CrossMark

Estimating Homelessness in the Netherlands Using a Capture-Recapture Approach

A. M. Coumans[1] · M. Cruyff[2] · P. G. M. Van der Heijden[2,3] ·
J. Wolf[4] · H. Schmeets[1,5]

Abstract This study focuses on the homeless population in the Netherlands, as an indicator of social exclusion. By applying the capture-recapture (CRC) methodology to three registers, not only the size of the homeless population could be estimated, but also its composition in terms of gender, age, place of living, and origin could be depicted. Because of the use of three registers and the availability of background characteristics for each of the registers, the usual stringent assumptions of capture recapture methodology is circumvented. This advanced application of CRC to estimate the homeless population on the national level, has led to official figures for five subsequent reference dates (January 1st of 2009, 2010, 2011, 2012 and 2013). In 2009 the size of the total homeless population in the Netherlands was estimated at 17,767, of which 5169 were registered on one of the three lists. Between 2009 and 2012 the estimated size of the population increased, which was largely due to the financial crisis. For all reference dates, the composition of this population showed that generally more men than women were registered and that homeless people in the age category of 30–49 years old were registered more than the younger or older age groups. Compared to the general Dutch population, the homeless population includes relatively many men, many people aged 30–49 years and people with a non-western background.

✉ A. M. Coumans
m.coumans@cbs.nl

[1] Department of Social-Economic and Spatial Statistics, Statistics Netherlands, P.O. Box 4481, 6401 CZ Heerlen, The Netherlands

[2] Department of Methodology and Statistics, Faculty of Behavioural and Social Sciences, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands

[3] Southampton Statistical Sciences Research Unit, University of Southampton, Southampton, UK

[4] Onderzoekscentrum Maatschappelijke Zorg (Omz), Public Health, UMC St Radboud, Radboud University Nijmegen, Huispost 117 Omz, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

[5] Maastricht University, Maastricht, The Netherlands

1 Introduction

This paper presents a methodological illustration of a simple, cost-effective and replicable technique to estimate the homeless population on the national level. In the Netherlands, for 5 subsequent years estimates have been conducted and the results show an increase in the homeless population from 17.8 thousand in 2009 to 27.3 thousand in 2012. In 2013 no further increase was found. This trend is completely congruent with the development of the financial crisis in this time span. It is the first time that this promising technique to assess the homeless population has been successfully applied on a national level. This approach is highly recommendable for other countries as it saves time, reduces costs and yields into reliable and plausible figures about the homeless. However, adoption of this methodology is only worthwhile if some requirements are met, including the availability of national population registers of sufficient quality, and the possibility to uniquely identify people. We will outline this new methodology and put the homeless in the context of the difficult to survey groups.

Homeless people are a hard-to-reach population: in the Netherlands, for example, they are not identifiably included in the basic municipal registration and are thus excluded from sample frames and general population statistics. As a result, they are under-represented in large scale social surveys conducted by national statistical institutes. And—even if feasible—approaching the homeless for interviews using standard sample survey methods would result in excessive non-response rates and consequently in non-response bias.

For a number of reasons, it is important to know how many people in a society are homeless. First, the number of homeless people is potentially an indicator of social exclusion within a society (Coumans and Schmeets 2014; Schmeets and Coumans 2013). An indication of the size of the homeless population can play a pivotal role in determining whether the problem will be included on a government's policy agenda in the first place (Amore et al. 2011). Second, homeless people often have multiple health problems—physical, mental and substance abuse problems; large numbers of homeless—and especially roofless people may pose a threat to public health. Second, homeless people may trigger feelings of insecurity among fellow-citizens and may cause nuisance in public spaces as a result of their behaviour, public drinking and drug use and/or petty crime. And fourthly, homeless people have specific characteristics, particularly in terms of quality-of-life indicators such as health, lifestyle, and social participation. Because non-coverage of this specific group results in biased population statistics, an effort should be made to estimate the number of homeless people and their characteristics.

An accurate and reliable identification and classification of homeless people is also necessary to develop policies for different manifestations of homelessness and to monitor the effectiveness of such interventions (Amore et al. 2011). An estimate of roofless people may, for example, provide insight into the number of housing units needed for homeless people.

