Estimation of Non-Registered Usual Residents in the Netherlands, ultimo September 2010[1]

Bart F.M. Bakker [2,3], Susanna Gerritse [4], Peter G.M. van der Heijden[4,5],
Jan van der Laan [2], Rik van der Vliet[2], Maarten Cruyff [4]

FIRST DRAFT. DO NOT CITE WITHOUT PERMISSION OF THE AUTHORS

Abstract

From the 2011 Census round, the population "usual residents" is defined as those who have lived or intend to live for a period of more than 12 months in their place of usual residence. Most of the register-based statistics describe the total number of registered persons in e.g. a population register at a particular moment or period. To estimate the non-registered population of usual residents, capture – recapture methods are available. By making a three list estimation, making restrictions to one day for the period-based registers and a short period for the event-based register, applying a stringent linkage method and deletion of erroneous records, most of the assumptions of the capture – recapture method are met. However, to estimate the number of usual residents, we need to divide the estimated total number of persons into those who stay longer than a year in The Netherlands and those who did not. In this paper we present a method to estimate the number of usual residents even though the residence duration is observed in only one of the lists. We apply the method to Dutch empirical data.

Keywords: population estimates, capture-recapture methodology, triple system estimates, undercount, undercoverage

Introduction

From the 2011 Census round, the population is the set of usual residents, defined as those who have lived or intend to live for a period of more than 12 months in their place of usual residence (European Parliament, 2008)[6]. This definition is hard to apply if statistics are register-based. Part of the problem is that intentions are not registered at all, but also the estimation of the number of residents who lived longer than a year is difficult. The actual population differs from the registered population (Bakker & Daas, 2012; Bakker, Van Rooijen & Van Toor, 2014), the latter also containing residents who immigrated from other countries of the European Union into the Netherlands and did not register as such, immigrants who (are planning to) stay shorter than a year and undocumented immigrants. An important part of the difference between the registered and the actual population in the Netherlands ultimo 2010, is the group of temporary workers from eastern Europe, in particular Poland, who do not register themselves in the population register (Gerritse, Van der Heijden, & Bakker, 2015). Within the European Union individuals with a European nationality are free to migrate and most of them (until January 1st 2014 Bulgarians and Romanians required a working permit) are free to work without a working permit. Temporary workers and other immigrants are required to register in the Dutch population register if they stay longer than four months or are planning to do so. Due to ignorance, a very small penalty and other reasons a large part of those who

---

should register do not. A second large group are refugees and asylum seekers who provisionally stay in anticipation of the decision of the authorities for a residence permit. They are not always allowed to register in the population register. Finally, a third category are the undocumented immigrants, for a great part former asylum seekers who did not receive a residence permit.

To estimate the non-registered population, capture – recapture methods are available (Fienberg, 1972; Bishop, Fienberg & Holland, 1975; IWGDMF, 1995;Van der Heijden *et al.* 2012; Baffour, Brown & Smith, 2013). Two or more registers that contain information on (parts of) the population are linked and from the shared units in the registers the part of the population missing from all registers can be estimated. When using two registers this method has the following assumptions: (a) the inclusion probability of being registered in the first register is independent of the inclusion probability of being registered in the second register; (b) the population should be closed (no elements appear or disappear) during the data collection period; (c) all elements of the population should have a positive probability of being registered in each of the registers; (d) the registers are perfectly linked; (e) the registers do not include erroneous captures (elements that do not belong to the population). Violation of these assumptions could lead to highly biased estimates. E.g. violation of the independence assumption leads to biased estimations (Brown *et al.* 2006; Gerritse *et al.,* 2015)

However, the total number of persons is not the population that we are interested in. Part of the total number of persons is not a usual resident, because they live here for a shorter period of time than a year and do not intend to stay for more than a year. In most administrative sources, information on the residence duration is not explicitly included. In none of the administrative sources, information on the intentions on the residence duration is available. In this paper, we present a procedure to estimate the total population of usual residents, even if only one source includes observed information on the residence duration. We apply the procedure to empirical data of the Netherlands.

For the time being, we assume that those who are registered in the Population Register (PR) are usual residents. We neglect the possibility of overcoverage in the PR, because to estimate this, more extensive research is needed with the use of additional data sources which are not available yet. Moreover, we restrict the estimation to the population 15-65 years of age because of lack of appropriate data sources.

The capture – recapture method

To estimate the population of usual residents, we apply the capture - recapture method. As just mentioned the capture-recapture method has a number of assumptions. The following paragraphs describe how these assumptions are met in our application of the method.

