  
<http://www.socialstyrelsen.se/publikationer/varastorapporter/miljohalsa>  
 (hämtat: 2012-05-06)
- [24] Länsstyrelsen Västra Götalands län. *Hälsorelaterad miljöövervakning*  
<http://www.lansstyrelsen.se/vastragotaland/sv/miljo-och-klimat/tilstandet-i-miljon/miljoovervakning/halsorelaterad-overvakning/Pages/index.aspx>  
 (hämtat: 2012-05-03).
- [25] Miljödepartementet. *Miljöbalk (1998:808)*  
<http://www.notisum.se/rnp/sls/lag/19980808.HTM> (hämtat: 2012-04-20).
- [26] Göteborgs Stad. *Miljötillsyn* <http://www.goteborg.se> (hämtat: 2012-04-24).
- [27] Göteborgs Stad, Miljö- och klimatnämnden. *Budget 2011*, ISSN 1401-243X. 2010.
- [28] Västra Götalandsregionens Miljömedicinska Centrum. *Bedömning av inomhusklimat i en bostadsfastighet i Majorna i Göteborgs kommun där hyresgästerna uppgett bostadsrelaterade symptom* [http://www.sahlgrenska.se/upload/SU/omrade\\_6/Arbets%20och%20Milj%C3%B6medicin/VMC/VMC/UtrednRapp/majorna\\_web.pdf](http://www.sahlgrenska.se/upload/SU/omrade_6/Arbets%20och%20Milj%C3%B6medicin/VMC/VMC/UtrednRapp/majorna_web.pdf)  
 (hämtat: 2012-05-08).
- [29] Arbets- och miljömedicin. *Om VMC*.  
<http://www.sahlgrenska.se/sv/SU/Omraden/6/arbetsochmiljomedicin/VMC/Om-VMC/> (hämtat 2012-05-08).
- [30] Arbets- och miljömedicin. *Ansamling av cancerfall - hur utreder vi?*  
[http://www.sahlgrenska.se/upload/SU/omrade\\_6/Arbets%20och%20Milj%C3%B6medicin/AMM/Rapporter/faktabladcluster100426.pdf](http://www.sahlgrenska.se/upload/SU/omrade_6/Arbets%20och%20Milj%C3%B6medicin/AMM/Rapporter/faktabladcluster100426.pdf)  
 (hämtat: 2012-05-08)
- [31] Strålsäkerhetsmyndigheten. *Hur farligt är radon?*  
<http://www.stralsakerhetsmyndigheten.se/Global/Publikationer/Broschyr/2009/faktablad-hur-farligt-ar-radon.pdf> (hämtat: 2012-02-28).
- [32] Socialstyrelsen. *Frågor och svar om radon*.  
<http://www.socialstyrelsen.se/fragorochsvar/radon> (hämtat: 2012-03-22)
- [33] Socialstyrelsen. *Varför inser inte Socialstyrelsen att radonsaneringar är onödiga? Rökare kommer ändå att dö av lungcancer*.  
<http://www.socialstyrelsen.se/halsoskydd/inomhusmiljo/radon/varforinserniinteattradonsaneringararodiga> (hämtat: 2012-03-15) .

