



Stockholms
universitet

Uppskattning av mörkertal i sjukanmälningar med hjälp av fångst-återfångstmetoden

Kaidi Oja

Kandidatuppsats 2009:10
Matematisk statistik
September 2009

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Uppskattning av mörkertal i sjukanmälningar med hjälp av fångst-återfångstmetoden

Kaidi Oja*

September 2009

Sammanfattning

För att begränsa förekomsten och vidare spridning av smittsamma sjukdomar måste enligt Smittskyddslagen alla anmälningspliktiga infektionssjukdomar rapporteras till Smittskyddsinstitutet (SMI). Enligt reglerna ska anmälan ske från laboratoriet och behandlande läkare till smittskyddsläkaren i respektive län och till SMI. Tyvärr görs inte detta alltid. I bästa fall görs åtminstone en anmälan, men tyvärr kan man misstänka att det finns ett mörkertal av diagnostiserade sjukdomsfall som inte rapporteras.

Syftet med detta arbete är att undersöka och jämföra olika statistiska metoder för att uppskatta mörkertalet i sjukanmälningar. Standard fångst-återfångstmetoden jämförs med bootstrap och bayesianska metoder. Alla metoder ger ett bra resultat men vid små datamängder är bayesianska metoder att föredra. För närvarande finns det 58 anmälningspliktiga sjukdomar. I denna uppsats beräknas mörkertalen för 8 av dem: campylobakterinfektion, harpest, hepatit A, meningokockinfektion, salmonellainfektion, sorkfeber, pneumokockinfektion (penicillinresistent) och tuberkulos.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: kaidioja@hotmail.com. Handledare: Mikael Andersson.

Förord

Denna uppsats utgör ett examensarbete om 15 högskolepoäng och leder till en kandidatexamen i matematisk statistik vid Matematiska institutionen, Stockholms Universitet. Examensarbetet har utförts på uppdrag av Smittskyddsinstitutet.

Jag vill rikta ett stort tack till Mikael Andersson, min handledare på matematiska institutionen vid Stockholms universitet, som har besvarat alla mina frågor och gett goda råd under arbetets gång.

Jag vill även tacka Anna Jansson, SMI , som har besvarat diverse frågor rörande epidemiologi.

Innehåll

1 Introduktion.....	3
1.1 Syfte och bakgrund.....	4
1.2 Metod	5
2 Teori.....	6
2.1 Fångst-återfångstmetoden.....	6
2.2 Bootstrapmetoden.....	8
2.3 Bayesiansk analys.....	9
2.3.1 Beta/binomial.....	11
2.3.2 Dirichlet/multinomial.....	12
3 Illustration av metoderna	12
3.1 Fångst-återfångstmetoden.....	12
3.2 Bootstrapmetoden.....	13
3.3 Bayesiansk analys.....	14
4 Resultat.....	15
5 Diskussion: argument kring de olika metoderna.....	19
6 Bilagor.....	21

1 Introduktion

Enligt Smittskyddslagen måste alla diagnostiserade anmälningspliktiga infektionssjukdomar rapporteras till Smittskyddsinstitutet (SMI). Syftet med lagen är att begränsa förekomsten av smittsamma sjukdomar inom landet genom övervakning och genom att försöka förhindra eller begränsa vidare spridning.

I dagsläget består listan av anmälningspliktiga sjukdomar av 58 sjukdomar och indelas på följande sätt ¹:

- (1) allmänfarlig sjukdom är sjukdom som kan vara livshotande, innebära långvarig sjukdom eller svårt lidande, t ex campylobakterinfektion, hepatit A-E, salmonellainfektion, tuberkulos. Det finns sammanlagt 25 allmänfarliga sjukdomar.
- (2) samhällsfarliga sjukdomar är sjukdomar som kan få en spridning i samhället som kan innebära en allvarlig störning. Det finns 2 samhällsfarliga anmälningspliktiga sjukdomar; smittkoppar och sars
- (3) sjukdomar utöver allmänfarliga sjukdomar som kan innebära hot mot människors hälsa. Det finns 31 sjukdomar som klassas som utöver allmänfarliga, t ex harpest, kikhosta, malaria, stelkramp.

Anmälningar kommer in via SmiNet som är ett system där sjukdomsfall rapporteras elektroniskt. Systemet introducerades 1997 och uppgraderades 2004.

Anmälan ska ske både av behandlande läkare (klinisk anmälan) och av den laboratorieläkare som diagnostiserar det smittämne som orsakat en anmälningspliktig sjukdom. Tyvärr görs inte detta alltid. Ibland görs endast en anmälan och man misstänker att ett visst antal sjukdomsfall inte blir registrerade hos SMI.

¹ SMI:s hemsida www.smi.se

1.1 Syfte och bakgrund

Syftet med denna uppsats är att undersöka och jämföra olika statistiska metoder för att uppskatta antal sjukdomsfall som varken blir rapporterade av behandlande läkare eller av laboratorieläkare och som SMI därmed inte får någon information om. Det så kallade mörkertalet uppskattas vanligtvis enligt fångst-återfångstmetoden som används hos ekologer för uppskattning av djurpopulationer. Analysen är utförd retrospektivt på data från åren 2002-2003 och 2007-2008.

Data som används för denna uppsats innehåller uppgifter om namn på sjukdom, landsting där den konstaterades samt datum för anmälan från behandlande läkare och från laboratorieläkare.

Denna uppsats behandlar 8 anmälningspliktiga sjukdomar som valdes ut att representera egenskaper som olika typer av smittspridning och farlighetsgrad samt välkända versus nyintroducerade sjukdomar. Med anledning av dessa olikheter valdes följande sjukdomar ut: campylobakterinfektion, harpest, hepatit A, meningokockinfektion, PRP², salmonellainfektion, sorkfeber och tuberkulos. Tuberkulos analyseras på data endast från 2007 och 2008 eftersom sjukdomen tidigare ingått i ett separat anmälningsystem. Sjukdomsinformation om de olika diagnoserna finns i Bilaga 1.

Icke-rapporteringen av anmälningspliktiga sjukdomsfall kan bero på olika faktorer. Tidsbrist, slarv eller okunskap kan vara möjliga orsaker. Därför är det även av intresse att se hur rapporteringen har förändrats genom åren.

² Pneumokockinfektion – penicillinresistent

1.2 Metod

Mörkertal i sjukanmälningar kan uppskattas med hjälp av den så kallade fångst-återfångst metoden. I korthet går den ut på följande:

Man fångar in ett visst antal individer (n_1), märker dessa på något sätt för att sedan släppa ut dem igen. Efter en viss tid har gått fångar man in ytterligare ett antal individer (n_2) och räknar antal märkta av dem (k). Om man antar att djuren har blandat sig slumpmässigt mellan fångstillfällena och att populationen är konstant kan man betrakta n_2 som en observation från en hypergeometrisk fördelning med parametrar n_1 , n_2 och N . ML-skattningen av hela populationen N blir

$$\hat{N} = \frac{n_1 \cdot n_2}{k}$$

I fallet med mörkertal i sjukanmälningar kan man betrakta läkaranmälningar som första och laboratorieanmälningar som andra fångsten och uppskatta totala antalet sjukdomsfall. Med sjukdomsfall menas fall med diagnos verifierat av laboratoriet.