For most homeless people providing housing in combination with intensive support will result in a considerable increase in their quality of life and participation in society, as well as in their sense of safety (Busch-Geertsema 2013; Maas et al. 2012; Patterson et al. 2013).

Also, from a more economic perspective, housing is a first condition to take part in education and the labour market, which will also reduce the number of people dependent on social benefits. Furthermore, by decreasing the number of roofless people the quality of public life will also improve.

Notwithstanding the importance and relevance of meaningful statistics on the homeless population for informed policy-making, an accurate and reliable estimate of the size of the total Dutch homeless population has not been available for years. Although reliable and accurate estimates of homeless populations have been constructed for several Dutch municipalities in the past years (Hulschbosch et al. 2003, 2005, 2006; Reinking et al. 2001; Van Zwieten et al. 2008; Vocks et al. 2008), the most recent estimation at national level was compiled in 2003 (De Bruin et al. 2003). This number was based on extrapolated estimates of roofless and homeless people from interviews with 95 experts of 44 munic-ipalities, and quick scans in 38 smaller municipalities (De Bruin et al. 2003). Older estimates were established in 1989 and 2000 by prognostic studies (PLOTT; Heydendael 2000; Heydendael and Brouwers 1990; Heydendael and Brouwers 1989) and in 1987 by the SGBO/VNG, the umbrella organisation of Dutch municipalities (Van der Zwet et al. 1990), which mainly used extrapolation of municipal estimations. Also, the Federatie Opvang, an umbrella organisation for social services, provides yearly counts of the homeless population using their services. However, since their target population differs substantially from that of the present study, a comparison is not possible.

This study presents an advanced application of the capture-recapture approach (CRC) to produce updated official figures on homeless people in The Netherlands for five subsequent reference dates (January 1st of 2009, 2010, 2011, 2012 and 2013). For all dates the same method as well as the same updated registers were used, which makes this study pre-eminently useful for signalising trends for the whole population and subpopulations. The Netherlands is the first country adopting this procedure to produce such national estimates. Moreover, since 2009 the figures are published by Statistics Netherlands on a yearly base as the official statistics on homelessness.

This paper also illustrates how the CRC approach can be applied to assess the homeless population in terms of background characteristics. By using covariates, we are able describe the estimated population by the levels of these covariates, giving insight into characteristics of individuals not included in any registers.

Thus the main aim of the current paper is to present a cost-effective strategy to estimate the homeless population on the national level, including information on various subpop-ulations. Therefore, as an illustration, data of 2009 were used. Furthermore, in order to test the robustness of the selected statistical model yielding the estimate of 2009, also the selected models and their resulting estimates of the years 2010–2013 are presented.

This paper is structured as follows. Section 2 outlines the definition of homelessness and the system of services provision for the homeless in the Netherlands. Section 3 offers a literature review on the various method for estimating homeless populations, including the CRC approach and its merits compared to other approaches. Section 4 details the registers used, as well as the CRC approach and its assumptions, and explains the process of model selection. Section 5 presents the results, and in Sect. 6 we make some concluding remarks and discuss both the application of CRC and its resulting estimates. Also, we shortly discuss some of the potential policy implications.

2 Definition of the Homeless Population in the Netherlands

How homelessness should be defined is a fundamental and persistent problem. Definitions of homelessness vary considerably across the world and only a few are based on a defined concept (Amore et al. 2011). In the Netherlands there is no broadly accepted definition of homelessness. The authorities in the Netherlands (Department of Health, Welfare and Sport, central local authorities) have not formally established a definition of homelessness for policy aims. In principle, each of the municipalities is free to determine the way in which homelessness and groups of homeless people are defined. Furthermore, continuity in handled definitions has not been guaranteed in each municipality due to policy changes after municipal elections. However, the government and local authorities increasingly acknowledge the necessity to distinguish and to assess factual homeless people, residentially homeless people and marginally housed people. In the Netherlands, this distinction was introduced in a large scale homelessness study in The Hague (Wolf et al. 2002) and was derived from homelessness research internationally (Koegel et al. 1996). In the Homelessness Action Plan (Maas and Planije 2009), implemented between 2006 and 2013 by the Dutch government together with the municipal authorities of the four largest cities (Amsterdam, The Hague, Rotterdam and Utrecht), the distinction between factually homeless people and residentially homeless people was also implemented.

