

Populatsiooni suuruse hindamine

Korduspüügi meetod

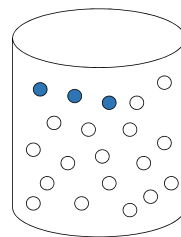
Korduspüügi meetod (Capture-Recapture)

Esimene teadaolev kasutaja oli Laplace, tuntakse ka kui Petersoni (1896, kalade arvukuse hindamine) meetod või Lincolni indeks (1930-ndad, haneliste arvukus).



Korduspüügi meetod (Capture-Recapture)

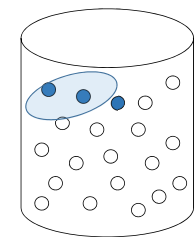
Esimene teadaolev kasutaja oli Laplace, tuntakse ka kui Petersoni (1896, kalade arvukuse hindamine) meetod või Lincolni indeks (1930-ndad, haneliste arvukus).



N – populatsiooni suurus
 M – märgiga isendite arv

Korduspüügi meetod (Capture-Recapture)

Esimene teadaolev kasutaja oli Laplace, tuntakse ka kui Petersoni (1896, kalade arvukuse hindamine) meetod või Lincolni indeks (1930-ndad, haneliste arvukus).



N – populatsiooni suurus
 M – märgiga isendite arv
 n – korduspüügi-valimi suurus
 m – korduspüügi-valimis märgistatud isendeid

$$\frac{m}{n} \approx \frac{M}{N}$$

$$N \approx \frac{M}{m} n$$

n – inimeste arv põhjalikumalt uuritud provintsisides
vastsündinute arv (M)
 $N \approx \frac{1\,000\,000}{71\,866} \cdot 2\,037\,615$
 m - vastsündinute arv põhjalikumalt uuritud provintsisides

Korduspüügi meetod

(Capture-Recapture)

Esimene teadaolev kasutaja oli Laplace, tuntakse ka kui Petersoni (1896, kalade arvukuse hindamine) meetod või Lincolni indeks (1930-ndad, haneliste arvukus).

$$\frac{m}{n} \approx \frac{M}{N}$$

$$N \approx \frac{M}{m} n$$

n - inimeste arv põhjalikumalt uuritud provintssides
 M - vastsündinute arv (Prantsusmaa elanike arv 1802. aastal)
 m - vastsündinute arv põhjalikumalt uuritud provintssides

$$N \approx \frac{1\,000\,000}{71\,866} \cdot 2\,037\,615 \approx 28\,328\,612$$

Teine valim ei pea tegelikult olema juhuslik. Piisab sellestki, kui tõenäosus 2. valimisse sattuda ei sõltu sellest, kas sa olid 1. valimis.

Nimelt juhuslikult valitud indiviidi korral kehtib (valimite sõltumatuse korral):

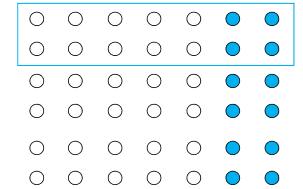
$$P(\text{valim } 1 \cap \text{valim } 2) = P(\text{valim } 1) P(\text{valim } 2)$$

$$\frac{m}{N} = \frac{M}{N} \cdot \frac{n}{N}$$

$$N = \frac{M}{m} n$$

$$N = \frac{12}{4} \cdot 14 = 42$$

teine valim



Teine valim ei pea tegelikult olema juhuslik. Piisab sellestki, kui tõenäosus 2. valimisse sattuda ei sõltu sellest, kas sa olid 1. valimis.

