Gevoeligheid van de vangst-hervangst methode voor geschonde aannames

S.C. Gerritse B.F.M. Bakker[2] P.G.M. van der Heijden[3]

Universiteit Utrecht

[2]Centraal Bureau voor de Statistiek/Vrije Universiteit

[3]Universiteit Utrecht/Univerisity of Southampton

Kwaliteitsrapportage Volkstelling 2011

- Onderdekking Gemeentelijke Basis Administratie (GBA, Population register)

Kwaliteitsrapportage Volkstelling 2011

- Onderdekking Gemeentelijke Basis Administratie (GBA, Population register)
- Usual residents

Kwaliteitsrapportage Volkstelling 2011

- Onderdekking Gemeentelijke Basis Administratie (GBA, Population register)
- Usual residents

Vangst hervangst begon met het schatten van dierpopulaties.

- Petersen methode, 1894
- Simpelste methode - twee steekproeven

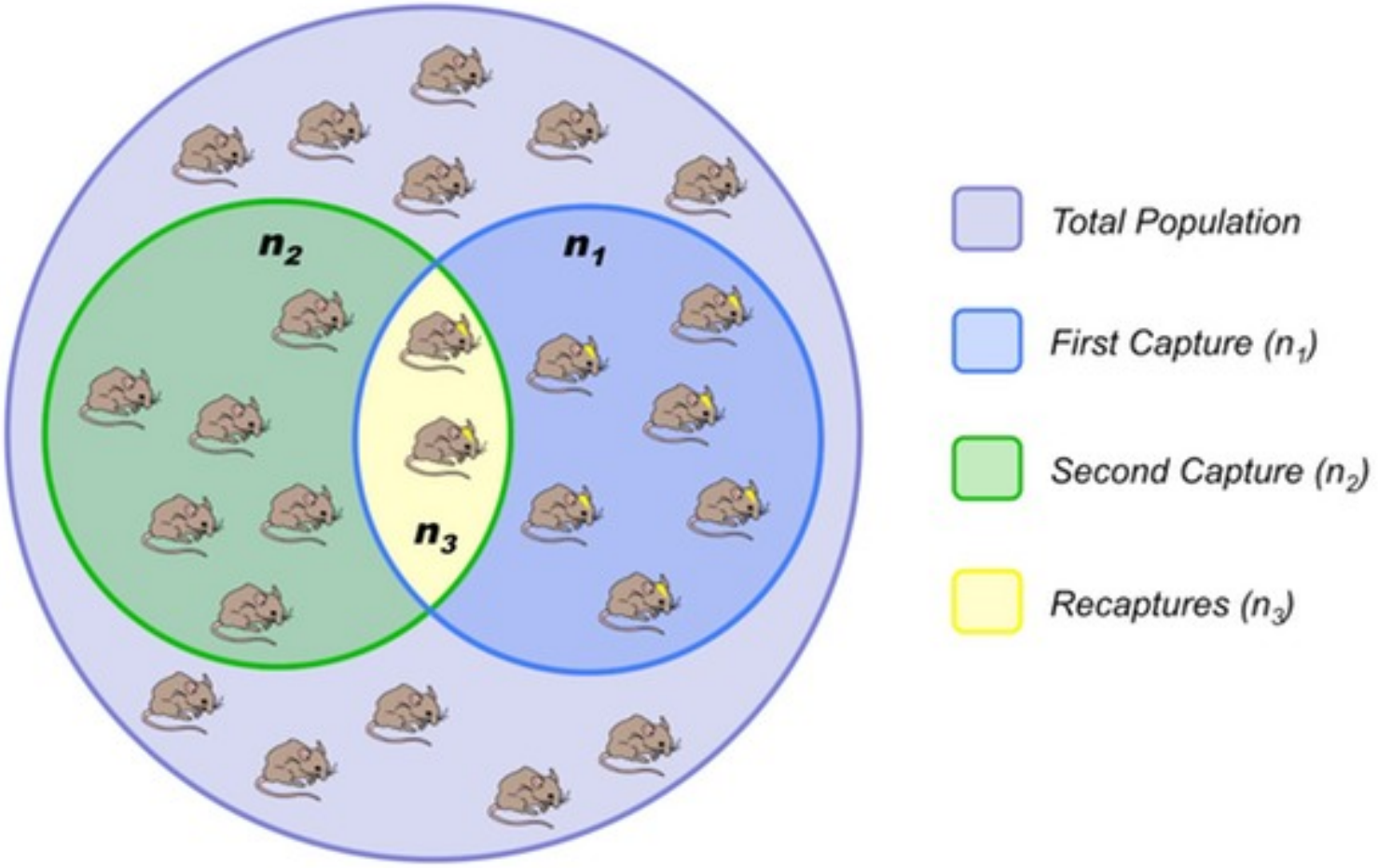
Kwaliteitsrapportage Volkstelling 2011

- Onderdekking Gemeentelijke Basis Administratie (GBA, Population register)
- Usual residents

Vangst hervangst begon met het schatten van dierpopulaties.