För att kunna tillämpa standard fångst-återfångstmetoden måste man anta att

- 1) händelsen att ett fall rapporteras av ett laboratorium är oberoende av händelsen att fallet rapporteras av en läkare
- 2) anmälningssannolikheter är konstanta; sannolikheten att ett fall rapporteras är p_1 för alla läkare och p_2 för alla laboratorier i Sverige

För att undersöka om antagandet om konstanta rapporteringssannolikheter stämmer för hela nationen uppskattas mörkertalet på nationsnivå och på landstingsnivå. Nationsnivå innebär att sjukdomsfall summeras över alla landsting och därefter uppskattas antal icke-rapporterade sjukdomsfall. Landstingsnivå innebär att mörkertalet skattas separat för varje landsting för att sedan summeras. Stora skillnader i resultaten skulle indikera på stora skillnader i rapporteringssannolikheter. Små skillnader kan ses som ett tecken på stabila sannolikheter.

För skattningar på nationell nivå finns det tillräckligt stora datamängder och antalet

observationer är oftast inte ett problem. Däremot kan det bli problematiskt när mindre bestånd (till exempel enskilda landsting) behöver analyseras. Därför har även datorintensiva metoder som bootstrap och bayesianska procedurer tillämpats för att komma till rätta med dessa problem och öka säkerheten att skattningarna håller.

2 Teori

I detta kapitel presenteras bakgrundsteori för att enklare se hur olika metoder implementeras vid uppskattning av mörkertal.

2.1 Fångst-återfångstmetoden

Mörkertalestimeringen har utförts enligt fångst-återfångstmetoden som härstammar från biologin där den används för att skatta djurpopulationer, till exempel antal fiskar i en sjö.

Metoden fungerar på följande sätt: man fångar in ett visst antal individer, märker dessa på något sätt och släpper ut dem igen. Efter en viss tidsperiod fångar man in ytterligare ett antal djur och räknar antalet av dessa som är märkta.

För att uppskatta mörkertal i sjukankmälningar kan man betrakta läkaranmälningarna som första och laborationsanmälningarna som andra fångsten.

I tabell 1 nedan beskrivs fördelningen av totala antalet sjukdomsfall med hjälp av en 2x2 kontingenstabell. I tabellen betecknar n_{11} antal sjukdomsfall rapporterade av både läkare och klinik, n_{12} antal rapporterade av laboratorium men inte av klinik och n_{21} av klinik men inte laboratorium. Det sökta antalet sjukdomsfall som varken rapporteras av klinik eller lab betecknas n_{22} .

Beteckna det totala antalet sjukdomsfall med N . Enligt Tabell 1 är $N = n_{11} + n_{12} + n_{21} + n_{22}$

Tabell 1: schematisk beskrivning av totala antalet sjukdomsfall med hjälp av en 2x2 kontingenstabell

		Klinik		Σ
		Ja	Nej	
Laboratorium	Ja	n_{11}	n_{12}	$n_{1.}$
	Nej	n_{21}	n_{22}	$n_{2.}$
Σ		$n_{.1}$	$n_{.2}$	$n_{..}$

Om man antar att händelserna att ett fall rapporteras av en läkare och att ett fall rapporteras av ett laboratorium är oberoende gäller det att $p_{\text{klinik/lab}} = p_{\text{klinik/ej lab.}}$

Enligt tabell 1 medför detta att

$$\frac{\hat{p}_{11}}{\hat{p}_1} = \frac{\hat{p}_{21}}{\hat{p}_2}$$

vilket är likvärdigt med

$$\frac{n_{11}}{n_{11} + n_{12}} = \frac{n_{21}}{n_{21} + n_{22}}$$

Löser man ut det sökta talet n_{22} från denna ekvation får man att

$$\hat{n}_{22} = \frac{n_{12} \cdot n_{21}}{n_{11}} \quad (1)$$

Totala antalet sjukdomsfall blir

$$\hat{N} = n_{11} + n_{12} + n_{21} + \frac{n_{12} \cdot n_{21}}{n_{11}} = \frac{n_{1.} \cdot n_{.1}}{n_{11}} \quad (2)$$

och variansen av totala antalet sjukdomsfall

$$Var(\hat{N}) = \frac{n_{.1} \cdot n_{.2} \cdot n_{12} \cdot n_{21}}{n_{11}^3} \quad (3)$$

\hat{N} är approximativt normalfördelat.

95 % konfidensintervall för \hat{N} ges av $\hat{N} \mp 1.96 \cdot \sqrt{Var(\hat{N})}$

2.2 Bootstrappmetoden

Bootstrappmetoden används för att granska normalapproximationen som ligger till grund för konfidensintervallen i fångst-återfångstmetoden. Metoden används alltså inte för själva skattningen av mörkertalet utan för att uppskatta varians och beräkna konfidensintervall.

Metoden som introducerades av Efron 1979 grundar sig på idén att ett bootstrappstickprov förhåller sig till ett stickprov som ett stickprov förhåller sig till populationen.

Bootstrappmetoden går i korthet ut på att genom återläggning dra ett antal observationer från ett befintligt urval konstruera ett nytt. För varje nykonstruerat urval kan man sedan skatta parametrar av intresse (till exempel medelvärde). Resultatet ger en bild av samplingsfördelningen som i sin tur ligger till grund för konfidensintervall.

Enligt Tabell 1 kan fördelningen av rapporterade sjukdomsfall betraktas som en tvåvägs kontingenstabell med två rader och två kolumner, där n_{ij} betecknar antal observationer på rad i och kolumn j och n_{22} är ej observerad.

Låt $\tilde{p}_{ij} = \frac{n_{ij}}{\tilde{n}}$ beteckna sannolikheten att en observation hamnar i cell (i,j) ,

$$\tilde{n} = n_{11} + n_{12} + n_{21} \text{ är fixt och } \tilde{p}_{11} + \tilde{p}_{12} + \tilde{p}_{21} = 1$$

3 A. Agresti: Categorical Data Analysis; Second Edition (2002)

Då gäller att celltalen n_{11} , n_{12} och n_{21} följer en multinomialfördelning med parametrar

$$\tilde{n}, \tilde{p}_{11}, \tilde{p}_{12}, \tilde{p}_{21}.$$

$$(n_{11}, n_{12}, n_{21}) \sim \text{Trinomial}(\tilde{n}, \tilde{p}_{11}, \tilde{p}_{12}, \tilde{p}_{21}) \quad (4)$$

Eftersom den empiriska sannolikhetsfördelningen är känd går bootstrappmetoden i det här fallet ut på att med återläggning och oberoende av varandra generera ett antal slumpstal från multinomialfördelningen. I nästa steg beräknas mörkertal enligt (1) för varje stickprov. Från detta urval kan man sedan skatta varians och beräkna konfidensintervall.