Nimelt juhuslikult valitud indiviidi korral kehtib (valimite sõltumatuse korral):

$$P(\text{valim } 1 \cap \text{valim } 2) = P(\text{valim } 1) P(\text{valim } 2)$$

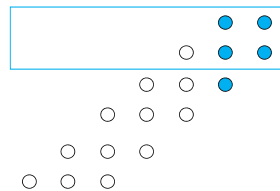
$$\frac{m}{N} = \frac{M}{N} \cdot \frac{n}{N}$$

$$N = \frac{M}{m} n$$

~~$$N = \frac{5}{4} \cdot 8 = 6,25$$~~

Sõltuvad valimid – meetod ei tööta!

teine valim



Ärkveloleku kontroll

Tahetakse teada, kui palju insulite toimus Lõuna-Londonis aastatel 1995-1996.

Paluti teavitada insuldihaigetest nii haiglaid kui perearste. Tulemused:

Haiglate poolt raporteeritud haigusjuhud:	169;
perearstide poolt raporteeritud haigusjuhte:	80;
neist kattuvaid haigusjuhte	11.

Andmed pärit: *Estimation of the incidence of stroke using a capture-recapture model including covariates*, K.Tilling, J. AC Sterne, C. DA Wolfe, *International Journal of Epidemiology* 2001; 30:1351-1359.

Märkus: Vaadeldud artiklis oli informatsiooni haigusjuhtude kohta kogutud enam kui kahest allikast.

$$\text{Insuldijuhtude arv} \approx \frac{M}{m} n = \frac{169}{11} \cdot 80 = 1229$$

Hinnangu täpsusest

$$\hat{N} = \frac{M}{m}n$$

Miks (Petersoni) hinnang on juhuslik?

Suurused M ja n on fikseeritud uuringu korraldajate poolt – vaid m sõltub juhusest. Milline on juhusliku suuruse m jaotus?

$$m \sim B(n, p) \quad p = \frac{M}{N} \quad N = \frac{M}{p}$$

$$\hat{p} = \frac{m}{n} \quad \hat{p} \pm 1,96 \cdot \sqrt{\hat{p}(1 - \hat{p})/n}$$

DeLury usaldusintervall

$$\frac{M}{\hat{p} \pm 1,96 \cdot \sqrt{\hat{p}(1 - \hat{p})/n}}$$
$$\frac{\frac{nM}{m + 1,96\sqrt{m(1 - m/n)}} \dots \frac{nM}{m - 1,96\sqrt{m(1 - m/n)}}}{M=200}$$

Hinnangu täpsusest

$$\hat{N} = \frac{M}{m}n$$

Miks (Petersoni) hinnang on juhuslik?

Suurused M ja n on fikseeritud uuringu korraldajate poolt – vaid m sõltub juhusest. Milline on juhusliku suuruse m jaotus?

$$m \sim B(n, p) \quad p = \frac{M}{N} \quad N = \frac{M}{p}$$

$$\hat{p} = \frac{m}{n} \quad > \text{binom.test}(10, 100)$$

95 percent confidence interval:
0.04900469 0.17622260

$$\frac{M}{p} = \frac{200}{0.176} = 1134,9 \quad \dots \quad \frac{M}{p} = \frac{200}{0.049} = 4081,6 \quad M=200$$

Hinnangu täpsusest

$$\hat{N} = \frac{M}{m}n$$

Miks (Petersoni) hinnang on juhuslik?

Suurused M ja n on fikseeritud uuringu korraldajate poolt – vaid m sõltub juhusest. Milline on juhusliku suuruse m jaotus?

$$m \sim B(n, p) \quad p = \frac{M}{N} \quad N = \frac{M}{p}$$

$$\hat{p} = \frac{m}{n} \quad > \text{library(exactci)}$$
$$> \text{binom.exact}(10, 100,$$

tsmethod="minlike")

95 percent confidence interval:
0.0534 0.1740

$$\frac{M}{p} = \frac{200}{0.174} = 1136 \quad \dots \quad \frac{M}{p} = \frac{200}{0.0534} = 3745 \quad M=200$$