We use three registers. This relaxes the assumption (a) of independency between the registers. We now only assume that the three factor interaction between the inclusion probabilities of the three registers is zero. The first register is the official Dutch Population Register (PR). It includes the total registered population of the Netherlands. The second one contains the employees in registered Dutch companies and is called the Employee Register (ER). The third one contains the crime suspects that are registered as such by the police (CSR). Only the PR contains actual information on the residence duration, because the first registration date is available.

The assumption (b) that the population is closed, is easily met for the PR and ER because both registers describe a period and any common date or period can be selected. The assumption is satisfied by restricting the data to the status of only one day: *ultimo* September 2010. *Ultimo* September is operationalized as the latest Friday in September. This date describes the status *ultimo* September best, because many jobs start or end in the weekend or on the first day of the month. This cannot be applied to the CSR because this register is event-based: crime suspects are registered of which the police make a report. The number of events on one specific day is not enough to apply the capture – capture method. In order to satisfy the assumption as well as possible, we restrict the period of the CSR to the second half of 2010. This period has been chosen because the data of the CSR of 2011 were not available yet by the start of this research.

Not all elements of the population have a positive probability of being registered (assumption c). The ER is restricted to the population of 15-65 years of age, while the CSR is restricted to the persons of 12 years and older. Because of these restrictions, we are not able to estimate the total population, but only the total population of 15-65 years of age. The youth and elderly are estimated in another way and this is not reported here.

The capture –recapture method is sensitive to linkage error. Therefore, the three registers are linked pairwise with much caution (assumption d). In the first step, the records are linked deterministically on a personal identification number that is widely available in administrative data sources in the Netherlands. The remaining records are linked probabilistically (Fellegi & Sunter, 1969; Ariel et al., 2014). To reduce the number of possible pairs, the data are blocked on variables that are assumed to be of very high quality. For the linkage of PR – ER and PR- CSR date of birth, sex, postal code, house number and extension are used after blocking on postal code or date of birth. The ER and CSR are linked in a slightly different way because a large number of records do not contain a Dutch postal code. Therefore the linkage is done on date of birth, sex, place of residence, address and house number after blocking on date of birth, or on place of residency and month of birth or day of birth. Table 1 shows the results of the linkage procedures. These are difficult to evaluate because the number of real links is unknown. However, the probabilistic linkage leads to an increase of 0.1% (PR-ER), 3.3% (PR-CSR) and 9.9% (ER-CSR) of the number of linked records.

Table 1. Linkage effectiveness

		source	linked	of which probabilistically	not linked
			%		
Linkage PR – ER	PR		42,4	0,1	57,6
	ER		96,9	0,1	3,1
Linkage PR – CSR	PR		1,1	3,3	98,9
	CSR		56,2	3,3	43,8
Linkage ER – CSR	ER		0,7	9,9	99,3
	CSR		19,8	9,9	80,2

Despite the attention paid to the linkage method and the careful execution of the method, it is still possible that not all records are linked that should be linked. One of the main reasons is that the linkage keys are incomplete or entirely missing. In Table 2, the amount of missing values is given for the different combinations of registers for the population selection as described in the next paragraph. From the last row in the table it is clear that the records in the CSR that do not link to either the PR nor the ER, contain large numbers of missing values in the linkage key. In the linkage key used for linkage to the PR this is 27.7% and in the linkage key used for linkage to the ER this is 37.7%. We will use this information to correct our initial capture-recapture estimates.

Table 2. Missing values in the linkage key in PR, ER and CSR linkage

linked to			gender	birth date	housenr	addition	postal code	country	street	place of residence	address1[a]	address2[b]
PR	ER	CSR	%									
y	y	y	0,0	0,0	0,3	0,0	0,1	0,1	0,3	0,1	0,3	0,3
y	n	y	0,0	0,0	0,6	0,0	0,2	0,3	0,6	0,3	0,6	0,6
n	y	y	0,0	0,0	3,7	0,0	0,6	1,1	4,5	0,9	3,7	4,8
n	n	y	0,5	0,0	27,7	0,0	0,2	19,7	26,2	11,2	27,7	37,7

a Address1 is a combination of: house number, postal code and addition. A missing value in one of these variables leads to a missing value for the linkage key as a whole. This linkage key is used for the linkage with the PR.
b Address2 is a combination of: street, house number, place of residence and country. A missing value in one of these variables leads to a missing value for the linkage key as a whole. This linkage key is used for the linkage with the ER.