Consensus also emerged in other European countries, and led to the European Typology of Homelessness and Housing Exclusion (ETHOS), which is classified on the basis of living situation. This typology, also adopted by Statistics Netherlands, consists of the four dimensions: roofless, houseless, insecure accommodation, and inadequate accommodation. The roofless and houseless dimensions together define homelessness; insecure and inadequate accommodation refer to housing exclusion (www.feantsa.org; Amore et al. 2011).

In this study we focus on roofless people, whom we define in accordance with Wolf et al. (2002) as individuals who had no permanent accommodation on the reference date. Also in line with Wolf et al. (2002), we distinguish the following categories:

- people who sleep outdoors, either in the open air or in covered public spaces such as doorways, bicycle sheds, railway stations, shopping centres or cars;
- people who spend the night indoors in transient accommodation run for the homeless, including emergency shelters;
- people who sleep indoors in the homes of friends, acquaintances or relatives, without knowing where they can sleep the following night.

In the Dutch institutional setting this population is mainly served by so called low-threshold services according to their basic needs such as a need for food, shelter and a safe haven during the day or night. These shelters also include drop-in services and night shelters. In addition, workers of outreach services make contact with this marginalized category on the streets and support them in finding services that match their needs.

Although the term *roofless* thus refers to the specific group of the literal or factual homeless, for the sake of clarity and recognisability, throughout this paper the term 'homeless' will be used. Focusing on the roofless means that this study will focus on a specific subgroup of the homeless population, i.e. those who often are publicly most visible. This also implies that our study will not provide estimates of the following houseless categories within the homeless population (see ETHOS): (1) people staying in residential services, such as a homeless hostel, temporary accommodation or transitional supported accommodation; (2) people in women's shelter; (3) people in accommodation

for immigrants; (4) people due to be released from institutions (for example, penal institutions); and (5) people receiving longer-term support (due to homelessness) in supported housing settings in which clients have their own apartment (sometimes shared with others) and receive targeted support on a regular basis from professional workers.

3 Literature Review on Size Estimations of the Homeless Population

Several methods have been developed to estimate the number of homeless people, particularly at the local level. This section offers an overview of the literature on several variants of population size estimations in the field of homelessness; it does not intend to be exhaustive. This review distinguishes the following methods for estimating or counting homeless populations: (1) link tracing methods, (2) indirect estimations, (3) single-contact censuses and their evaluation, and (4) capture-recapture methods. Although this classification reflects the existing literature on this broad subject, we are aware that the distinction is fairly rough as some methods partly overlap and can be applied simultaneously to estimate homelessness.

1. Link-tracing methods are well-known: they include all kinds of sampling techniques that use people from the target population as a starting point and ask them to nominate other members of their group, the so-called nominees. The nominees are in turn asked to do the same, and this is repeated until a predetermined point of saturation is reached (Spreen 1992; Thompson and Frank 2000). Specific examples of the use of this method among the homeless population are snowball sampling (Beata and Snijders 2002) and network sampling (Coumans and Spreen 2003; Spreen 1999; Spreen and Coumans 2001).

2. Other techniques to estimate the size of the homeless population are more indirect in nature and comprise document analysis, analysis of administrative data, key-informant interviews, or general population surveys (Benjaminsen and Dyb 2008; Collins 2010; De Bruin et al. 2003; Fitzpatrick et al. 2012; Hombs 2011; Metraux et al. 2001; Salo and Campanelli 1991; Wood et al. 2014). A less often referred to evaluative estimation technique is that of small area estimators (Hudson and Vissing 2010; Kriegler and Berk 2007, 2010). Some studies have combined several more or less evaluative approaches or sources (Collins 2010; Beata and Snijders 2002; De Bruin et al. 2003; D'Onise et al. 2007; Fitzpatrick et al. 2012; Pointing et al. 2013), which may be more valuable than opting for one single approach (D'Onise et al. 2007). General population surveys are often applied to study life-time or 5-year prevalence of homelessness (Link et al. 1994, 1995; Ringwalt et al. 1998; Tompsett et al. 2006; Toro et al. 2007).