- Petersen methode, 1894
- Simpelste methode - twee steekproeven

VANGST-HERVANGST



VANGST-HERVANGST

Koppelen van twee (incomplete) registers

VANGST-HERVANGST

Koppelen van twee (incomplete) registers

- Populatie register, Gemeentelijke Basis Administratie (GBA)

VANGST-HERVANGST

Koppelen van twee (incomplete) registers

- Populatie register, Gemeentelijke Basis Administratie (GBA)
- HKS, HerKenningsdienst Systeem politie

VANGST-HERVANGST

Koppelen van twee (incomplete) registers

- Populatie register, Gemeentelijke Basis Administratie (GBA)
- HKS, HerKenningsdienst Systeem politie

Table: Expected values

	HKS	
GBA	1	0
1	m_{11}	m_{10}
0	m_{01}	m_{00}

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

Loglineair model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

where $\lambda_1^A = \lambda_1^B = 0$.

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

Loglineair model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

where $\lambda_1^A = \lambda_1^B = 0$.

Odds ratio

$$\frac{m_{00}m_{11}}{m_{10}m_{01}} = 1$$

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

Loglineair model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

where $\lambda_1^A = \lambda_1^B = 0$.

Odds ratio

$$\frac{m_{00}m_{11}}{m_{10}m_{01}} = 1$$

Twee manieren om m_{00} te schatten:

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

Loglineair model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

where $\lambda_1^A = \lambda_1^B = 0$.

Odds ratio

$$\frac{m_{00}m_{11}}{m_{10}m_{01}} = 1$$

Twee manieren om m_{00} te schatten:

- Poisson loglinear regression: $\hat{m}_{00} = \exp(\hat{\lambda})$

SCHATTEN GEMISTE POPULATIE

Onder onafhankelijkheid

Loglineair model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

where $\lambda_1^A = \lambda_1^B = 0$.

Odds ratio

$$\frac{m_{00}m_{11}}{m_{10}m_{01}} = 1$$

Twee manieren om m_{00} te schatten:

- Poisson loglinear regression: $\hat{m}_{00} = \exp(\hat{\lambda})$
- Maximum likelihood estimate (mle):

$$\hat{m}_{00} = \frac{\hat{m}_{10}\hat{m}_{01}}{\hat{m}_{11}} = \frac{n_{10}n_{01}}{n_{11}}.$$

AANNAMES

Aannames

- Onafhankelijkheid tussen registers

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen
- Gesloten populatie

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen
- Gesloten populatie

Twee problemen:

- Hoe haal je de aannames?

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen
- Gesloten populatie

Twee problemen:

- Hoe haal je de aannames?
- Effect van schending van aannames op de populatie?

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen
- Gesloten populatie

Twee problemen:

- Hoe haal je de aannames?
- Effect van schending van aannames op de populatie?
  - Literatuur?

AANNAMES

Aannames

- Onafhankelijkheid tussen registers
- Perfect koppelen van registers
- Geen erroneous captures
- Heterogene inclusie kansen
- Gesloten populatie

Twee problemen:

- Hoe haal je de aannames?
- Effect van schending van aannames op de populatie?
  - Literatuur?
  - Sensitiviteitsanalyses

DATA

Vergelijken van twee verschillende nationaliteitsgroepen.

DATA

Vergelijken van twee verschillende nationaliteitsgroepen.

- Afghanistan, Irak, Iran
  - Hebben visa nodig om in Nederland te werken en verblijven

DATA

Vergelijken van twee verschillende nationaliteitsgroepen.

- Afghanistan, Irak, Iran
  - Hebben visa nodig om in Nederland te werken en verblijven
- Polen
  - EU $\rightarrow$ vrijheid van werk en verblijf

DATA

Vergelijken van twee verschillende nationaliteitsgroepen.