Låt $B_{(1)}, B_{(2)}, \dots, B_{(T)}$ beteckna det i storleksordning sorterade urvalet. Konfidensintervall för n_{22} är då

$$B_{(\text{undre})} < \hat{n}_{22} < B_{(\text{övre})} \quad \text{där}$$

$$\begin{aligned} (\text{undre}) &= \left[\frac{(T+1) \cdot \alpha}{2} \right] \\ (\text{övre}) &= \left[(T+1) \cdot (1 - \alpha/2) \right] \end{aligned}$$

[] betecknar avrundning till närmaste heltal.

Till exempel blir 95 % konfidensintervall för $T=10000$

$$B_{(250)} < \hat{n}_{22} < B_{(9750)} \quad ^4.$$

2.3 Bayesiansk analys

För skattningar av mindre vanliga sjukdomar och även för skattningar på landstingsnivå är små datamängder ett problem. Enligt formel (1) skattas mörkertalet till noll om en av n_{12} och n_{21} är lika med noll. För att komma till rätta med detta problem skattas

4 A.C Davison & D. V. Hinkley: Bootstrap Methods and their Application. (1997)

mörkertalet även enligt bayesianska metoder.

Bayesiansk analys består av praktiska metoder för att dra slutsatser från data med utgångspunkt från sannolikhetsmodeller och innebär oftast omfattande numeriska beräkningar. Skillnaden från den ”klassiska” statistiken är att man betraktar data som fixa medan parametrar som inte kan observeras är slumpmässiga. En icke-Bayesian ser parametern som fix och data som utfall av stokastiska variabler.

Bayesiansk analys kan delas upp i tre steg:

- 1) Formulering av likelihood för data givet parametrarna
- 2) Formulering av a priori fördelning för parametrarna
- 3) Med hjälp av Bayes teorem beräkna fördelningen av de icke observerade parametrarna givet observerade data, det vill säga härledning av a posteriorifördelningen

Bayesiansk statistik bygger på användning av Bayes sats som är en matematisk formel formulerad redan på 1700-talet och som kan anses som en formel för att 'vända på' betingning.

Bayes formel (kontinuerlig)

$$p(\theta/y) = \frac{p(y/\theta) \cdot p(\theta)}{\int_{-\infty}^{\infty} p(y/\theta') \cdot p(\theta') d\theta'} \quad (5)$$

Eftersom nämnaren är en normeringskonstant så att $p(\theta/y)$ blir en sannolikhetsfördelning skriver man för enkelhets skull

$$p(\theta/y) \propto p(y/\theta) \cdot p(\theta) \quad (6)$$

där symbolen \propto står för 'proportionell mot'.

$p(y/\theta)$ är betingad sannolikhet för mätdata givet θ och kallas likelihoodfunktionen, $p(\theta)$ betecknar fördelningen för den stokastiska variabeln θ och kallas a-priori-fördelningen. $p(\theta/y)$ kallas a-posteriori-fördelningen och kan tolkas som korrigeringen av a-priori-fördelningen med hjälp av den information observerad data ger.

En a-priori-fördelning bör vara konjugerad med en likelihoodfunktion vilket innebär att a-posteriori-fördelningen tillhör samma fördelningsfamilj som a-priori-fördelningen.

Som illustration av metoden följer nedan härledning av a-posteriori-fördelningen för mätdata som är Bin(n, θ)-fördelade. Härledning av a-posteriori-fördelningen för en multinomial är en generalisering av detta.

2.3.1 Beta/binomial

Den konjugerade a-priori-fördelningen för binomial är betafördelning.

Täthetsfunktionen för en $Beta(\alpha, \beta)$ -fördelning är

$$f(\theta) = \frac{1}{B(\alpha, \beta)} \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

där $0 < \theta < 1, \alpha > 0, \beta > 0$ och $B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} d\theta = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}$.

För en Beta(s, t) är väntevärdet $\frac{\alpha}{\alpha + \beta}$ och variansen $\frac{\alpha \cdot \beta}{(\alpha + \beta)^2 \cdot (\alpha + \beta + 1)}$.

Om vi nu låter θ ha a-priori-fördelningen $Beta(\alpha, \beta)$ och vill räkna ut a-posteriori-fördelningen för Bin(n, θ)-fördelade data använder vi Bayes formel (6)

I detta fall blir resultatet

$$p(\theta/y) \propto \theta^y \cdot (1-\theta)^{n-y} \cdot \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

som identifieras som en $Beta(\alpha + y, \beta + n - y)$.

2.3.2 Dirichlet/multinomial

Dirichletfördelningen är multivariat generalisering av betafördelningen och konjugerade a-priori-fördelningen för multinomial.

Täthetsfunktionen för Dirichletfördelningen med parametrar $\alpha = (\alpha_1, \dots, \alpha_n)$ är

$$p(\theta) = \text{Dirichlet}(\theta; \alpha) = \frac{1}{B(\alpha)} \cdot \prod_{i=1}^n \theta_i^{\alpha_i - 1}$$

där $\theta_1, \dots, \theta_n \geq 0$; $\sum_{i=1}^n \theta_i = 1$; och $\alpha_1, \dots, \alpha_n > 0$.

$$B(\alpha) = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^n \alpha_i)}$$
 är normeringskonstant.

Vid multikategorisk data är likelihoodfunktionen $\text{Multinomial}(N, p_1, p_2, \dots, p_n)$.

Genom att tillämpa Bayes formel (6) får man a-posteriori-fördelningen $\text{Dirichlet}(\alpha_1 + x_1, \dots, \alpha_n + x_n)$.

Speciellt gäller det att om man låter $\alpha = 1$ (det vill säga att om vi har en likformig a-priori-fördelning på $[0,1]$) så blir a-posteriori-fördelningen en $\text{Dirichlet}(x_1+1, \dots, x_n+1)$ -fördelning.

3 Illustration av metoderna

Mörkertal har uppskattats enligt standard fångst-återfångstmetoden, bootstrapmetoden och genom att tillämpa bayes formel för åren 2003, 2004, 2007 och 2008. Här fokuseras dock på en av dessa, 2008. Avsnittet utgörs av jämförelse av metoderna som presenteras i tabellform.

3.1 Standard fångst-återfångstmetoden

I tabellerna nedan presenteras resultat för mörkertal estimerats enligt standard fångst-

återfångstmetoden. Mörkertalet har uppskattats enligt formel (1) och variansen enligt formel (3). För att räkna ut 95 % konfidensintervall för mörkertalet har denna räknats ut först för det estimerade totalantalet enligt $\hat{N} \pm 1.96 \cdot \sqrt{Var(\hat{N})}$ och sedan subtraheras totalantalet rapporterade sjukdomsfall. Om 2.5 % kvantilen för totalantalet sjukdomsfall uppskattas vara mindre än totalt registrerade sjukdomsfall blir denna negativ. Som man kan se i tabellerna nedan så inträffar detta ganska ofta vilket kan tas som en indikation på att normalapproximationen som ligger till grund för konfidensintervallen är dålig.