To prevent erroneous captures (assumption e), we removed the records from the ER and CSR of persons who do not belong to the population. (For the PR, see the next section.) From the ER and CSR we have removed the following categories: (a) the few persons with the Dutch nationality not registered in the PR because we expect them to be expats working in another country and therefore not belonging to the population; (b) persons with an address in Belgium or Germany, the neighbouring countries of The Netherlands, because it is likely that they live in Belgium or Germany and are

only temporary in The Netherlands to work, to go to school, to shop or to have a short holiday; (c) persons who are reported for a crime by the border police at the airport or elsewhere because they did not enter the country at all. Despite the removal of these categories, it is still possible that there are persons that committed a crime and are registered as such in the CSR but do not belong to the population. For example, drugs runners living in France, Luxembourg, Switzerland, etc., gangs of pickpockets from Eastern Europe, or tourists arrested for drunk driving. In the estimation of the non-registered population, we have to take into account that their numbers are unknown and it is reasonable to assume that they overlap with the category that did not link to the other registers because of incomplete or unknown linkage keys.

The residence duration

Most of the assumptions of the capture – recapture method are met by making a three-register estimation, making restrictions to one day for the period-based registers and a short period for the event-based register, applying a stringent linkage method and deletion of erroneous records,. However, to estimate the number of usual residents, we need to divide the estimated total number of persons into those who stay longer than a year in The Netherlands and those who do not. First, we estimate the total population and secondly we estimate the number of usual residents by adding the covariate "Residence duration" to the estimation model. However, residence duration is not available in all of our three registers.

For the time being, we assume that persons who are registered in the population register (intend to) stay longer than a year in the Netherlands because data sources needed for the estimation of overcoverage of the PR are not available yet. Most of the registered persons in the ER and CSR are also registered in the PR. So, we only have to estimate the residence duration for those not in the population register. We start the procedure with the 113 thousand persons in the ER who are not registered in the PR, because it is likely that those who have a job in the Netherlands also stay in the Netherlands during the time they have the job. Therefore, the residence duration will be derived from the jobs they had consecutively. After that, we impute the residence duration in the remaining CSR-records by Predictive Mean Matching (PMM; Van Buuren, 2012).

Crucial in this operation is to derive the residence duration from employment records. These records contain information on the starting and ending dates of sequential jobs. The residence duration of employees with only one job is estimated as the duration of that particular job. For employees with more than one job, the residence duration is defined as the period of continuous stay. However, if there are gaps in between jobs, we have to decide on which duration of a gap is acceptable to assume continuous stay. In order to decide on this issue, we investigated six scenario's: 8, 15, 22, 31, 62 and 93 days. For each scenario, we determined the number of employees that would have a residence duration longer than one year (Table 3).

Table3. Six scenario's for residence duration derived from employment records by nationality

	not registered in PR						
	total	residence duration >1 jr in scenario					
		8 days	15 days	22 days	31 days	62 days	93 days
Nationality	x 1.000	%					
EU15 (excl. Netherlands)	18,7	42	42	43	47	53	58
Polish	80,7	19	20	21	25	32	36
Other EU	10,8	21	22	22	24	29	32
Other Western	0,5	43	43	43	44	45	48
Turkish, Morrocan, Antillean, Surinam	0,6	54	55	56	59	66	71
Iraqi, Iranian, Afghan, Asylum seeker countries Africa	0,3	53	54	57	61	65	71
Other Balkan, former Soviet Union, other Asian, Latin American	0,9	36	36	37	38	41	46
Not mentioned elsewhere	0,4	46	46	47	50	56	60
Total	113,0	24	25	25	29	35	40

Of the 730,000 persons in the union of PR and ER who do not have the Dutch nationality, 113,000 are not registered in the PR. We assume that the 617,000 persons who are registered in both are usual residents. For the 113,000 we choose the scenario of 31 days, because:

- The biggest groups are European, in particular Eastern European. The probability that they return to their homeland if the gap is larger than 31 days becomes larger.

- Return to their homeland will be more like a holiday than a return to live there if the period is restricted to 31 days.
- Financially, it is rather easy to bridge a gap of one month.
- The differences with the scenarios of 22 and 62 days are relatively large.
- The majority of the persons with more than one job have had two jobs consecutively (not in table). If the start of the first of those two jobs is more than a year ago, they have stayed at least eleven month in the Netherlands.