- Afghanistan, Irak, Iran
  - Hebben visa nodig om in Nederland te werken en verblijven
- Polen
  - EU $\rightarrow$ vrijheid van werk en verblijf

Table: Geobserveerde waardes en mle $\hat{m}_{00}$

	HKS	
GBA	1	0
1	1,085	26,254
0	255	6,170.3

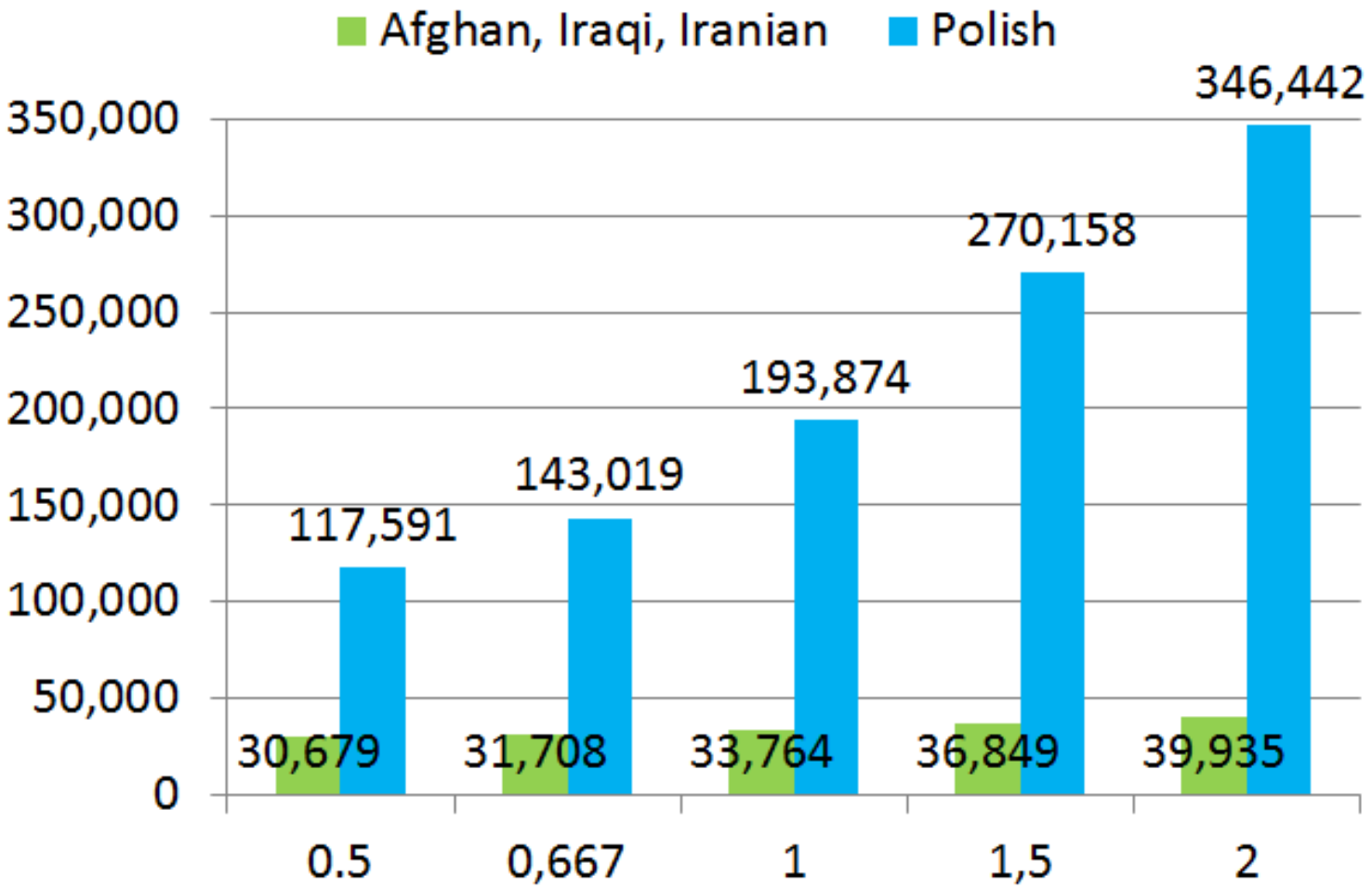
(a) Afghanistan, Irak and Iran verblijvend in Nederland in 2007