Tabell 2 :mörkertal (\hat{n}_{22}) beräknat enligt standard fångst-återfångstmetoden på *nationsnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	1,68	1,32	-0,91	4,26
Harpest	0,26	0,54	-0,80	1,31
Hepatit A	0,00	0,00	0,00	0,00
Meningokockinfektion	0,00	0,00	0,00	0,00
PRP	2,28	1,67	-0,99	5,54
Salmonellainfektion	0,69	0,84	-0,96	2,34
Sorkfeber	0,49	0,72	-0,92	1,89
Tuberkulos	0,30	0,63	-0,93	1,54

Tabell 3: mörkertal (\hat{n}_{22}) beräknat enligt standard fångst-återfångstmetoden på *landstingnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	1,51	1,27	-0,99	4,00
Harpest	0,33	0,77	-1,18	1,84
Hepatit A	0,00	0,00	0,00	0,00
Meningokockinfektion	0,00	0,00	0,00	0,00
PRP	3,14	2,53	-1,81	8,08
Salmonellainfektion	0,94	1,01	-1,03	2,91
Sorkfeber	0,17	0,43	-0,67	1,00
Tuberkulos	0,13	0,39	-0,63	0,88

3.2 Bootstrapmetoden

För att estimeras mörkertal i sjukanmätningar genom att tillämpa bootstrapmetoden

simuleras slumpstal från $Multinomial(\tilde{n}, \tilde{p}_{12}, \tilde{p}_{21}, \tilde{p}_{11})$

$$\text{där } \tilde{p}_{12} = \frac{n_{12}}{\tilde{n}}, \quad \tilde{p}_{21} = \frac{n_{21}}{\tilde{n}}, \quad \tilde{p}_{11} = \frac{n_{11}}{\tilde{n}}.$$

För att få god precision i resultaten har 10000 oberoende slumpstal genererats för beräkningar i denna uppsats.

I nästa steg beräknas mörkertal enligt (1) för varje stickprov. Från detta urval kan man sedan skatta medelvärde, varians och konfidensintervall.

Tabell 4: mörkertal (\hat{n}_{22}) beräknat enligt bootstrapmetoden på *nationsnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	1,68	0,25	1,22	2,18
Harpest	0,26	0,19	0,00	0,69
Hepatit A	0,00	0,00	0,00	0,00
Meningokockinfektion	0,00	0,00	0,00	0,00
PRP	2,28	0,71	1,04	3,84
Salmonellainfektion	0,69	0,15	0,42	1,00
Sorkfeber	0,49	0,18	0,20	0,88
Tuberkulos	0,30	0,31	0,00	1,01

Tabell 5: mörkertal (\hat{n}_{22}) beräknat enligt bootstrapmetoden på *landstingsnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	1,51	0,58	0,91	2,25
Harpest	0,33	0,95	0,00	4,00
Hepatit A	0,00	0,00	0,00	0,00
Meningokockinfektion	0,00	0,00	0,00	0,00
PRP	3,14	1,37	0,52	7,56
Salmonellainfektion	0,94	0,51	0,48	1,52
Sorkfeber	0,17	0,35	0,01	0,46
Tuberkulos	0,13	0,45	0,00	0,69

3.3 Bayesiansk analys

För att uppskatta mörkertalet enligt bayesianska metoder har 10000 oberoende slumpstal genererats från $Dirichlet(n_{11}+1, n_{12}+1, n_{21}+1)$

Sedan räknas mörkertalet för varje set av slumpstal enligt (1). Från denna fördelning skattas medelvärde och variansen för mörkertal.

Tabell 6: mörkertal (\hat{n}_{22}) beräknat enligt bayesianska metoder på *nationsnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	1,71	0,36	1,07	2,49
Harpest	0,38	0,33	0,00	1,23
Hepatit A	0,08	0,14	0,00	0,46
Meningokockinfektion	0,04	0,10	0,00	0,29
PRP	2,47	1,04	0,82	4,88
Salmonellainfektion	0,72	0,21	0,37	1,19
Sorkfeber	0,54	0,27	0,15	1,20
Tuberkulos	0,61	0,62	0,00	2,18

Tabell 7: mörkertal (\hat{n}_{22}) beräknat enligt bayesianska metoder på *landstingnivå*

Diagnos	\hat{n}_{22}	Std	95 % konfidensintervall	
Campylobakter	2,25	0,78	1,25	3,62
Harpest	3,00	1,50	0,29	8,79
Hepatit A	0,65	0,95	0,00	2,74
Meningokockinfektion	0,64	0,96	0,00	2,78
PRP	8,36	2,09	2,19	18,55
Salmonellainfektion	1,58	0,73	0,75	2,77
Sorkfeber	1,30	1,14	0,14	4,68
Tuberkulos	6,21	1,99	1,41	16,32

4 Resultat

Eftersom man ville studera hur antal sjukdomsfall som varken rapporteras av laboratorieläkare eller av klinik skiljer sig mellan de olika sjukdomar presenteras resultatet i form av andel rapporterade sjukdomsfall som är definierat som antal rapporterade fall dividerat med det uppskattade totalantalet sjukdomsfall. Resultatet presenteras i Tabell 8 nedan. För beräkningarna användes mörkertal estimerat på nationsnivå enligt bayesianska metoder.

Enligt Tabell 8 har campylobakter, salmonellainfektion och tuberkulos högst andel rapporterade sjukdomsfall (99.91 %, 99.94 % respektive 99.88 %). Även hepatit A, sorkfeber och meningokockinfektion har genomsnittlig rapportering som ligger över 99 %. Sjukdomen som har lägst andel rapporterade fall är harpest med 97.92 %.

Tabell 8 andel rapporterade sjukdomsfall

Diagnos	Totalt rapporterade sjukdomsfall	Uppskattat mörkertal	Uppskattat totalantal	Andel rapporterade sjukdomsfall
Campylobakter				
2008	7692	1,708	7693,708	99,98%
2007	7106	1,514	7107,514	99,98%
2003	7147	11,939	7158,939	99,83%
2002	7134	9,972	7143,972	99,86%
Totalt	29079	25,132	29104,132	99,91%
Harpest				
2008	382	0,384	382,384	99,90%
2007	174	0,657	174,657	99,62%
2003	698	28,223	726,223	96,11%
2002	160	0,818	160,818	99,49%
Totalt	1414	30,083	1444,083	97,92%
Hepatit A				
2008	78	0,079	78,079	99,90%
2007	68	0,121	68,121	99,82%
2003	123	0,703	123,703	99,43%
2002	75	0,521	75,521	99,31%
Totalt	344	1,424	345,424	99,59%
Meningokockinfektion				
2008	49	0,041	49,041	99,92%
2007	49	0,128	49,128	99,74%
2003	56	0,236	56,236	99,58%
2002	47	0,454	47,454	99,04%
Totalt	201	0,858	201,858	99,57%
PRP				
2008	564	2,468	566,468	99,56%
2007	668	3,948	671,948	99,41%
2003	561	13,206	574,206	97,70%
2002	524	8,080	532,080	98,48%
Totalt	2317	27,702	2344,702	98,82%
Salmonellainfektion				
2008	4180	0,716	4180,716	99,98%
2007	3931	1,053	3932,053	99,97%
2003	3793	1,217	3794,217	99,97%
2002	3886	6,870	3892,870	99,82%
Totalt	15790	9,856	15799,856	99,94%
Sorkfeber				
2008	569	0,543	569,543	99,90%
2007	2194	1,289	2195,289	99,94%
2003	179	5,151	184,151	97,20%
2002	262	1,494	263,494	99,43%
Totalt	3204	8,478	3212,478	99,74%
Tuberkulos				
2008	588	0,611	588,611	99,90%
2007	507	0,718	507,718	99,86%
Totalt	1095	1,33	1096,329	99,88%