3. A specific part of the literature concerning estimations of the homeless comprises the use of techniques to monitor and evaluate the count or enumeration of homeless people during S-night, the census in the USA and Australia (Cousineau and Ward 1992; Devine and Wright 1992; Edin 1992; Hopper 1992; Hopper et al. 2008; Stark 1992; Williams 2011; Wright and Devine 1992). Cowan et al. (1986) referred to this approach as a single-contact census. Critical observation of the enumerators, interviews with homeless people, post-count survey methods and a plant-capture[1]

---

[1] Plant-capture is a modified capture-recapture method, in which the proportion of observers or "plants" that are "captured" by the enumerators during S-night, is used to evaluate the accuracy of the count.

methodology were most common as evaluative approaches. Also ethnography is often used as an additional source of information (Salo and Campanelli 1991).

4. Finally, probably one of the best-known and most widely discussed method to estimate the number of homeless people is the capture-recapture (CRC) approach (Beata and Snijders 2002; Berry 2007; Brent 2007; Chao et al. 2001; D'Onise et al. 2007; Félix-Medina and Thompson 2004; Fisher et al. 1994; International Working Group For Disease Monitoring and Forecasting, IWGDMF 1995; Shaw et al. 1996; Van der Heijden et al. 2009; Van der Heijden et al. 2012; Williams and Cheal 2002). This estimation technique has its origins in biology and refers to the estimation of an unobserved part of a certain population, originally populations of animals. The approach has evolved into a useful technique with applications in epidemiological research and the social sciences. The methodology has proven to be especially useful for estimating hidden populations, such as drug users and homeless people. In the field of homelessness, there are two main variants of CRC applications: (1) using two or more independent observations of the study population in their own environment, and (2) using incomplete lists with members of the study population. In both cases, assuming we have two observers or two lists, the unobserved part of the population is calculated by using the ratio between individuals observed or registered only once and individuals observed by both observers or registered on both lists. In this study, we applied CRC of linked administrative sources, which will be elaborated in the next section.

4 Methods

4.1 CRC of Linked Administrative Sources

Compared with many of the approaches discussed above, CRC of linked administrative sources has the advantage that it is cost-effective for a statistical bureau in need of a national estimate of the number of homeless. Many of the above approaches may be used successfully to provide local or regional estimates, but even at this level are time-consuming and demanding. Also it is difficult, if not impossible, to carry out a series of local surveys and generalise these to a national estimate. Another major advantage is that this approach deals with incomplete lists, which is often an evident problem using registers of this population.

However, CRC of linked administrative sources assumes that the administrative sources are of high quality. If the number of sources to be linked is two, there is also the assumption that inclusion of a homeless person in one source is independent of him/her being registered in the other. However, as we shall discuss in more detail in Sect. 4.3, if the number of sources to be linked is larger than two, this restrictive independence assumption is not necessary to find unbiased estimators, but is replaced by the less restrictive assumption that, in the case of k registers, the so-called k-factor interaction is absent. Another assumption is that the population is closed, and that there are no problems in linking the individuals in the different registers to each other. For this latter assumption each of the registers must contain the relevant information to do this, but also privacy regulations may not obstruct the linking of individuals from different registers. For the Netherlands, each of the registers used contains a key variable as well as detailed information on the homeless, which is derived from the same official source, the national population register. By law, Statistics Netherlands is entitled to link the registers used.