Hinnangu täpsusest

$$\hat{N} = \frac{M}{m}n$$

M=200, n=100, m=10

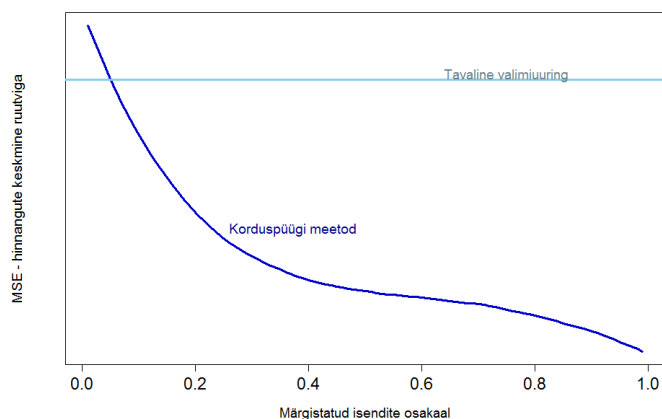
$$\hat{N} = 2000$$

Ligikaudne (DeLury): 1259...4854
binom.test: 1134...4082
binom.exact: 1136...3745

Kuna kasutada? Üks näide

Kui teaksime teatud tüüpi isendite arvu – näiteks narkomaanide arvu Eestis, siis saaksime hinnata ka narkomaanide osakaalu Eestis.

Kas mõistlikum oleks võtta juhuslik valim Eestlastest ja uurida narkomaanide osakaalu seal, või tasuks pigem märgistada narkomaane (näiteks „märgiga“ narkomaanid on üledoosi tõttu haiglasse sattunud narkomaanid) ja kasutada siis kordupüügi meetodit?



Petersoni hinnangu nihkest

Hinnangu $\hat{N} = \frac{M}{m}n$ puhul on tegemist nihkega hinnanguga:

$$E(\hat{N} - N) = 0$$

Pahanduste põhjustajaks on võimalus, et me ei suuda loomade teistkordsel püüdmisel kätte saada ühtegi märgistatud isendit.

Sestap on pakutud välja hinnanguid, mis peaksid hinnangu nihet vähendama – näiteks lisame kunstlikult ühe märgiga isendi populatsiooni juurde, kes ka kindlasti teise valimisse sattub. Lõpliku populatsiooni suuruse rehkendamisel võtame selle kunstlikult lisatud isendi populatsiooni arvukusest maha (Chapmani hinnang).

$$\hat{N}_C = \frac{(M+1)}{(m+1)}(n+1) - 1$$

Eeldustest

1. Märgid ei lähe kaduma ja on alati loomal tuvastatavad;
2. populatsioon on kinnine (st loomi ei tule juurde ega lahku);
3. võetavad valimid on teineteisest sõltumatud.

eeldust 2 võib muuta kahel viisil:

2a. Loomad lahkuvad populatsioonist peale märgistamist, märgistatud loomade ja märgistamata loomade kadumiskiirus (suremiskiirus) on aga sama. Loomi populatsiooni ei lisandu. Sellisel juhul ei muutu märgistatud loomade osakaal populatsioonis ja toodud hinnang on mõistlik, kusjuures iseloomustab loomade arvukust märgistamise ajal.

2b. Loomi lisandub populatsiooni märgistamise ja taaspüüdmise aja vahepeal, kuid lahkujaid ei ole. Sellisel juhul tohib toodud hinnangut kasutada. Hinnatakse sellisel juhul loomade arvukust perioodi lõpul ehk taaspüüdmise ajal.

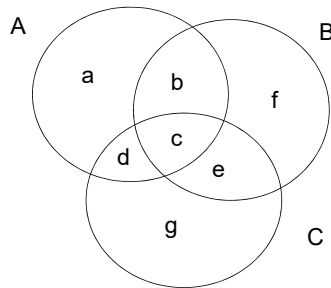
Rohkem kui kaks valimit

Vahel on võimalik kasutada enam kui kahe valimi andmeid. Näiteks võib olla võimalik saada infot patsientide kohta kolmest või enamast allikast (perearstid, haiglad, kindlustus,...); rände ajal rõngastatud loomi on võimalik vaadelda mitmel vaatluspäeval; pikaajalisi linde saab märgistada ja varem märgistatud lindude osakaalu leida igal aastal (näiteks viie aasta jooksul) vms.