För att enklare se hur andel av mörkertal har förändrats genom åren har histogram använts. Staplarna står för andel icke-rapporterade sjukdomsfall (= 1- andel rapporterade sjukdomsfall).

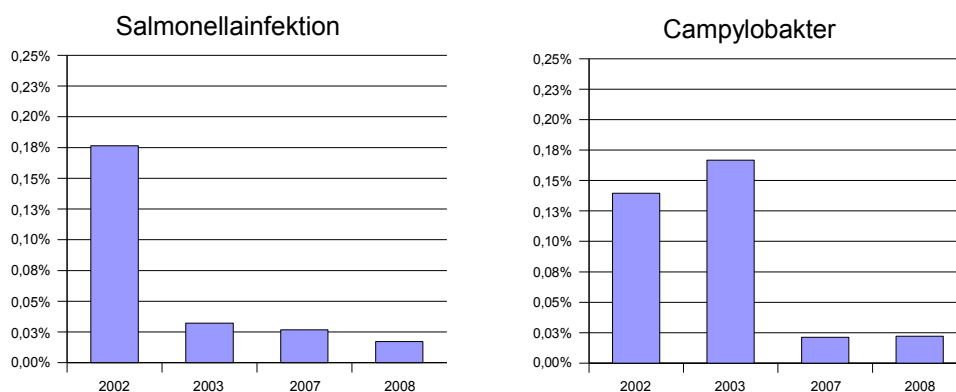
Från diagrammen kan man konstatera att antal icke-rapporterade fall har minskat med åren och 2008 ligger den under 0.5 % för samtliga åtta diagnoser. När man jämför åren 2007-2008 med 2002-2003 ser man en tydlig minskning av andel mörkertal.

Förbättringen i rapporteringen kan förklaras av det internet-baserade rapporteringssystemet som introducerades 2004 och som innebar enklare rapportering av sjukdomsfall.

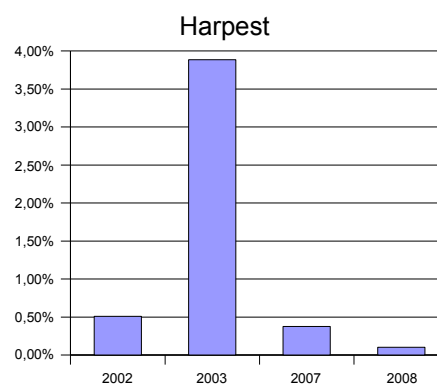
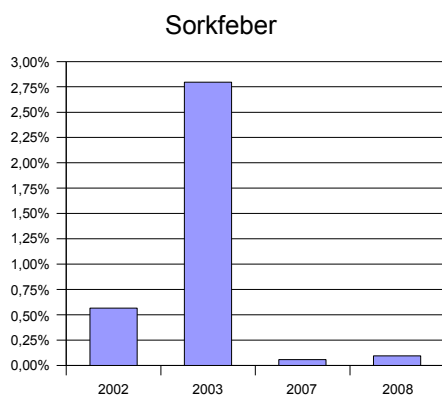
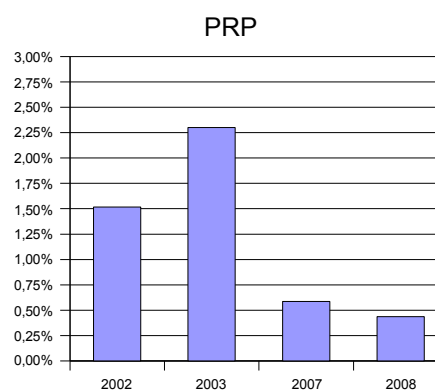
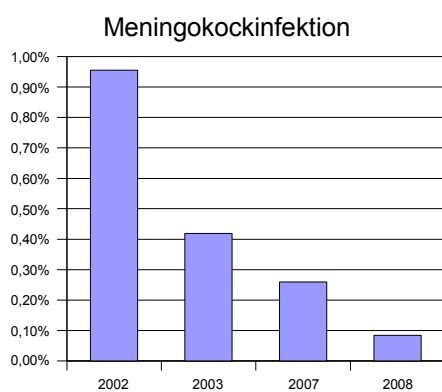
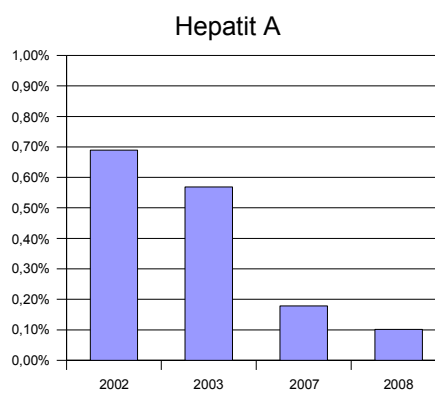
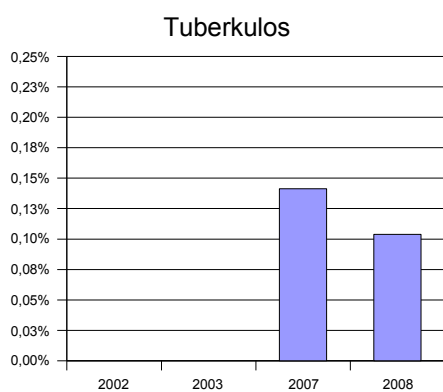
Skillnader mellan sjukdomarna kan förklaras med att sjukdomar som är ”välkända” och som har varit anmälningspliktiga en längre tid har i genomsnitt också lägre andel av fall som man missar. Till exempel har campylobakterinfektion och salmonellainfektion konstant mycket låga andel mörkertal och det är just dessa två sjukdomar som diagnostiseras oftast (i genomsnitt 7270 respektive 3948 sjukdomsfall per år).

Figur 1: Andel icke-rapporterade sjukdomsfall för samtliga åtta diagnoser för åren 2002, 2003, 2007, 2008.

Observera att y-axeln ej visar samma skala i diagrammen nedan.