According to this scenario, 29% of the 113.000 persons are usual residents.

After this step, we have a measure of residence duration of all persons who are registered in the PR and ER. Thus, all persons registered in the CSR who link to the PR or the ER, have assigned a residence duration too. However, for remaining persons registered in the CSR it is not possible to derive the residence duration because events (apparently committed crimes) are registered. Therefore, we have to impute the residence duration in the records in the CSR that do not link to the PR and ER. We will use the people that are registered in the ER but not in the PR as donor for the residence duration. We have chosen this subpopulation because we assume that this subpopulation will resemble the individuals that are only in the CSR the most. A worse alternative would be to use the PR for donors, but those registered in the PR most of the time have settled long time ago and for other reasons than the temporary workers or asylum seekers. We make use of PMM (Van Buuren, 2012), and use nationality group, age and sex as the predictors. PMM assumes independence of the process of missingness conditional on the predictor variables. Nationality group has 8 categories: (1) EU15 (excl. Netherlands) (2) Polish (3) Other EU (4) Other western (5) Turkish, Moroccan, Antillean, Surinam (6) Iraqi, Iranian, Afghan, asylum seeker countries Africa (7) Other Balkan, former Soviet Union, other Asian, Latin American, and (8) not mentioned elsewhere. The countries are clustered according to main migration motives, migration legislation, regulations of the PR and size. For age, we use four categories: (1) 15-24 (2) 25-34 (3) 35-49 and (4) 50-64 years of age. Sex has the categories (1) male and (2) female.

Using the method above we can now derive for each person in the PR, ER and CSR whether or not they are a usual resident:
- Those registered in the PR are all assumed to be usual residents.
- Those registered in the ER who are not registered in the PR are usual residents if they work continuously (or with gaps of less than 32 days between jobs) for more than a year in the Netherlands.
- Those registered in the CSR who are not registered in the PR or the ER, are usual residents if their imputed value of residence duration is longer than a year.

We will use the residence duration as a covariate in the log-linear model of the capture – recapture method.

Capture – recapture estimates

After imputing the residence duration in the remaining part of the CSR, all information is available to apply the capture – recapture method. The saturated log-linear model, assuming the three factor interaction to be zero, has seven parameters and only seven observed counts:

(1) $\text{Log } m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$

A, B and C are the variables denoting being an element in the registers PR, ER and CSR. A, B and C are indexed respectively, by i, j and k, where a subscript =1 if the element is in the register and 0 if the element is not in the register. Expected values are denoted by m_{ijk}. Table 4 shows the observed values for the combination of the three registers. The m_{000} in the table has to be estimated and divided into usual residents and non-usual residents. According to model (1), i.e. a model without covariates, the total population is estimated at 1,064 thousand persons who are not registered in the PR: 113 thousand persons who are registered in the ER but not in the PR, 5 thousand persons registered in the CSR but not in the PR or ER, and an estimation of 946 thousand who are not registered in either register. To distinguish between usual residents and non-usual residents, we add the covariate Usual residence to model 1. To improve the plausibility of the assumption that the three factor interaction of being observed in the registers A, B and C is zero, we add covariates nationality group, age and sex (see the former section for an explanation of these variables).[7] To simplify estimation, we estimate models for all eight nationality groups separately. To prevent that we overfit the model, we search for a model

---

[7] An additional advantage is that this makes it possible to estimate the distribution of these variables in the total population.

that fits the data well and is as parsimonious as possible with the Bayesian Information Criterion (BIC). The BIC criterion has a larger penalty for the number of parameters in the model than the Akaike Information Criterion (AIC) and therefore leads to more parsimonious models when the sample size is large. Using the BIC criterion, the model selected for each nationality group and the population size estimate is shown in Table 5.

Table 4. The observed values for the three registers

PR	ER	CSR		
		y	n	Total
PR	ER	$x 1.000$		
y	y	2,1	259,8	261,9
y	n	4,9	350,6	355,4
n	y	0,4	112,5	112,9
n	n	5,1	m_{000}	
Total		12,4		