Juhul, kui kasutada on enam kui kahe valimi andmeid, saab sageli mõnda loetletud eeldustest pehmedada või siis saab kasutada täiendavaid andmeid tehtud eelduste paikapidavuse kontrollimiseks.

Vaatleme järgnevalt, kuidas saaks (saab) kontrollida valimite sõltumatust kasutades kolmanda valimi andmeid.

Olgu meil kasutada 3 valimit. Soovime kontrollida, et tegemist on sõltumatute valimitega.



Kui valimid A, B ja C on sõltumatud, siis peaksime kasutades *d*, *c* ja *e*-d saama hinnata *g*-d (või valimi C kogusuurust) kasutades tavalist Petersoni hinnangut:

$$c+d+e+g = (d+c)/c \cdot (c+e)$$

Kui saadud hinnang pole hea, võime kahtlustada, et tegemist on sõltuvate valimitega.

Juhul, kui valimid on sõltumatud, siis $c/d=e/g$ ehk $OR = c \cdot g / (d \cdot e) \approx 1$

$OR > 1$, positiivne sõltuvus, *N* alahinnatud;

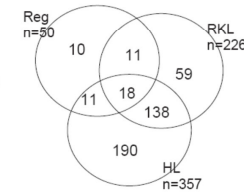
$OR < 1$, negatiivne sõltuvus, *N* ülehinnatud.

Näide

Prantsusmaal peavad arstid alates 1987.a. teavitama igast Legionella juhtumist. Selliseid teateid tuleb umbes 60 haigusjuhu kohta aastas. Sooviti kontrollida, kui efektiivne on loodud süsteem tegelikult haigusjuhtudest teavitamisel.

Andmeid 1995. a. aset leidnud haigusjuhtumite kohta koguti kolmest allikast

- Riiklik register (Reg), kuhu arstid on kohustatud saatma teabe iga esinenud haigusjuhtumi kohta. Teateid haigusjuhtumitest 1995.a. oli kokku $n = 50$;
- Riiklik kesklaboratoorium (RKL), mis ainsana on Prantsusmaal võimeline panema lõplikku (ametlikku) Legionella diagnoosi, $n = 226$;
- Haiglate laboratooriumid (HL), (Legionella diagnoosi puhul peaksid proovi reeglite järgi edasi saatma kesklaboratooriumisse lõpliku kinnituse saamiseks), $n = 357$.



Reg + HL -> $n_{RKL} \quad (11+18)/18 \cdot (138+18) = 251$

Reg + RKL -> $n_{HL} \quad (11+18)/18 \cdot (138+18) = 251$

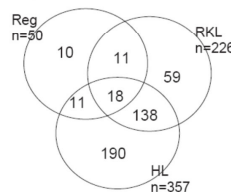
HL + RKL -> $n_{Reg} \quad (11+18)/18 \cdot (11+18) = 47$

Näide

Prantsusmaal peavad arstid alates 1987.a. teavitama igast Legionella juhtumist. Selliseid teateid tuleb umbes 60 haigusjuhu kohta aastas. Sooviti kontrollida, kui efektiivne on loodud süsteem tegelikult haigusjuhtudest teavitamisel.

Andmeid 1995. a. aset leidnud haigusjuhtumite kohta koguti kolmest allikast

- Riiklik register (Reg), kuhu arstid on kohustatud saatma teabe iga esinenud haigusjuhtumi kohta. Teateid haigusjuhtumitest 1995.a. oli kokku $n = 50$;
- Riiklik kesklaboratoorium (RKL), mis ainsana on Prantsusmaal võimeline panema lõplikku (ametlikku) Legionella diagnoosi, $n = 226$;
- Haiglate laboratooriumid (HL), (Legionella diagnoosi puhul peaksid proovi reeglite järgi edasi saatma kesklaboratooriumisse lõpliku kinnituse saamiseks), $n = 357$.