	HKS	
GBA	1	0
1	374	39,488
0	1,445	152,567.3

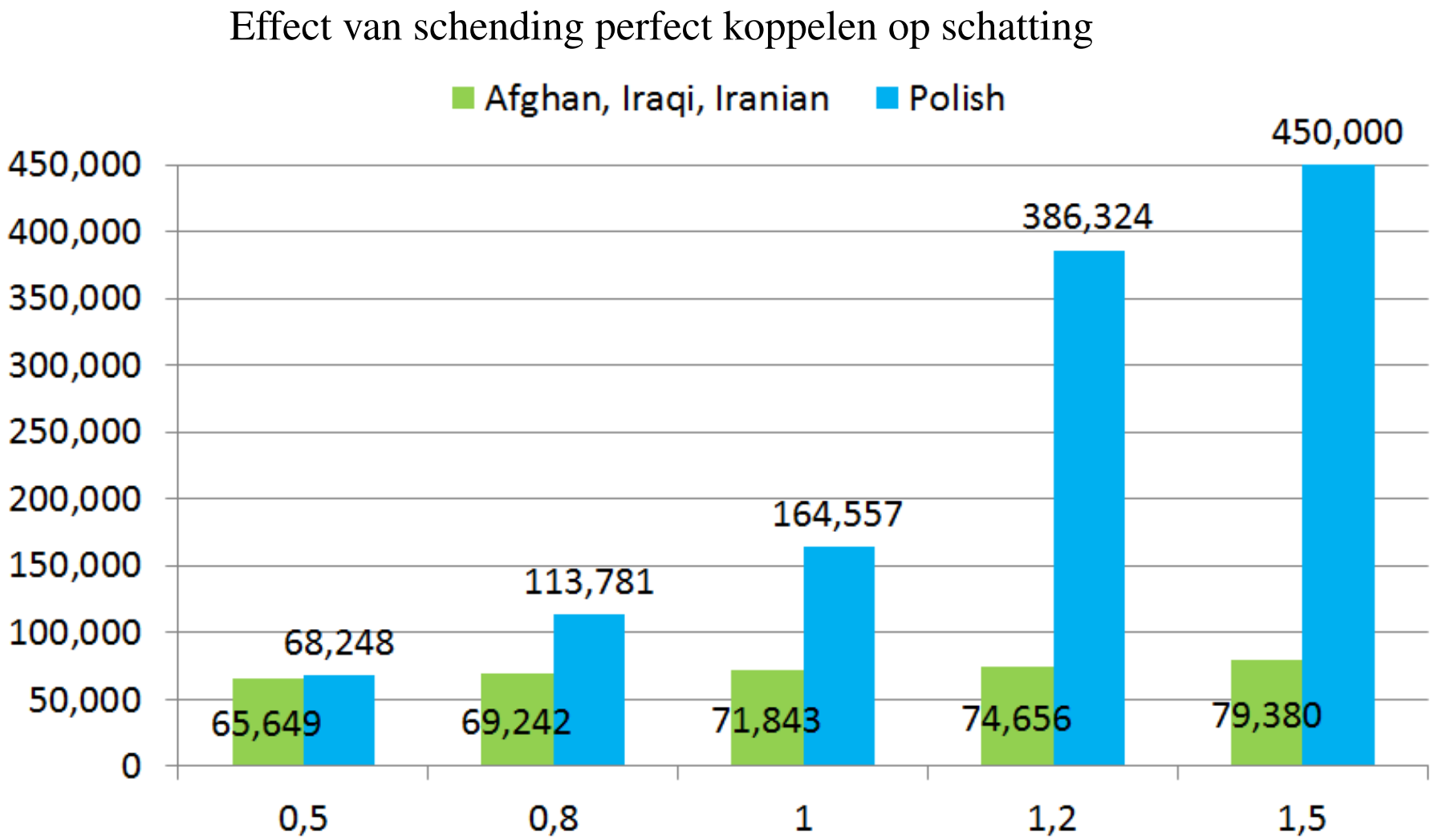
(b) Polen verblijvend in Nederland in 2009