4.2 Data and Selections

For all reference dates, i.e. January 1 of 2009–2013 the estimations were based on the same three data sources, i.e.:

1. A list compiled especially for this study, consisting of individuals who, according to the official Dutch population registration, the basic municipal administration system (GBA), are residing at an address where a low-threshold service (shelter) for the homeless is located. Homeless people can use these services as a postal address, in order to receive a social benefit. To identify these individuals uniquely, we matched a list of shelter addresses with the GBA on the reference date. The addresses were provided yearly by the main umbrella organisations offering low threshold services for the homeless, i.e. the Federatie Opvang, the Army and the Centrum voor Dienstverlening (located in Rotterdam). The coverage of the initial list was an estimated 90–95 %. In addition, each consecutive year Statistics Netherlands updated this address list by internet research and contacting institutions. This list is further referred to as Shelter.

2. A list of individuals between 18 and 65 years old who received income support (WWB), but did not have a permanent residence according to the official WWB registration on the reference date of January 1. Apart from not having a permanent address, there are some specific conditions to get WWB-support. First of all, one is supposed not to be able to support oneself, which means that someone has not enough money to pay the rent, health insurance, food, and bills for electricity, gas and light. Other criteria are that the (household) income has to be below a certain level, that the applicant has the Dutch Nationality and lives or stays in the Netherlands, and has a dwelling way of life. People who are (expectantly) temporarily homeless, or residentially homeless are excluded from this arrangement. Over the years, the quality of this register is consistent. This list is called WWB below.

3. A list of individuals registered as being homeless in the National Alcohol and Drugs Information System (Ladis; source: Stichting Informatie Voorziening Zorg, SIVZ). This list is further referred to as Ladis. It has to be noted that the housing situation of individuals was recorded at the moment of their registration in the institute. This means that possibly their housing situation had been changed between registration and our reference date from homeless to having a residence or the other way around. Unfortunately, it was not possible to determine the exact size of this potential bias to our estimates.

Each person on all three registers was assigned a unique identification number based on the same Dutch population registration, the GBA, which aims to register all inhabitants in the Netherlands. This number is assigned to every citizen at birth or in the case of permanent residence. Moreover, we were also able to link for each homeless person the officially registered background characteristics according to the GBA.

Within these lists, further selections were made which are important for correct interpretation of the data. This study focuses on homeless people aged between 18 and 65. This selection was made as by definition people who receive income support according to the WWB are between 18 and 65 years (see also point 2 above). Second, the two other lists comprise hardly any people younger than 18 or older than 65 years. For the youngest category this is mainly due to Dutch policy on aid for the homeless, which has a separate circuit for homeless youngsters. Older people generally have changed street life for living in residential services, such as homeless hostels, or they stay in nursing homes for

homeless people. Therefore, their number is not expected to be substantial among the factual homeless group.

Lastly, people registered at addresses at which more than one kind of service was located were not selected for this study. Otherwise the list would be biased as it would also include residentially homeless people or people living in women's refuges. After all, we are only interested in estimating the size of the actually roofless population. The only exception to this was when people registered at these addresses were also registered as receiving income support according to the WWB. As income support is only provided under strict conditions (see above), it is plausible to assume that these people were actually roofless.

In the analysis the variables Shelter (S), WWB (W) and Ladis (L) were used to indicate the three lists.

Gender (G), age (A; 18–29/30–49/50–64 years), place of living (P; in one of the four largest cities,[2] i.e. Den Hague, Amsterdam, Rotterdam and Utrecht vs. other) and origin (O; native Dutch/western foreign or non-western foreign) were used as covariates that enable description of the estimated population in terms of these variables (see Sect. 5.2). For the coding scheme see Table 5 of appendix.

Table 1 presents the cross-classification of the variables Shelter, WWB and Ladis. The total number of identified homeless is 5169. The cell (no, no, no) is empty by definition and its value has to be estimated. Subsequently, the identified 5169 homeless added to this estimate will yield an estimate of the total population size.

4.3 CRC and the Model Selection Process

A well-known CRC technique for estimating the size of a population is to select two or more registers of this population, link the individuals in the registers, and estimate the number of individuals that occur in neither register (Bishop et al. 1975; Fienberg 1972; Cormack 1989; International Working Group for Disease Monitoring and Forecasting, IWGDMF 1995; Van der Heijden et al. 2012). For example, with two registers A and B, linkage gives a count of individuals in A but not in B, a count of individuals in B but not in A, and a count of individuals both in A and B. These counts form a contingency table denoted by A × B with the variable labeled A short for 'inclusion in register A', taking the levels 'yes' and 'no', and likewise for register B. In this table the cell 'no, no' has a zero count by definition, and the statistical problem is to estimate this value in the population. A population size estimate is obtained by adding this estimated count of missed individuals to the counts of individuals found in at least one of the registers.