Table 5. Estimated number of usual residents not registered in the PR for each nationality group

	model*	m_{000}	
Nationality		$x 1000$	
EU15 (excl. Netherlands)	PU+PE+AU+CS+EC+SU+PC+ESA+PSA+PCA	89	
Polish	PU+EC+CS+EA+CA+PE+PC+PSA+SAU	81	
Other EU	PU+AU+EC+SU+PA+ES+PEA+PSA+PCS	50	
Other Western	PU+PE+AU+PS+CS+ESA	7	
Turkish, Morrocan, Antillean, Surinam	PU+SA+PC+ESA+ECA+ECS+PES	3	
Iraqi, Iranian, Afghan, Asylum seeker countries Africa	PU+PE+CS+CA+PS+EC+ESA	7	
Other Balkan, former Soviet Union, other Asian, Latin American	PU+CS+ESA+PES+ECA	19	
Not mentioned elsewhere	PU+PE+CS+PS+EC+CA+ESA	6	
Total m_{000}		262	
In ER, not in PR			33
IN CSR, not in ER and PR			2
total estimation not in PR			297

*) P=in PR; E=in ER; C=in CSR; A=Age; S=Sex; U=usual residence; all main effects are in the models; PU means that the interaction between P and U is part of the model

After adding the covariates, the estimated number of usual residents not registered in the PR is 297 thousand. Previous research by Hoogteijling (2002) and Bakker (2009) showed much smaller estimates. It is not entirely clear what their definitions of the estimated populations are, but presumably they used a definition of at least four month residence. The first one estimated the non-registered for 2001 between 46 and 116 thousand. The second one estimated the number of non-registered at 236.5 thousand in 2006. We expect that the number of not-registered persons increased since then, because of the changed position of the inhabitants of new EU-countries. This has led to a large migration flow to the Netherlands from, in particular, Poland and other eastern European countries. On the other side we expect a decline in the estimated undercount, because the current definition of usual residence with a period of one year is more strict than the ones used before. If we take in everything in a glance, we expect that the estimation of 297 thousand is the absolute upper bound of the real number of usual residents not registered in the PR.

As was already mentioned, the number of records in the CSR that cannot be linked to the two other registers is very large for which there are two possible reasons: erroneous captures and missed links. There are erroneous captures, because in this cell, persons who do not live in the Netherlands are captured because they have committed a crime (see earlier section

directly after Table 2). The experts we consulted were not able to give estimates for the number of "criminal tourists". In order to investigate the effect of both reasons we simulate that 10%,20% or 30% of the unlinked persons in the CSR is a criminal tourist (Table 6). As expected, the estimated number of usual residents becomes smaller with each reduction of 10%. The effect of the 10% reductions varies between 27 and 31 thousand. We estimate a number of 210,000 usual residents who are not registered in the PR if we assume that 30% of the not linked CSR-records are in fact criminal tourists.

Table 6. Estimated number of usual residents not registered in the PR for each nationality group, random reduction of 10%, 20% and 30% in the CSR-records not linked to PR and ER*

	0%		10%		20%		30% reduction	
Nationality	m $_{000}$ x1000							
EU15 (excl. Netherlands)	89		79		67		56	
Polish	81		73		64		57	
Other EU	50		45		40		35	
Other Western	7		7		6		5	
Turkish, Morrocan, Antillean, Surinam	3		2		2		2	
Iraqi, Iranian, Afghan, Asylum seeker countries Africa	7		6		6		5	
Other Balkan, former Soviet Union, other Asian, Latin American	19		18		15		13	
Not mentioned elsewhere	6		5		5		4	
Total m $_{000}$	262		234		203		176	
In ER, not in PR		33		33		33		33
IN CSR, not in ER or PR		2		2		1		1
total estimation not in PR		297		269		238		210

*) The models are almost the same as in table 5

The second reason for the large number of records that cannot be linked to the other registers is that we missed links. This is very likely because of the missing values in the linkage keys of part of the CSR-records. However, we believe that there is a large overlap between the "criminal tourists" and the missed links because it is much harder to identify criminal tourists than usual inhabitants. Therefore, we simulated that in addition to 30% criminal tourists, we have 10% missed links. Moreover, we expect that the vast majority of the missed links should actually link to the ER and not to the PR because police officers have access to the PR to check the identities of criminal suspects. If 10% of these CSR-records are moved to the cell ER but not linking to the PR, the estimated number of usual residents is 184 thousand, a further reduction of 26 thousand usual residents.