- [34] Darby S et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ* 2005;330:223. doi: 10.1136/bmj.38308.477650.63 (hämtat: 2012-03-13).
- [35] Strålsäkerhetsmyndigheten. *Radonkällor*. [www.stralsakerhetsmyndigheten.se](http://www.stralsakerhetsmyndigheten.se) (hämtat: 2012-02-28).
- [36] Socialstyrelsen. *Förekomst av radon*.  
<http://www.socialstyrelsen.se/halsoskydd/inomhusmiljo/radon/forekomstavraddon>  
(hämtat 2012-03-22)
- [37] Strålsäkerhetsmyndigheten *Rikt- och gränsvärden för radon*.  
<http://www.stralsakerhetsmyndigheten.se/Allmanhet/Radon/Rikt-och-gransvar-den-for-radon/> (hämtat 2012-03-22)
- [38] Socialstyrelsen Tillsyn av radon i skolor och förskolor.  
<http://www.socialstyrelsen.se/halsoskydd/tillsyn/resultat-tillsynsprojekt/radon>  
(hämtat: 2012-03-22)
- [39] Nationalencyklopedin. Radonmätning. 2012 [hämtat: 2012-02-28]. Tillgänglig: [www.ne.se](http://www.ne.se)
- [40] Strålsäkerhetsmyndigheten. *Mätningar*.  
[www.stralsakerhetsmyndigheten.se/Allmanhet/Radon/Matningar](http://www.stralsakerhetsmyndigheten.se/Allmanhet/Radon/Matningar) (hämtat: 2012-02-28).
- [41] Strålsäkerhetsmyndigheten. *Radon*  
<http://www.stralsakerhetsmyndigheten.se/Allmanhet/Radon/> (hämtat 2012-04-27)
- [42] Strålsäkerhetsmyndigheten. *Åtgärder mot radon i bostäder*  
[http://www.boverket.se/Global/Webbokhandel/Dokument/2007/atgarder\\_%20mot%20\\_radon\\_%20i%20\\_bostader.pdf](http://www.boverket.se/Global/Webbokhandel/Dokument/2007/atgarder_%20mot%20_radon_%20i%20_bostader.pdf) (hämtat: 2012-03-30)
- [43] Radonguiden. *Kostnader för radonåtgärder*  
[http://www.radonguiden.se/kostnader\\_radon.asp](http://www.radonguiden.se/kostnader_radon.asp) (hämtat: 2012-03-30)



## A Radon

Radon ( $Rn$ ) är den tyngsta av ädelgaserna och bildas genom radioaktivt sönderfall av radium. De tre radonisotoperna genomgår själva  $\alpha$ -sönderfall till radioaktiva isotoper av polonium, vismut och bly som kallas radondöttrar[1].

### A.1 Skada

Vid inandning av radonhaltig luft kan det fastna radondöttrar i luftvägarna och därifrån avge  $\alpha$ -strålning som kan orsaka skador på lungorna. Radon ökar därför risken för lungcancer[31]. Det som avgör hur stor skada radon orsakar är den totala strålningsdosen man utsätts för genom hela livet. Kortvarig exponering av hög radonhalt har inte visat sig farligt, utan för att påverkas måste man exponeras dagligen under en längre period[32].

### A.2 Riskökning

Radon är, efter rökning, den näst vanligaste orsaken till lungcancer och beräknas orsaka cirka 500 av totalt cirka 2800 fall av sjukdomen i Sverige varje år. Av dessa 500 personer som får lungcancer är omkring 450 rökare och 50 icke-rökare[33]. Effekten av radon tar dock lång tid och det kan ta 15-40 år från det att en person utsätts för radon till att lungcancer kan diagnostiseras[31]. Vid radon i bostaden gäller att den proportionella EER, *Excess relative risk*), vilket är den relativa risken minus 1[8], ökar med cirka 0,16 per 100 Bq/m<sup>3</sup>. För en person som inte röker och inte är utsatt för radon är risken att drabbas av lungcancer innan 75 års ålder cirka 0,41% medan den för en ej radonutsatt rökare (15-24 cigaretter om dagen) är cirka 10,1%[34]. Eftersom grundrisken är mycket högre för rökare är även ökningen i risk från radon större för dem i absoluta tal.

### A.3 Förekomst

Radonhalten i en byggnad påverkas av byggnaden själv som var byggnaden står. I Skandinavien är berggrunden ofta mer radioaktiv än i många andra länder[1] och luft i marken kan därför ha hög radonhalt (5 000 – 2 000 000 Bq/m<sup>3</sup>). Då lufttrycket inomhus ofta är lägre än utomhus kan, särskilt om huset har en otät grund eller marken är extra luftgenomsläpplig, markluft med mycket radon sugas in i byggnaden[35].

Radon tillförs även inomhusluften från alla stenbaserade byggnadsmaterial. I de flesta fall är detta bidrag litet, men blå lättbetong kan orsaka skadligt höga radonhalter. Alunskifferbaserad blå lättbetong tillverkades 1929-1975 men kan ha använts så sent som 1980. Om materialet används i såväl inner- som ytterväggar och bjälklag kan det ge en radonhalt på 1000 Bq/m<sup>3</sup>[35].