Reg + HL -> $n_{RKL} \quad (11+18)/18 \cdot (138+18) = 251 \quad OR = 59/11 / (138/18) = 0,7 \quad (0,3..1,6)$

Reg + RKL -> $n_{HL} \quad (11+18)/18 \cdot (138+18) = 251 \quad OR = 190/11 / (138/18) = 2,25 \quad (1,0..4,9)$

HL + RKL -> $n_{Reg} \quad (11+18)/18 \cdot (11+18) = 47 \quad OR = 10/11 / (11/18) = 1,5 \quad (0,5..4,6)$

Hinnangud kahe allika põhjal:

Reg – RKL	389
Reg – HL	615
RKL – HL	715

Aga – allikad on positiivselt sõltuvad:

$OR(Reg - RKL; HL) = 2,3 \quad [1,0-5,7]$

$OR(Reg - HL; RKL) = 0,7 \quad [0,3-1,8]$

$OR(HL - RKL; Reg) = 1,5 \quad [0,4-5,4]$

Otsus: kombineerime kaks sõltuvat allikat kokku Reg+RKL:

Reg+RKL – HL $528 \quad [495-561]$ haigusjuhtu

Legionärihaiguse näide võetud:

Infuso et al Eurosurveillance 1998;3(5):48-50;

Korduv püüdmiss-tabamismeetod

(Mark-Recapture; Capture-Recapture)



Lisamaterjal – eksamil ei küsita

Korduv püüdmiss-tabamismeetod

(Mark-Recapture; Capture-Recapture)

Tähistame:

n_{11} – isendite arv, kes esinesid nii esimeses kui ka teises valimis;

n_{01} – isendid, kes esinesid ainult teises valimis;

...

p_1 – tõenäosus sattuda 1. valimisse;

p_2 – tõenäosus sattuda 2. valimisse;

$\log(EY)=...$

$$En_{11} = N p_1 p_2$$

$$En_{10} = N p_1 (1-p_2)$$

$$En_{01} = N (1-p_1) p_2$$

$$En_{00} = N (1-p_1) (1-p_2)$$

Korduv püüdmiss-tabamismeetod

(Mark-Recapture; Capture-Recapture)

Tähistame

$$\beta_0 := \log(N) + \log(1-p_1) + \log(1-p_2);$$

$$\log(E n_{00}) = \beta_0$$

Parameetrid μ , a ja b hinnatavad Poissoni regressiooni abil

$\log(EY)=...$

$$En_{11} = N p_1 p_2$$

$$\rightarrow \log(E n_{11}) = \log(N) + \log(p_1) + \log(p_2)$$

$$En_{10} = N p_1 (1-p_2)$$

$$\rightarrow \log(E n_{10}) = \log(N) + \log(p_1) + \log(1-p_2)$$

$$En_{01} = N (1-p_1) p_2$$

$$\rightarrow \log(E n_{01}) = \log(N) + \log(1-p_1) + \log(p_2)$$

$$En_{00} = N (1-p_1) (1-p_2)$$

$$\rightarrow \log(E n_{00}) = \log(N) + \log(1-p_1) + \log(1-p_2)$$

Antud juhul 3 võrrandit ja 3 tundmatut. Ühe vaatluse korral suurima tõepära hinnang Poissoni jaotuse parameetritele on antud vaatlus ise:

$$\begin{aligned} \log(E n_{00}) &= \beta_0 && \Rightarrow \log(n_{00}) \approx \beta_0 \\ \log(E n_{10}) &= \beta_0 + \beta_1 && \Rightarrow \log(n_{10}) \approx \beta_0 + \beta_1 \\ \log(E n_{01}) &= \beta_0 + \beta_2 && \Rightarrow \log(n_{01}) \approx \beta_0 + \beta_2 \\ \log(E n_{11}) &= \beta_0 + \beta_1 + \beta_2 && (+\beta_{1,2}) \Rightarrow \log(n_{11}) \approx \beta_0 + \beta_2 + \beta_1 \end{aligned}$$