Discussion

In this paper, we estimated the number of non-registered usual residents in the Netherlands by using the capture-recapture methods. By making a three-list estimation, making restrictions to one day for the period-based registers (PR and ER) and a short period for the event-based register (CSR), applying a stringent linkage method and deletion of erroneous records, most of the assumptions of the capture – recapture method are met. We derived the residence duration of the employed from their employment history. If they held consecutive jobs for longer than a year and the period between jobs is shorter than 32 days, they are assumed to be usual residents. For those in the CSR who are not registered in the PR and ER, we imputed the residence duration based on the employment histories of the persons in the ER not registered in the PR. Finally, we estimated the number of non-registered usual residents by applying capture-recapture methodology. We divided the observed population into eight nationality groups and searched for well-fitting and parsimonious models. Under the assumption that the individuals in the CSR who are not in the PR and ER are erroneous captures, we brought down this number with 10, 20 and 30%. Moreover, we simulated that 10% of the

remaining records in this cell are missed links. We estimated under each of these scenarios the population size again. In the end, we conclude that the number of usual residents (aged 15-65) not registered in the PR is not higher than 297,000, and probably will not be lower than 184,000.

It is complex to apply capture – recapture methodology, because of the assumptions that are easily violated and the severe sensitivity of the outcomes to these violations. If you are planning to use the method, e.g. to estimate the undercount of a register-based census, be sure to form a multidisciplinary team, in which knowledge on the sources, the methods and the real world are combined. In our case, we combined knowledge on migration laws and regulations, migration patterns, labour market behaviour, criminality, linkage methodology and capture-recapture methodology, and peculiarities of the sources. Without that, it would have been impossible to finish this research in a proper way. And even now, the result is based on a large number of choices, of which some are quite arbitrary. Simulations could give insight into the consequences of violating the assumptions (Gerritse et al., 2015). Similarity with results from other methods, would give more confidence in the outcomes (Bakker, 2009). Statistics Netherlands will continue the work to improve the estimation of the size of the non-registered usual residents.

References

Ariel, A., M.de Groot, G. van Grootheest, J. van der Laan, J. Smit, B. Verkerk, B. Bakker, 2014, *Record linkage in health data: a simulation study* (The Hague / Heerlen: Statistics Netherlands)

Baffour, B., J.J. Brown, P.W.F Smith, 2013, An investigation of triple system estimators in censuses. *Statistical Journal of the International Association for Official Statistics*, 29, 53-68

Bakker, B.F.M., 2009, Trek alle registers open! (Amsterdam: VU University Press)

Bakker, B.F.M., & P. Daas, 2012, Some Methodological Issues of Register Based Research, *Statistica Neerlandica,* vol. 66, nr. 1, pp. 2-7

Bakker, B.F.M., J. van Rooijen & L. van Toor, 2014, The System of social statistical datasets of Statistics Netherlands: an integral approach to the production of register-based social statistics, forthcoming in *Statistical Journal of the International Association of Official Statistics*

Bishop, Y., Fienberg, S., & Holland, P., 1975, *Discrete multivariate analysis, theory and practice*. New York: McGraw-Hill.

Brown, J.J., O. Abott & I.D. Diamond, 2006, Dependence in the 2001 one-number census project. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 169, 883-902

European Parliament, 2008, Regulation (EC) No 763/2008 of the European Parliament and of the council of 9 July 2008 on population and housing censuses, In: *Official Journal of the European Union*, 13.8.2008, pp. L 218/14-L 218/20

Fellegi , I.P. & A.B. Sunter, 1969, A Theory for Record Linkage, *Journal of the American Statistical Association*, Vol. 64, No. 328, pp. 1183-1210

Fienberg, S., 1972, The multiple recapture census for closed populations and incomplete 2k contingency tables. Biometrika, 59, 409-439.

Gerritse, S.C., P.G.M. van der Heijden & B.F.M. Bakker, 2015, Sensitivity of population size estimation for violating assumptions in loglinear models, *Journal of Official Statistics*

Hoogteijling, E.M.J., 2002, Raming van het aantal niet in de GBA geregistreerden, Rapport 177-02-SOO (Voorburg: Centraal Bureau voor de Statistiek)

IWGDMF (International Working Group for Disease Monitoring and Forecasting), 1995, Capture- recapture and multiple record systems estimation. Part 1. History and theoretical development. *American Journal of Epidemiology*, 142, 1059-1068

Van Buuren, S., 2012*, Flexible imputation of missing data* (Boca Raton: Chapman & Hall/CRC Press)

Van der Heijden, P.G.M., J. Whittaker, M. Cruyff, B. Bakker & R. van der Vliet, 2012, People born in the Middle East but residing in the Netherlands: invariant population size estimates and the role of active and passive covariates, *Annals of Applied Statistics,* vol. 6, nr. 3, pp. 831-852