En tredje vanlig radonkälla är dricksvattnet. Vatten från jordlager och berggrund innehåller alltid radon och därför kan vatten från bergborrade brunnar, kalkkällor och grävda brunnar som fylls på från sprickor i berget ha radonhalter som är högre än flera hundra Bq/l. Speciellt kan bergborrade brunnar ge vatten som har en radonhalt över 1000 Bq/l. Detta sker framförallt i områden där uranhalten i berget är hög, men låg uranhalt i berget är dock ingen garanti för låg radonhalt i vattnet. Däremot är radonhalten i grävda brunnar mycket lägre (10-300 Bq/l). Vid användning av radonhaltigt vatten avges mycket till inomhusluften. På ett ungefär gäller att om vattnet har en radonhalt på 1000 Bq/l höjer det radonhalten i inomhusluften med 100 Bq/m<sup>3</sup>[35].

Radonhalten i en byggnad är inte konstant utan varierar över tiden. Dessa variationer beror framför allt på vädret vilket påverkar lufttrycket i huset, luftväxlingarna och radoninflödet från marken[1].

Socialstyrelsens riktvärde för radon i bostäder och lokaler för allmänna ändamål ligger på 200 Bq/m<sup>3</sup>. Det uppskattas att cirka 400 000 svenska bostäder har en radonhalt som överstiger detta riktvärde, vilket innebär att cirka 10% av Sveriges invånare lever i ett hem där radonhalten är högre än gränsvärdet[36].

Under 2006/2007 mätte Socialstyrelsen radonhalten i skolor och förskolor i 145 av Sveriges 290 kommuner, vilka representerade cirka 70% av alla barn i åldern 1-17 år. Av 3714 skolor hade 394 för radonvärden över 200 Bq/m<sup>3</sup> och för 542 saknas uppgifter. För förskolor gäller att radonhalten var för hög i 218 av de 6819 förskolorna i kommunerna, men för 1048 av dem finns inga uppgifter alls[38].

#### A.4 Mätning och åtgärder

Radonhalten i inomhusluft mäts normalt med en spårfilm i dosa. Det är en plastfilm i vilken det bildas spår då den påverkas av alfastrålning (som radon avger)[39]. Eftersom radonhalten inte är konsant blir en mätning noggrannare desto längre tid den utförs och enligt strålsäkerhetsmyndigheten bör den vara minst tre månader. Om en förhöjd radonhalt konstateras måste man först av allt lokalisera källan så att lämpliga åtgärder kan sättas in[40]. Att utföra en mätning kostar vanligtvis några hundra kronor[41]. De lösningar som används för att sänka radonhalten i ett hus kan, i grova drag, delas in i fem olika typer av åtgärder:

- Tätning av huset för att hindra att det läcker in radon från marken
- Förbättring av ventilationssystemet
- Installation av F-/FT-ventilation
- Installation av radonsug, radonbrunn eller liknande åtgärder som sänker lufttrycket i marken under huset
- Om radonet kommer från byggnadsmaterial går det i vissa fall att byta ut materialet

[42] Kostnaden för de åtgärder som krävs för att sänka radonhalten kan variera från några tusen kronor till över 100 000 kronor beroende på vilken eller vilka åtgärder som behöver sättas in[43].

## B Kod

I denna bilaga återfinns kod till de program i MATLAB och Mathematica som har använts för att utföra beräkningar och producera diagram.