Seega, sõltumatute valimite korral ($\beta_{1,2}=0$):

$$\begin{aligned} \log(n_{10}) + \log(n_{01}) - \log(n_{11}) &\approx (\beta_0 + \beta_1) + (\beta_0 + \beta_2) - (\beta_0 + \beta_1 + \beta_2) && (= \beta_0) \\ &\approx \log(n_{00}) \end{aligned}$$

$$n_{00} \approx \exp(\log(n_{10}) + \log(n_{01}) - \log(n_{11})) = n_{10} n_{01} / n_{11}$$

Hinnang populatsiooni suurusele:

$$N = n_{10} n_{01} / n_{11} + n_{11} + n_{10} + n_{01}$$

Võrdle Petersoni hinnanguga:

$$\begin{aligned} N &= (n_{10} + n_{11})(n_{01} + n_{11}) / n_{11} \\ &= n_{10} n_{01} / n_{11} + n_{11} n_{01} / n_{11} + n_{10} n_{11} / n_{11} + n_{11} n_{11} / n_{11} \\ &= n_{10} n_{01} / n_{11} + n_{01} + n_{10} + n_{11} \end{aligned}$$

Hindamine R-is

```
register valim n
0 1 5
1 1 5
1 0 105
```

Table 1 Prevalence of type 2 diabetes mellitus in Oguni in men and women aged 50–69 method of capture-recapture

	ODR ^a		SS ^b		No. overlapped cases
	No. cases	Prevalence (%)	No. cases	Prevalence (%)	
Men	110	7.1	10	8.2	5
Women	124	7.0	16	9.2	9
Total	234	7.1	26	8.8	14

^a Oguni Diabetes Registry.
^b Sample study.

```
> m1=glm(n~valim+register, family=poisson())
```

```
> summary(m1)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.654e+00	6.399e-01	7.272	3.53e-13 ***
valim	-3.045e+00	4.577e-01	-6.651	2.91e-11 ***
register	4.494e-15	6.325e-01	0.000	1

Hindamine R-is

```
> exp(coef(m1))
```

```
(Intercept)      valim      register
105.00000000    0.04761905    1.00000000
```

Kui palju oli selliseid mees-diabeetikuid, kes pole kirjast ei diabeediregistris ning ei sattunud ka valimisse

```
> 5+5+105+105
```

```
[1] 220
```

```
> confint(m1)
Waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept)  3.360721  5.946963
valim        -4.085251 -2.252321
register     -1.279722  1.279722
```

Hindamine R-is

```
> exp(coef(m1))
```

```
(Intercept)      valim      register
105.00000000    0.04761905    1.00000000
```

Kui palju oli selliseid mees-diabeetikuid, kes pole kirjast ei diabeediregistris ning ei sattunud ka valimisse

```
> 5+5+105+105
```

```
[1] 220
```

```
> sum(n)+exp(confint(m1))
Waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept) 28.80995730 382.5897617
valim        0.01681892  0.1051549
register      0.27811472  3.5956386
```

Hindamine R-is

Table 1 Prevalence of type 2 diabetes mellitus in Oguni in men and women aged 50-69 by the registry, the method of capture-recapture

	ODR ^a		SS ^b		No. overlapped cases	Estimated no. total cases 95% CI
	No. cases	Prevalence (%)	No. cases	Prevalence (%)		
Men	110	7.1	10	8.2	5	220 (144-499)
Women	124	7.0	16	9.2	9	220 (161-376)
Total	234	7.1	26	8.8	14	435 (329-661)

^a Oguni Diabetes Registry.
^b Sample study.