Figur 1 fortsättning



5 Diskussion: argument kring de olika metoderna

För att kunna tillämpa standard fångst-återfångstmetoden på mörkertalestimeringen i sjukankmälningar måste man anta att sannolikheten att ett fall rapporteras är p_1 för alla läkare och p_2 för alla laboratorier i Sverige samt att händelserna att ett fall rapporteras av en läkare och att fallet rapporteras av ett laboratorium är oberoende.

För att verifiera antagandet om konstanta anmälningssannolikheter utfördes beräkningarna även på landstingnivå. Skillnaden mellan skattningarna gjorda på nationsnivå respektive landstingnivå skiljer sig inte så mycket åt för fångst-återfångstmetoden och bootstrap. Det kan ses som en indikation på att anmälningssannolikheterna är ganska stabila över hela landet.

När det gäller oberoendeantagandet så är det omöjligt att verifiera detta med endast två datakällor; det skulle behövas en tredje källa för att kunna bevisa detta.

Om man betraktar data som en 2x2 kontingenstabell är rad- och kolumnsummorna vid standard fångst-återfångstmetoden fixa och fördelningen av data kan beskrivas av en hypergeometrisk fördelning. Fixa rad- och kolumnsummor skulle i det här fallet innebära fixa antal laboratorie- och klinikanmälningar som ju inte överensstämmer med verkligheten. När man däremot tillämpar bootstrapmetoden är $\tilde{n} = n_{11} + n_{12} + n_{21}$ fixt; det vill säga antal rapporterade fall betraktas som fixt och man simulerar från den empiriska sannolikhetsfördelningen. Celltalen följer då en trinomialfördelning där parametrarna står för sannolikheterna att ett fall blir rapporterad antingen endast av klinik, endast av laboratorium eller av båda.

Skillnaderna mellan de två modellerna yttrar sig även i stora skillnader i variansskattningarna. Vid tillämpning av bootstrapmetoden finns slumpen i hur totalantalet sjukdomsfall fördelas mellan cellerna, givet en sannolikhet för varje cell. Vid standard fångst-återfångstmetoden där det enligt modellen är det rad- och kolumnsummorna som är fixa blir det ju också större variationer i fördelningen av rapporterade sjukdomsfall.

För att kunna beräkna konfidensintervall antar man att \hat{N} är approximativt normalfördelat i fångst-återfångstmetoden. Men som man ser av resultatet är det en dålig approximation eftersom det händer ganska ofta att 2.5 % kvantilen blir negativ som skulle kunna tolkas som att totalantalet rapporterade fall beräknas vara lägre än det egentligen är. Därför är bootstrapmetoden bättre vid konfidensintervall beräkningar.

För skattningar av mindre vanliga sjukdomar och även för skattningar på landstingnivå är små datamängder ett problem. Enligt formel (1) skattas mörkertalet till noll om en av n_{12} och n_{21} är lika med noll. När man studerar resultatet för bayesianska metoder kan man tro att problemet är löst; eftersom man får en skattning större än noll för varje diagnos. Dock bör man vara skeptisk mot resultatet. Bayesiansk analys grundar sig på att nya data justeras med hjälp av kännedom av a-priori-fördelningen. Eftersom denna kan vara osäker kan resultatet ibland ifrågasättas. I fallet med en likformig a-priori-fördelning innebär detta att sannolikheten för ett fall att rapporteras av både laboratorium och klinik överskattas medan sannolikheterna för rapportering av laboratorium men inte av klinik och sannolikheten för rapporteringen av klinik men inte av laboratorium underskattas. Detta resulterar i överskattning av mörkertalet. Överskattningen blir större ju fler landsting man summerar och därför bör man vara extra kritisk mot skattningarna gjorda på denna nivå. Övervärderingen är mindre för skattningarna gjorda på nationsnivå.

Bilaga 1

Avsnittet utgörs av sjukdomsinformation av de diagnoser som tas upp i rapporten. Informationen om dessa sjukdomar kommer från SMI:s hemsida www.smi.se.

Campylobakterinfektion'

Campylobakterinfektion är en zoonos, det vill säga en sjukdom som kan överföras mellan djur och människor. Sjukdomen sprids huvudsakligen via förorenade livsmedel och vanliga symptom är diarré, magsmärtor, illamående, kräkningar och feber. För att förebygga smitta bör man undvika dricka opastöriserad mjölk och även vatten ur bäckar och älvar. Livsmedel som kan tänkas innehålla campylobakter (till exempel kyckling, får- och fläskkött) bör tillagas så att köttet uppnår minst +65°C.

Harpest (tularemi)

Drabbar oftast olika gnagare men kan även överföras till människan genom direkt beröring med smittat djur, bett av smittad mygga eller fästing, intag av smittat vatten eller inandning av förorenat damm. Vanliga symptom är hög feber, huvudvärk, illamående. Smittas man via luftvägarna insjuknar man i lunginflammation.

Hepatit A

Virus som orsakar inflammation i lever sprids via avloppsförorenat vatten och kan anrikas i till exempel musslor och ostron. Smittan kan spridas direkt från person till person vid mycket nära kontakt inom till exempel samma hushåll eller på daghem men smittan kan också spridas sexuellt. Vanliga symptom är feber, illamående, kräkningar och gulhet i huden.

Meningokockinfektion (epidemisk hjärnhinneinflammation)

Meningokocken är en bakterie som infekterar enbart människan. Bakterien sprids som en luftburen droppsmitta genom nära kontakt. Trots att det är många som utsätts för smittan är det vanligt att man inte insjuknar. Omkring 10 % av frisk befolkning bär på

bakterien. Sjukdomen kännetecknas av svår huvudvärk, feber, kräkningar och nackstyvhet. Patienten är ofta omtöcknad och förvirrad, sjukdomen kan i svåra fall leda till medvetslöshet. Den allvarligaste formen kännetecknas av blodförgiftning med hudblödningar och sjunkande blodtryck.

PRP

Pneumokocker är den vanligaste orsaken till bakteriella luftvägsinfektioner såsom lunginflammation, öroninflammation och bihåleinflammation. Bakterien är också en av de vanligaste orsakerna till hjärnhinneinflammation och blodförgiftning (sepsis). PRP är infektion/bärarskap med pneumokocker med nedsatt känslighet för penicillin G. Den viktigaste åtgärden för att minska spridningen av PRP är att undvika onödig antibiotikaanvändning.

Salmonellainfektion

Den så kallade zoonotiska salmonellabakterien är vanligt förekommande hos t.ex. nötkreatur, gris, hönsfågel, vilda fåglar och sällskapsdjur som hund, katt, orm och sköldpadda. För att förebygga smitta rekommenderas allmänt god hygien. Livsmedel som kan tänkas innehålla bakterien bör lagas på hög temperatur.