### B.1 Exempel med polishuset i Hagfors

Koden nedan är den som användes för beräkningarna i avsnitt 3:

```
%%
% F r att testa ifall nollhypotesen kan f rkastas
% ver 20 r
n=100;
p=0.0058*20 %antal personer som drabbas av cancer per r multiplicerat med antal r
alpha=0.05;
k=17;

P=zeros(k,1);

%P(K>=17)

for i=1:k
    P(i)=((factorial(n)/(factorial(i-1)*factorial(n-(i-1))))*(p^(i-1))*((1-p)^(n-(i-1))));
end

Pr=1-sum(P)

if (Pr>alpha)
    disp('Kan ej f rkasta')
else
    disp('F rkasta hypotes')
end

%%
% F r att testa ifall nollhypotesen kan f rkastas
% ver 32 r
n=100;
p=0.0058*32 %antal personer som drabbas av cancer per r multiplicerat med antal r
alpha=0.05;
k=17;

P=zeros(k,1);

for i=1:k
    P(i)=((factorial(n)/(factorial(i-1)*factorial(n-(i-1))))*(p^(i-1))*((1-p)^(n-(i-1))));
end

Pr=1-sum(P)

if (Pr>alpha)
    disp('Kan ej f rkasta hypotes')
else
    disp('F rkasta hypotes')
end

%%
% F r att hitta hur m nga som beh ver drabbas f r att nollhypotesen ska
% f rkastas

% 20 r
n=100;
p=0.0058*20 %55000/9000000=0.0061, antal pers som drabbas av cancer varje r
alpha=0.05; %Bara Hagforspolisen
alpha=9.63e-7 % Fallet med Sveriges befolkning mellan 30 och 75 r
alpha=1.36654e-6 % Arbetsplatser i Sverige med r tt antal anst llda
alpha=4.99922e-5 % Arbetsplatser i Vrmland med r tt antal anst llda
k=0;
```

```

P=zeros(n,1);
while k<n

    for i=1:k
        P(i)=((factorial(n)/(factorial(i-1)*factorial(n-(i-1))))*(p^(i-1))*((1-p)^(n-(i-1))));
    end

    Pr=1-sum(P);

    if Pr<alpha
        disp('Antalet sjukdomsfall som kr vs f r att frkasta nollhypotesen r ')
        k
        k=n;
    else
        k=k+1;
    end

end

%%
% F r att hitta hur m nga som beh ver drabbas f r att nollhypotesen ska
% frkastas

```

```

% 32 r
n=100;
p=0.0058*32 %55000/9000000=0.0061, antal pers som drabbas av cancer varje r
%alpha=0.05; %Bara Hagforspolisen
%alpha=9.63e-7 % Fallet med Sveriges befolkning mellan 30 och 75 r
%alpha=1.36654e-6 % Arbetsplatser i Sverige med r tt antal anst llda
alpha=4.99922e-5 % Arbetsplatser i Vrmland med r tt antal anst llda
k=0;

```

```

P=zeros(n,1);
while k<n

    for i=1:k
        P(i)=((factorial(n)/(factorial(i-1)*factorial(n-(i-1))))*(p^(i-1))*((1-p)^(n-(i-1))));
    end

    Pr=1-sum(P);

    if Pr<alpha
        disp('Antalet sjukdomsfall som kr vs f r att frkasta nollhypotesen r ')
        k
        k=n;
    else
        k=k+1;
    end

end

```

För beräkningarna till tabell 2, det vill säga testen för om det är konstigt att Hagfors incidenten händer någon grupp i Sverige, Värmland etc. användes Mathematica

```

(*grupper om 100 pers mellan 30-75 år i Swe*)
n = 53239;
p = 0.0684;
alfaprime = 0.05;
k = 0;
P0 = (n!/(k!*((n - k)!)))*(p^k)*(1 - p)^(n - k);
P = 1 - P0
alfa = 1 - (1 - alfaprime)^(1/n)

```

```

(*Arbetsställen i Swe*)
n = 37535;
p = 0.0684;

```

```

alfaprime = 0.05;
k = 0;
P0 = (n!/(k!*((n - k)!)))*(p^k)*(1 - p)^(n - k);
P = 1 - P0
alfa = 1 - (1 - alfaprime)^(1/n)

(*Arbetsställen i Värmland*)
n = 1026;
p = 0.0684;
alfaprime = 0.05;
k = 0;
P0 = (n!/(k!*((n - k)!)))*(p^k)*(1 - p)^(n - k);
P = 1 - P0
alfa = 1 - (1 - alfaprime)^(1/n)

(*Poliser i Swe*)
n = 280;
p = 0.0684;
alfaprime = 0.05;
k = 0;
P0 = (n!/(k!*((n - k)!)))*(p^k)*(1 - p)^(n - k);
P = 1 - P0
alfa = 1 - (1 - alfaprime)^(1/n)