```
> exp(coef(ml))
(Intercept)      valim      register
105.00000000    0.04761905    1.00000000
```

Kui palju oli selliseid mees-diabeetikuid, kes pole kirjast ei diabeediregistris ning ei sattunud ka valimisse

```
> 5+5+105+105
[1] 220
> sum(n)+exp(confint(ml))[1,]
Waiting for profiling to be done...
      2.5 %    97.5 %
143.8100 497.5898
```

Kergesti laiendatav enam kui kahe valimi juhule; sellisel juhul võimaldab hinnata ka sõltuvusi valimite vahel:

Log-lineaarne analüüs, 3 valimit:

$$\begin{aligned} \log(n_{000}) &= \mu \\ \log(n_{100}) &= \mu + a \\ \log(n_{010}) &= \mu + b \\ \log(n_{001}) &= \mu + c \\ \log(n_{110}) &= \mu + a + b \quad (+ ab) \\ \log(n_{011}) &= \mu + b + c \quad (+ bc) \\ \log(n_{101}) &= \mu + a + c \quad (+ ac) \\ \log(n_{111}) &= \mu + a + b + c \quad (+ ab + ac + bc + abc) \end{aligned}$$

Kasutusnäiteid - Süstivate narkomaanide populatsiooni suurus

Analüüsi teostas: Kristiina Rajaleid

https://intra.tai.ee/images/prints/documents/132040857284_The_prevalence_of_injecting_drug_use_in_estonia_ENG.pdf

Andmebaasidel põhinevat korduspüügi (*capture-recapture*) meetodit rakendades määrati süstivate narkomaanide populatsiooni hinnangulist suurust. Selleks kasutati narkomaaniaga seotud infot kolmest riiklikust andmebaasist: Eesti Haigekassa (üledoosid ja narkomaania ravi), HIV-nakkuse Referentslabor (HIV-positiivseks osutunud narkootiliste ainete tarvitajad), Eesti Politsei POLIS andmebaas (narkootilise või psühhotroopse aine omamine või tarvitamine).

Toodud andmed võeti välja 2004. aasta ja vanusevahemiku 15-44 kohta. Kokku identifitseeriti erinevates andmebaasides 6 704 juhtu. Unikaalse identifitseerimiskoodiga juhte oli neist 3 264 ning peale andmete puhastamist kaasati analüüsi 3 024 kirjet. Antud andmetest lähtuvalt leiti matemaatilise modelleerimise kaudu nende süstivate narkomaanide oletuslik hulk, kes ei esine üheski andmebaasis. Taolise meetodi rakendamise tulemusena saadi, et 2004. aastal oli Eestis 15-44-aastasteid süstivaid narkomaane hinnanguliselt 13 800.

Table 9. Numbers of IDUs and overlap between four data sets.

Data sources				Number of records by gender		
HIV ref lab	Police arrests	EHIF Drug treatment	EHIF Overdoses	Males	Females	Total
no	no	no	no			
no	no	no	yes	34 (1.31%)	16 (3.70%)	50 (1.65%)
no	no	yes	no	146 (5.36%)	48 (11.11%)	194 (6.42%)
no	no	yes	yes	3 (0.12%)	2 (0.46%)	5 (0.17%)
no	yes	no	no	2193 (84.61%)	312 (72.22%)	2505 (82.84%)
no	yes	no	yes	36 (1.39%)	3 (0.69%)	39 (1.29%)
no	yes	yes	no	110 (4.24%)	29 (6.71%)	139 (4.60%)
no	yes	yes	yes	5 (0.19%)	2 (0.46%)	7 (0.23%)
yes	no	no	no	41 (1.58%)	11 (2.55%)	52 (1.72%)
yes	no	no	yes	1 (0.04%)	0 (0.00%)	1 (0.03%)
yes	no	yes	no	3 (0.12%)	2 (0.46%)	5 (0.17%)
yes	no	yes	yes	0 (0.00%)	1 (0.23%)	1 (0.03%)
yes	yes	no	no	9 (0.35%)	4 (0.23%)	13 (0.43%)