ONAFHANKELIJKHEID

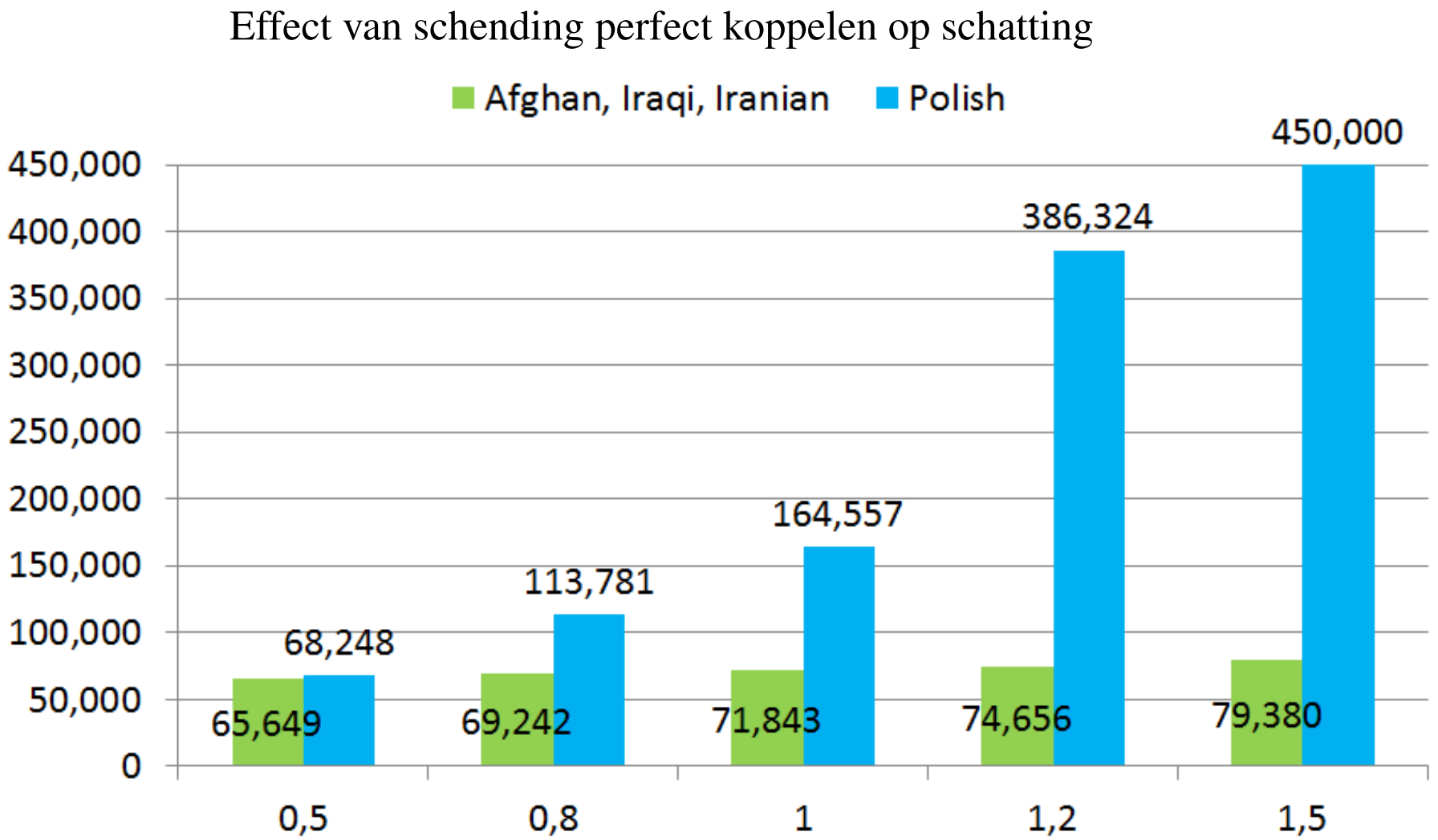
Effect van schending onafhankelijkheidsaanname op schatting