```

## B.2 Okända miljöfaktorer

Koden för att producera figurerna i avsnitt 4 finns nedan

```

close all
%%
%3D plot
n=100;
k=17;

PA=linspace(0,1,50);
r=linspace(1,2,50);

p=0.0058*20; %Sannolikheten att en person f r cancer f re 75 rs lder

[PAmesh, R]=meshgrid(PA,r);

lambda=(1-PAmesh)+R.*PAmesh;

PABbin=((((R*p./lambda).^k).*(1-R*p./lambda).^(n-k))./(((R*p./lambda).^k).*((1-R*p./lambda).^(n-k)).*PAmesh+((p./lambda).^k).*((1-p./lambda).^(n-k)).*(1-PAmesh))).*PAmesh;

figure
mesh(PAmesh,R, PABbin)
colormap(Cool)
xlabel('A priori f rdelning f r P(A)')
ylabel('Olika v rden p den relativa risken, r')
title('P(A|17 sjuka i Hagfors)')

figure
contourf(PAmesh,R,PABbin)
colorbar
xlabel('A priori f rdelning f r P(A)')
ylabel('Olika v rden p den relativa risken, r')
title('P(A|17 sjuka i Hagfors)')

%%
%Fixa RR

```

```

n=100;
k=17;

PA=linspace(0,1,50);
RR=1.1;
p=0.0058*20;
lambda=(1-PA)+RR*PA;

PABbin=((((RR*p./lambda).^k).*(1-RR*p./lambda).^(n-k))./(((RR*p./lambda).^k).*((1-RR*p./lambda).^(n-k)).*PA+((p./lambda).^k).*((1-p./lambda).^(n-k)).*(1-PA))).*PA;

figure
plot(PA,PABbin)
xlabel('A priori fördelning för P(A)')
ylabel('P(A|17 sjuka i Hagfors)')
title('Relativa risken r=1,1')

%%
%Fix P(A)

n=100;
k=17;

PA=0.5;
RR=linspace(1,2,50);
p=0.0058*20;
lambda=(1-PA)+RR*PA;

PABbin=((((RR*p./lambda).^k).*(1-RR*p./lambda).^(n-k))./(((RR*p./lambda).^k).*((1-RR*p./lambda).^(n-k)).*PA+((p./lambda).^k).*((1-p./lambda).^(n-k)).*(1-PA))).*PA;

figure
plot(r,PABbin)
xlabel('Relativa risken r')
ylabel('P(A|17 sjuka i Hagfors)')
title('P(A)=0,5')

```

### B.3 Exempel med radon som miljöfaktor

Här är den kod som använts i avsnitt 6:

```

close all
%Kostnader
Ku=500;
Ka=50000;
Ks=22000000;

%Enklaste fallet
figure()
N=1;
n=100;
k=0:n/2;
pA=(0.1*0.0269382.^k.*0.973062.^(-k+n))./(0.1*0.0269382.^k.*0.973062.^(-k+n)+0.9*0.023118.^k.*0.976882.^(-k+n));
plot(k,pA)
grid on
title('Sannolikheten för en radonhalt p vid 200 Bq/m^3')
xlabel('Antalet sjuka av totalt 100 personer')
ylabel('Sannolikhet')
p=0.0235;
lambda=1.016525;
r=1.16525;

Ehandling=Ku+Ka*pA;
EejHandling=Ks*N*p*(r-1)/lambda*pA;
figure()
plot(k,Ehandling,k,EejHandling)
legend('Handling','Ej Handling')