yes	yes	no	yes	3 (0.12%)	1 (0.23%)	4 (0.13%)
yes	yes	yes	no	4 (0.15%)	1 (0.23%)	5 (0.17%)
yes	yes	yes	yes	4 (0.15%)	0 (0.00%)	4 (0.13%)
				2592 (100.00%)	432 (100.00%)	3024 (100%)

APPENDIX 7. Model characteristics for capture-recapture analysis.

Model (all)	CI1	Estimate	CI2	ll(model)	df	AIC	BIC	Deviance
d1 d2 d3 d4	3535.73	6270.71	10124.89	-71.46	5	152.93	156.47	73.14
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	9198.44	15196.94	39899.88	-36.44	11	6.32	-7.75	3.08
d1 d2 d3 d4 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	6897.75	12665.12	28781.86	-36.89	10	6.25	-9.56	3.98
d1 d2 d3 d4 d1*d2 d1*d4 d2*d3 d2*d4 d3*d4	6602.70	9719.22	24102.57	-47.74	10	7.70	12.14	25.68
d1 d2 d3 d4 d1*d2 d1*d3 d2*d3 d2*d4 d3*d4	8260.55	12570.32	33715.81	-49.86	10	7.98	16.38	29.92
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d4 d3*d4	3759.12	6751.27	11394.27	-42.71	10	7.03	2.08	15.62
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d3*d4	4351.13	7033.16	13026.14	-43.85	10	7.18	4.37	17.91
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4	8298.76	12457.38	33855.98	-45.36	10	7.38	7.39	20.93
Model (men)								
d1 d2 d3 d4	3056.22	5215.47	8402.96	-66.02	5	9.47	40.68	67.76
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	10843.62	16480.46	50699.88	-33.82	11	5.96	-7.49	3.35
d1 d2 d3 d4 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	7119.18	12386.93	30599.77	-34.42	10	5.92	-8.98	4.56
d1 d2 d3 d4 d1*d2 d1*d4 d2*d3 d2*d4 d3*d4	7292.39	10204.54	30104.17	-43.46	10	7.13	9.10	22.64
d1 d2 d3 d4 d1*d2 d1*d3 d2*d3 d2*d4 d3*d4	9949.30	14178.72	45921.09	-45.98	10	7.46	14.13	27.67
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d4 d3*d4	3280.14	5713.25	9717.27	-40.50	10	6.73	3.17	16.71
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d3*d4	3609.61	5402.97	9808.97	-43.81	10	7.17	9.79	23.33
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4	11343.19	16265.75	54248.30	-39.37	10	6.58	0.91	14.45
Model (women)								
d1 d2 d3 d4	597.90	869.74	1747.40	-31.73	5	4.90	-11.10	15.98
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	1378.97	1691.78	6765.22	-26.04	11	4.94	-6.23	4.60
d1 d2 d3 d4 d1*d3 d1*d4 d2*d3 d2*d4 d3*d4	1059.12	1413.64	4131.76	-26.16	10	4.82	-8.70	4.85
d1 d2 d3 d4 d1*d2 d1*d4 d2*d3 d2*d4 d3*d4	937.28	1087.87	3944.50	-27.72	10	5.03	-5.57	7.97
d1 d2 d3 d4 d1*d2 d1*d3 d2*d3 d2*d4 d3*d4	1117.19	1307.75	5332.36	-27.61	10	5.01	-5.80	7.74
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d4 d3*d4	635.00	971.63	1836.69	-26.79	10	4.91	-7.43	6.11
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d3*d4	1219.84	1687.25	5038.25	-26.04	10	4.81	-8.94	4.60
d1 d2 d3 d4 d1*d2 d1*d3 d1*d4 d2*d3 d2*d4	790.84	922.29	2751.04	-28.08	10	5.08	-4.85	8.69

d1 = HIV reference laboratory; d2 = police arrest data; d3 = HIV treatment; HIV overdoes