Sorkfeber

Sjukdomen sprids via skogssork genom att utsöndra virus i saliv, avföring och urin. Smittspridning från skogssork till människa sker framför allt via inandning av damm som förorenats av sorkars urin och avföring eller via direktkontakt med saliv, urin eller avföring. Vanliga symptom är hög feber, muskelvärk, allmän sjukdomskänsla och ofta svåra buk- och ryggsmärtor. Njurarna angrips och nedsatt njurfunktion kan ses under 1–2 veckor. Ibland kan hudblödningar uppträda.

Tuberkulos

Tuberkulos är en av de mest spridda infektionssjukdomarna i världen och man räknar med att en tredjedel av jordens befolkning är bärare av tuberkulosbakterien. Varje år

insjuknar minst nio miljoner människor i aktiv tuberkulos och c:a två miljoner dör. Flest sjukdomsfall och dödsfall inträffar i Afrika och Asien.

Tuberkulos är en luftburen smitta som sprids genom upphostningarna från en patient med tuberkulos i främst lungorna men ibland även i struphuvudet. Bakterien växer mycket långsamt men är också mycket långlivade och kan under gynnsamma omständigheter överleva i årtal. I människokroppen kan den förekomma latent under personens hela livstid.

Vanliga symptom vid tuberkulos är feber, nattsvett, avmagring och trötthet. För lungtuberkulos är långvarig hosta mer än tre veckor typisk med eller utan upphostningar av slem och ibland blod.

Bilaga 2

Campylobakterinfektion

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
2008 (7692)					
Fångst-återfångst	Nation	1,68	1,32	-0,91	4,26
	Landsting	1,51	1,27	-0,99	4,00
Bootstrap	Nation	1,68	0,25	1,22	2,18
	Landsting	1,51	0,58	0,91	2,25
Dirichlet	Nation	1,71	0,36	1,07	2,49
	Landsting	2,25	0,78	1,25	3,62
2007 (7106)					
Fångst-återfångst	Nation	1,48	1,24	-0,94	3,91
	Landsting	1,56	1,28	-0,95	4,07
Bootstrap	Nation	1,48	0,22	1,07	1,95
	Landsting	1,56	0,54	1,02	2,16
Dirichlet	Nation	1,51	0,32	0,94	2,22
	Landsting	2,30	0,76	1,37	3,59
2003 (7147)					
Fångst-återfångst	Nation	11,87	3,61	4,79	18,96
	Landsting	12,52	3,76	5,14	19,90
Bootstrap	Nation	11,87	1,07	9,84	14,02
	Landsting	12,52	1,14	10,01	15,16
Dirichlet	Nation	11,94	1,55	9,11	15,17
	Landsting	14,33	1,44	10,59	18,76
2002 (7134)					
Fångst-återfångst	Nation	9,87	3,29	3,41	16,33
	Landsting	8,36	3,07	2,34	14,38
Bootstrap	Nation	9,87	0,99	8,00	11,92
	Landsting	8,36	1,02	6,42	10,51
Dirichlet	Nation	9,97	1,42	7,44	12,96
	Landsting	9,86	1,28	6,90	13,31

Mörkertal (\hat{n}_{22}), standard avvikelser och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt standard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Harpest

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
<i>2008 (382)</i>					
Fångst-återfångst	Nation	0,26	0,54	-0,80	1,31
	Landsting	0,33	0,77	-1,18	1,84
Bootstrap	Nation	0,26	0,19	0,00	0,69
	Landsting	0,33	0,95	0,00	4,00
Dirichlet	Nation	0,38	0,33	0,00	1,23
	Landsting	3,00	1,50	0,29	8,79
<i>2007 (174)</i>					
Fångst-återfångst	Nation	0,42	0,72	-0,99	1,83
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,42	0,32	0,00	1,17
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,66	0,58	0,00	2,13
	Landsting	2,80	1,55	0,05	9,04
<i>2003 (698)</i>					
Fångst-återfångst	Nation	27,63	6,99	13,92	41,33
	Landsting	46,68	12,44	22,30	71,06
Bootstrap	Nation	27,63	4,52	19,48	37,17
	Landsting	46,68	3,07	28,63	66,09
Dirichlet	Nation	28,22	6,47	17,22	42,34
	Landsting	50,70	3,77	27,23	83,22
<i>2002 (160)</i>					
Fångst-återfångst	Nation	0,60	0,85	-1,08	2,27
	Landsting	0,77	1,09	-1,36	2,90
Bootstrap	Nation	0,60	0,38	0,00	1,50
	Landsting	0,77	0,98	0,00	3,00
Dirichlet	Nation	0,82	0,64	0,00	2,50
	Landsting	4,01	1,81	0,39	11,82

Mörkertal (\hat{n}_{22}), standard avvikelser och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt standard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Hepatit A

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
2008 (78)					
Fångst-återfångst	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,08	0,14	0,00	0,46
	Landsting	0,65	0,95	0,00	2,74
2007 (68)					
Fångst-återfångst	Nation	0,05	0,22	-0,39	0,48
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,05	0,06	0,00	0,20
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,12	0,17	0,00	0,63
	Landsting	0,76	1,01	0,00	3,50
2003 (123)					
Fångst-återfångst	Nation	0,56	0,80	-1,02	2,14
	Landsting	0,27	0,56	-0,84	1,38
Bootstrap	Nation	0,56	0,30	0,13	1,30
	Landsting	0,27	0,58	0,32	1,61
Dirichlet	Nation	0,70	0,52	0,09	2,05
	Landsting	2,07	1,18	0,45	5,52
2002 (75)					
Fångst-återfångst	Nation	0,32	0,62	-0,89	1,52
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,32	0,26	0,00	0,98
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,52	0,53	0,00	1,86
	Landsting	1,82	1,51	0,00	6,50

Mörkertal (\hat{n}_{22}), standard avvikelser och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt standard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Meningokockinfektion

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
2008 (49)					
Fångst-återfångst	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,04	0,10	0,00	0,29
	Landsting	0,64	0,96	0,00	2,78
2007 (49)					
Fångst-återfångst	Nation	0,04	0,22	-0,38	0,47
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,04	0,06	0,00	0,21
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,13	0,20	0,00	0,64
	Landsting	0,87	1,11	0,00	4,00
2003 (56)					
Fångst-återfångst	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,00	0,00	0,00	0,00
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,24	0,39	0,00	1,30
	Landsting	1,65	1,39	0,00	6,83
2002 (47)					
Fångst-återfångst	Nation	0,21	0,51	-0,79	1,21
	Landsting	0,00	0,00	0,00	0,00
Bootstrap	Nation	0,21	0,25	0,00	0,88
	Landsting	0,00	0,00	0,00	0,00
Dirichlet	Nation	0,45	0,57	0,00	2,00
	Landsting	1,27	1,40	0,00	6,00