```

```

grid on
title('Kostnaden f r olika ageranden om N=1')
xlabel('Antalet sjuka av totalt 100 personer')
ylabel('Kostnad (SEK)')

%N stan enklaste
%1=r kare 2=icke-r kare
figure()
n1=100;
k1=0:n1/2;
n2=100;
k2=0:n2/2;
[K1, K2]=meshgrid(k1,k2);
pA2=(0.1* 0.005412.^K2.* 0.13332.^K1.* 0.86668.^(-K1 + n1).*0.994588.^(-K2 + n2))
./((0.1*0.005412.^K2.* 0.13332.^K1.*0.86668.^(-K1 + n1).*0.994588.^(-K2 + n2) +0.9*
0.00464448.^K2.*0.114413.^K1.*0.885587.^(-K1 + n1).*0.995356.^(-K2 + n2));
mesh(K1, K2, pA2)
grid on
figure()
contourf(K1,K2,pA2)
xlabel('Antalet sjuka av totalt 100 r kare')
ylabel('Antalet sjuka av totalt 100 icke-r kare')
title('Sannolikheten f r en radonhalt p 200 Bq/m^3')

%Flera m jliga radonhalter

%Uppdelat i r kare och icke-r kare
n1=100;
k1=0:n1/2;
n2=100;
k2=0:n2/2;
[K1, K2]=meshgrid(k1,k2);

A1=[0; 200; 400; 1000]; %Olika m jliga radonhalter
PA1=[0.877; 0.097; 0.024; 0.002];
r100=0.16;

p1A1=zeros(4,1); %risken f r r kare
p1A1(1)=0.101;

p2A1=zeros(4,1); %risken f r icke-r kare
p2A1(1)=0.0041;
for l=2:4
    p1A1(l)=(1+r100*A1(l)/100)*p1A1(l);
    p2A1(l)=(1+r100*A1(l)/100)*p2A1(l);
end
p1=0;
for l=1:4
    p1=p1+p1A1(l)*PA1(l);
end

PB=0;
for i=1:4
    PB=PB+binopdf(K1,n1,p1A1(i)).*binopdf(K2,n2,p2A1(i))*PA1(i);
end

pAlB=zeros(length(k1),length(k2),4);

for i=1:4
    pAlB(:, :, i)=(binopdf(K1,n1,p1A1(i)).*binopdf(K2,n2,p2A1(i)))./PB*PA1(i);
end

maxA=zeros(size(PB)); %Tar fram mest sannolika radonhalt
for i=1:length(k1)
    for j=1:length(k2)
        [C, index]=max(pAlB(i, j, :));
        maxA(i, j)=A1(index);
    end
end

```

```

end
figure()
mesh(K1,K2,maxA)

figure()
contourf(K1,K2,maxA)
xlabel('Antalet sjuka av totalt 100 r kare')
ylabel('Antalet sjuka av totalt 100 icke-r kare')
title('Den mest sannolika radonhalten i Bq/m^3')

figure()
contourf(K1,K2,pAlB(:, :, 1))
figure()
mesh(K1,K2,pAlB(:, :, 2))

%Flera m jliga radonhalter, en grupp av personer
N=1;
n=100;
k=0:n/2;
Al=[0; 200; 400; 1000];
PA1=[0.877; 0.097; 0.024; 0.002];
r100=0.16;
p=0.0235; %Sannolikheten att slumpm ssig person blir sjuk
rAl=(1+r100*Al/100); %Den relativa risken vid olika Al

lambda=dot(PA1,rAl);

pAl=p/lambda*rAl; %Sannolikheten att bli sjuk om man r utsatt f r Al;

PB=0;
for i=1:4
    PB=PB+binopdf(k,n,pAl(i))*PA1(i);
end

pAlB=zeros(length(k),4);
for i=1:4
    pAlB(:,i)=binopdf(k,n,pAl(i))./PB*PA1(i);
end
figure()
plot(k,pAlB(:,1),k,pAlB(:,2),k,pAlB(:,3),k,pAlB(:,4))
legend('0 Bq/m^3', '200 Bq/m^3', '400 Bq/m^3', '1000 Bq/m^3')
xlabel('Antalet sjuka av toltalt 100 personer')
ylabel('Sannolikhet')
title('Sannolikheten for olika radonhalter')

N=1;
Ehandling=Ku;
EejHandling=0;
for i=2:4
    Ehandling=Ehandling+Ka*pAlB(:,i);
    EejHandling=EejHandling+Ks*N*p*(rAl(i)-1)/lambda*pAlB(:,i);
end
figure()
plot(k,Ehandling,k,EejHandling)
legend('Handling','Ej Handling')
ylabel('Kostnad (SEK)')
xlabel('Antalet sjuka av totalt 100 personer')
title('Kostnaden f r olika ageranden om N=1')
grid on

```