The frequency of the missing 'no, no' cell can be obtained by fitting a log-linear model to the incomplete contingency table. Log-linear models describe (the logarithm of) the observed cell frequencies of the contingency table in terms of main and interaction effects of the variables in the model. To distinguish between different log-linear models we use the notation by Bishop et al. (1975). In this notation, variables that interact with each other are enclosed by a single set of square brackets, while variables that do not interact are enclosed by different sets of square brackets.

For example, consider a 2 × 2 contingency table of the registers A and B. The log-linear model [AB] for these two registers is given by

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_{ab}^{AB}, \tag{1}$$

Table 1 Cross-classification of homeless included in Shelter, WWB and Ladis

Shelter		WWB			
		Yes		No	
		Ladis		Ladis	
		Yes	No	Yes	No
	Yes	30	495	24	999
	No	175	2,792	654	-

where m_{ab} denotes the expected frequency of the cell ab, for a, $b = \{$'yes', 'no'$\}$. The parameter λ denotes the intercept, the parameters λ_a^A and λ_b^B the main effects of A and B, and the parameter λ_{ab}^{AB} the interaction effect of A and B. The inclusion of λ_{ab}^{AB} in the model implies that the probability to be in A depends on whether the subject is in B or not, and vice versa. The model is called *saturated* since it has as many parameters as cell frequencies. However, because the cell $m_{no,no}$ is not observed, model [AB] has one parameter too many, and can therefore not be estimated.

The *independence* model [A][B] on the other hand, as given by

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B, \tag{2}$$

has only three parameters, and the absence of the interaction parameter λ_{ab}^{AB} implies that the inclusion probabilities of the registers A and B are assumed to be independent. Note that for a 2×2 contingency table with one unobserved cell the model [A][B] is actually a saturated model, since it has exactly as many parameter as observed cell frequencies. After fitting this model to the three observed cell frequencies, we can use the parameter estimates to derive an estimate of the frequency of the missing 'no, no' cell, and hence of the total population size.

Aside to the independence assumption, the population size estimate is derived under the assumptions that in at least one of the two registers the inclusion probabilities are homogeneous (see Chao et al. 2001; Zwane et al. 2004), that the population is closed and that it is possible to perfectly link the individuals in registers A and B. However, it is generally agreed that these assumptions are unlikely to hold for human populations. We discuss two approaches that may be adopted to make the impact of possible violations less severe.

The first approach is to include covariates in the model, in particular covariates whose levels have heterogeneous inclusion probabilities for both registers (Baker 1990; see Bishop et al. 1975). For example, with the covariate X we can then expand the two-way contingency table to a three-way contingency table, and fit a log-linear model [AX][BX], as given by

$$\log m_{abx} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_x^X + \lambda_{ax}^{AX} + \lambda_{bx}^{BX}, \tag{3}$$

where the two-factor interaction parameters λ_{ax}^{AX} and λ_{bx}^{BX} denote interactions between the covariant X and the registers A and B, respectively. The restrictive independence assumption is replaced by a less restrictive assumption of independence of A and B conditional on the covariates; and sub-population size estimates are derived (one for every level of the covariates) that add up to the total population size estimate.

The second approach is to include a third register C, and to analyze the three-way contingency table with log-linear models that may include one or more two-factor interactions. The saturated model is then given

$$\log m_{abc} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{bc}^{BC}, \tag{4}$$

which in short-hand notation is expressed by [AB][AC][BC], which is the commonly used and more comprehensible notation for log linear modeling.

Note that this model allows for the pairwise dependence of registers but—given the absence of the parameter λ_{abc}^{ABC}—not for the three-factor interaction. Unfortunately, it is not always possible to include a third register, either because such a register is not available, or because there is no information that enables linkage of individuals in the third register to both other registers.