Mörkertal (\hat{n}_{22}), standard avvikelsen och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt tandard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Metod	Nivå	PRP \hat{n}_{22}	Std	95 % konfidensintervall	
2008 (564)					
Fångst-återfångst	Nation	2,28	1,67	-0,99	5,54
	Landsting	3,14	2,53	-1,81	8,08
Bootstrap	Nation	2,28	0,71	1,04	3,84
	Landsting	3,14	1,37	0,52	7,56
Dirichlet	Nation	2,47	1,04	0,82	4,88
	Landsting	8,36	2,09	2,19	18,55
2007 (668)					
Fångst-återfångst	Nation	3,74	2,14	-0,47	7,94
	Landsting	3,36	2,22	-0,99	7,71
Bootstrap	Nation	3,74	0,93	2,11	5,76
	Landsting	3,14	1,27	1,25	7,14
Dirichlet	Nation	3,95	1,37	1,75	7,06
	Landsting	7,98	2,00	2,58	18,26
2003 (561)					
Fångst-återfångst	Nation	12,82	4,29	4,41	21,24
	Landsting	9,96	3,85	2,42	17,50
Bootstrap	Nation	12,82	2,34	8,63	17,84
	Landsting	9,96	1,53	6,27	15,31
Dirichlet	Nation	13,21	3,40	7,57	20,81
	Landsting	18,86	2,63	9,07	34,51
2002 (524)					
Fångst-återfångst	Nation	7,76	3,27	1,35	14,17
	Landsting	7,19	3,71	-0,08	14,46
Bootstrap	Nation	7,76	1,72	4,68	11,43
	Landsting	7,19	1,64	3,37	13,48
Dirichlet	Nation	8,08	2,47	3,97	13,73
	Landsting	13,96	2,38	5,71	27,57

Mörkertal (\hat{n}_{22}), standard avvikelser och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt tandard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Salmonellainfektion

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
2008 (4180)					
Fångst-återfångst	Nation	0,69	0,84	-0,96	2,34
	Landsting	0,94	1,01	-1,03	2,91
Bootstrap	Nation	0,69	0,15	0,42	1,00
	Landsting	0,94	0,51	0,48	1,52
Dirichlet	Nation	0,72	0,21	0,37	1,19
	Landsting	1,58	0,73	0,75	2,77
2007 (3931)					
Fångst-återfångst	Nation	1,02	1,03	-1,00	3,03
	Landsting	0,74	0,89	-0,99	2,48
Bootstrap	Nation	1,02	0,20	0,65	1,44
	Landsting	0,74	0,58	0,66	1,98
Dirichlet	Nation	1,05	0,29	0,57	1,70
	Landsting	2,14	0,86	1,03	3,85
2003 (3793)					
Fångst-återfångst	Nation	1,17	1,11	-1,00	3,34
	Landsting	1,31	1,19	-1,02	3,65
Bootstrap	Nation	1,17	0,24	0,73	1,68
	Landsting	1,31	0,59	0,72	2,09
Dirichlet	Nation	1,22	0,35	0,63	1,99
	Landsting	2,37	0,87	1,20	4,08
2002 (3886)					
Fångst-återfångst	Nation	6,79	2,72	1,45	12,13
	Landsting	11,35	3,91	3,69	19,02
Bootstrap	Nation	6,79	0,79	5,32	8,41
	Landsting	11,35	1,40	7,84	15,53
Dirichlet	Nation	6,87	1,13	4,88	9,29
	Landsting	12,52	1,70	7,48	18,92

Mörkertal (\hat{n}_{22}), standard avvikelsen och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt tandard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Sorkfeber

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
2008 (569)					
Fångst-återfångst	Nation	0,49	0,72	-0,92	1,89
	Landsting	0,17	0,43	-0,67	1,00
Bootstrap	Nation	0,49	0,18	0,20	0,88
	Landsting	0,17	0,35	0,01	0,46
Dirichlet	Nation	0,54	0,27	0,15	1,20
	Landsting	1,30	1,14	0,14	4,68
2007 (2194)					
Fångst-återfångst	Nation	1,24	1,14	-1,00	3,48
	Landsting	1,06	1,08	-1,06	3,17
Bootstrap	Nation	1,24	0,25	0,81	1,78
	Landsting	1,06	0,59	0,51	1,88
Dirichlet	Nation	1,29	0,36	0,70	2,11
	Landsting	2,53	1,20	0,80	6,18
2003 (179)					
Fångst-återfångst	Nation	4,78	2,61	-0,34	9,89
	Landsting	10,32	6,25	-1,92	22,56
Bootstrap	Nation	4,78	1,40	2,50	7,88
	Landsting	10,32	2,76	3,52	31,59
Dirichlet	Nation	5,15	2,15	2,01	10,27
	Landsting	15,53	3,23	3,44	39,85
2002 (262)					
Fångst-återfångst	Nation	1,34	1,25	-1,10	3,78
	Landsting	2,30	1,93	-1,47	6,08
Bootstrap	Nation	1,34	0,47	0,60	2,41
	Landsting	2,30	1,40	0,54	9,27
Dirichlet	Nation	1,49	0,73	0,46	3,22
	Landsting	5,27	1,88	0,80	14,18

Mörkertal (\hat{n}_{22}), standard avvikelsen och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt standard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

Tuberkulos

Metod	Nivå	\hat{n}_{22}	Std	95 % konfidensintervall	
<i>2008 (588)</i>					
Fångst-återfångst	Nation	0,30	0,63	-0,93	1,54
	Landsting	0,13	0,39	-0,63	0,88
Bootstrap	Nation	0,30	0,31	0,00	1,01
	Landsting	0,13	0,45	0,00	0,69
Dirichlet	Nation	0,61	0,62	0,00	2,18
	Landsting	6,21	1,99	1,41	16,32
<i>2007 (507)</i>					
Fångst-återfångst	Nation	0,36	0,70	-1,01	1,72
	Landsting	0,27	0,60	-0,90	1,44
Bootstrap	Nation	0,36	0,36	0,00	1,21
	Landsting	0,27	0,57	0,00	1,13
Dirichlet	Nation	0,72	0,73	0,00	2,59
	Landsting	8,63	2,29	2,04	21,59

Mörkertal (\hat{n}_{22}), standard avvikelsen och 95 % konfidensintervall uträknat på nations- och landstingnivå enligt standard fångst-återfångstmetoden, bootstrapmetoden och bayesianska principer (Dirichlet).

6 Källförteckning

- [1] Alan Agresti (2002) *Categorical Data Analysis*, John Wiley And Sons Ltd
- [2] Andrew Gelman, John B. Carlin, Hal S. Stern & Donald B. Rubin (2003)
Bayesian Data Analysis, Chapman & Hall/CRC
- [3] Bradley Efron, R J Tibishirani (1994) *Introduction To The Bootstrap*, Taylor & Francis Ltd
- [4] A. C Davison, & D. V. Hinkley (1997) *Bootstrap Methods and their Application*, CambridgeUniversity Press.
- [5] Bernard W. Lindgren (1993) *Statistical Theory* , Chapman & Hall /CRC
- [6] Anna Jansson, M. Arneborn, K. Ekdahl (2004) *Sensitivity of the Swedish surveillance system for communicable diseases 1998-2002, assessed by the capture-recapture method*