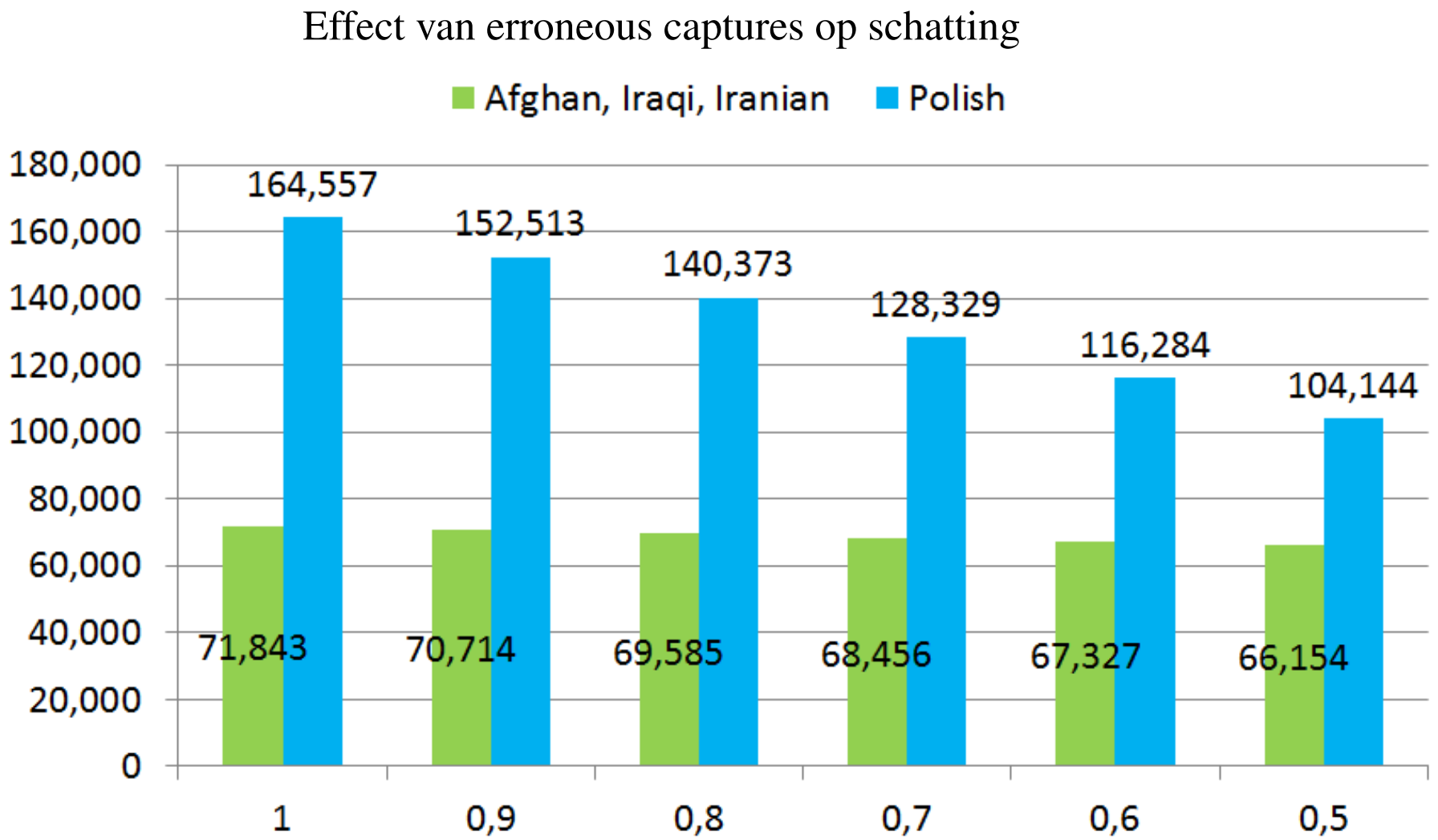
PERFECT KOPPELEN



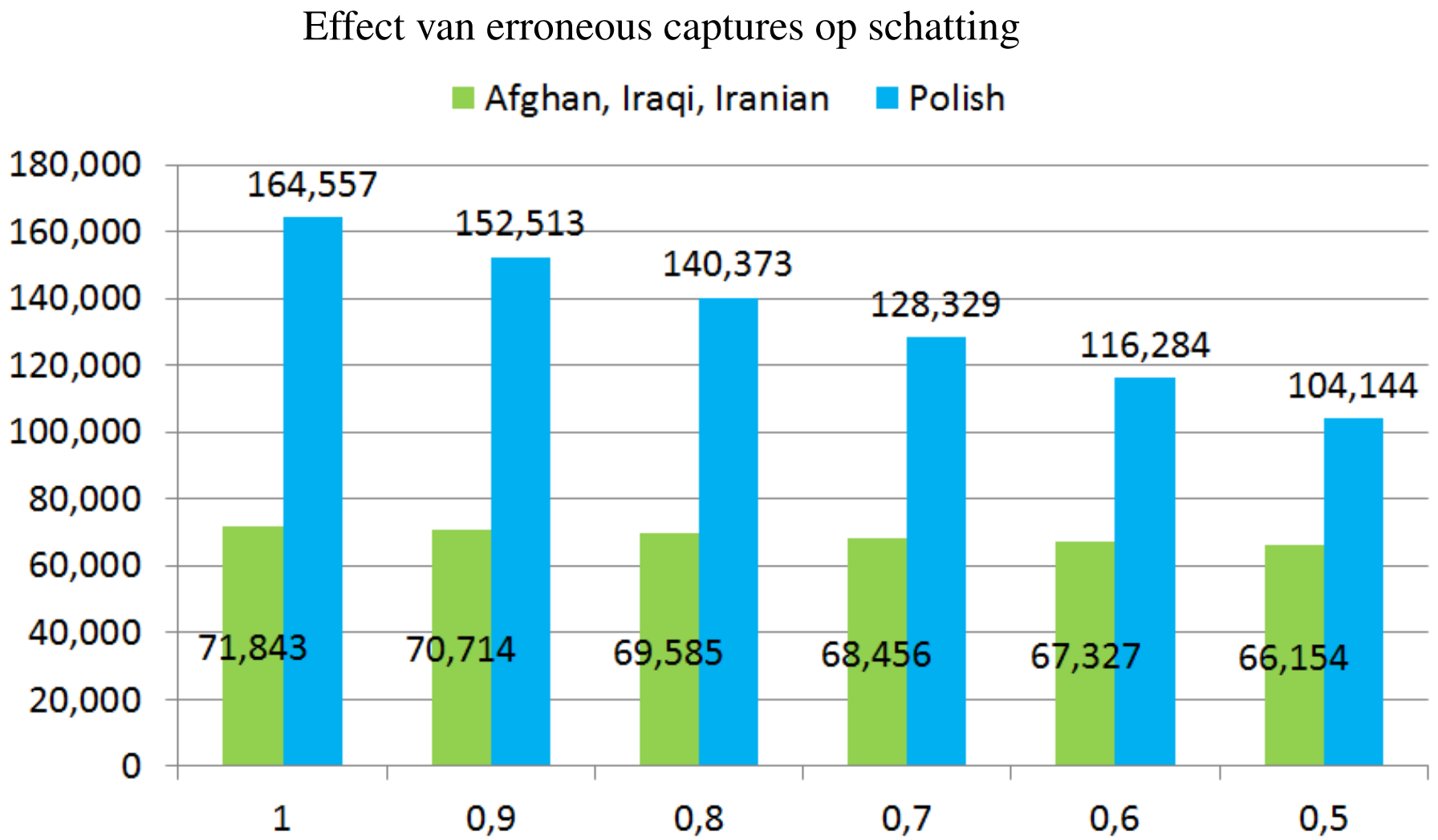
PERFECT KOPPELEN



ERRONEOUS CAPTURES



ERRONEOUS CAPTURES



Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Usual residents:

- Vangst hervangst: 184,336 gemist door alle drie registers

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Usual residents:

- Vangst hervangst: 184,336 gemist door alle drie registers
- 33,000 usual residents in HKS en WNB

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Usual residents:

- Vangst hervangst: 184,336 gemist door alle drie registers
- 33,000 usual residents in HKS en WNB
- Onderdekking GBA 217,336

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Usual residents:

- Vangst hervangst: 184,336 gemist door alle drie registers
- 33,000 usual residents in HKS en WNB
- Onderdekking GBA 217,336
- GBA geregistreerde usual residents: 16,638,805

Vangst hervangst voor usual residents:

- Derde register (WerkNemers Bestand, Polis)
- Meerdere covariaten (leeftijd, geslacht, nationaliteitsgroep en usual residence)
- Determistisch en probabilistisch linkage
- Verwijderen Erroneous captures
- Korte steekproefperiode

Usual residents:

- Vangst hervangst: 184,336 gemist door alle drie registers
- 33,000 usual residents in HKS en WNB
- Onderdekking GBA 217,336
- GBA geregistreerde usual residents: 16,638,805
- Totaal aantal usual residents: 16,856,141, waarvan 1.3% onderdekking.

Thank you!

- Email: s.c.gerritse@uu.nl, sc.gerritse@cbs.nl
- Website: http://scgerritse.wordpress.com/