In this study, we dispose over a third register as well as a number of covariates, and the assumptions underlying population size estimation can therefore be substantially relaxed. Since we have three registers, we can test model pairwise dependencies between the registers by including the interaction terms λ_{ab}^{AB}, λ_{ac}^{AC} and λ_{bc}^{BC}, and testing whether these are significant. Furthermore, since we have covariates, we no longer need to assume homogeneity of inclusion probabilities. For example, we can investigate whether the inclusion probabilities of the registers Shelter and Ladis are independent, and whether the inclusion probabilities for males in each of the three registers differs from those for females. As noted above, by using covariates we are also able to give insight into characteristics of individuals not included in any of the registers.

For model selection we follow the standard approach in log-linear modeling. The goodness-of-fit of a model is evaluated by the deviance. The deviance compares the fit of the fitted model against that of the saturated model, and a non-significant deviance indicates that the model fits adequately. The deviance is Chi square distributed with degrees of freedom equal to the number of observed cell frequencies minus the number of estimated parameters. The relative fit of a model is evaluated by the AIC, which is the standard measure of fit for this kind of model testing. The AIC allows for the comparison of non-nested models; the model with the lowest AIC is to be preferred (for an example in the context of population size estimation, see Burnham and Anderson 2002).

Although the availability of the three register variables S, W and L and the four covariates G, A, P and O allows for substantial relaxation of the assumptions, it severely complicates model selection. With seven variables the number of potential log-linear models is extremely large. In combination with the absence of a definite theoretical explanation, the use of some exploratory model selection procedure is indispensable. To this end we use the STEP function of the statistical software package R (see Hastie and Pregibon 1992; Venables and Ripley 2002). Note that we use the coding scheme as introduced in Sect. 4.2 to indicate the variables in our model (see also Table 5 of appendix).

The STEP function searches for the best model by starting with independence model [S][W][L][G][A][P][O], and adding and deleting parameters in a stepwise fashion on the basis of the AIC, with the restriction that the saturated model is [SWGAPO][SLGA-PO][WLGAPO] and not [SWLGAPO]. It is well known, however, that the use model selection procedures as the STEP procedure are no guarantee for finding the model with the lowest AIC. In an attempt to further improve the AIC, we performed an ad hoc stepwise model search on the basis of the model selected by STEP. The same procedure to optimize the model was used for model selection in 2010–2013. The resulting model was used to obtain a point estimate estimation of the population size, and a 95 % confidence interval of

the population size estimate was based on the parametric bootstrap (compare Van der Heijden et al. 2012).

5 Results

5.1 Model Selection

Table 2 shows the results of our model search. Model 1 is the independence model. As expected it does not fit the data, with a deviance statistic $X^2 = 1653, df = 242, p < .001$. Model 2 is the model selected by the STEP procedure. It fits the data well, $X^2 = 175, df = 198, p = .88$, but yields an unrealistically high population size estimate of 31,415 and an excessively large confidence interval, ranging from 23,682 to 47,948. In an effort to explain this awkward result, we found that the interaction term [WL] is largely responsible for the high estimate. We suspect that the interaction term [WL] is either due to sampling error, or that it is part of the more general interaction term [SWL]. In the former case, we would expect that there is a model without the term [WL] that fits the data almost equally well. In the latter case, the omission of the three-factor interaction parameter λ_{swl}^{SWL} might have resulted in a biased estimate. Unfortunately, this assumption cannot be verified, since the parameter λ_{swl}^{SWL} cannot be estimated due to the missing cell frequency of not being in any of the three registries.

An alternative model was found by rerunning the STEP procedure with the restriction that the selected models should not include the interaction term [WL]. This resulted in model 3 of Table 2. This model also fits the data well, $X^2 = 180, df = 196, p = .79$, and yields a population size estimate of 17,584. Our own model search entailed some small adjustments to Model 3. The most important refinements of Model 3 include the introduction of the four-factor interaction terms [SWGP], [WOGP] and [WGAP] and the three-factor interaction terms [LGA] and [GAO], and the elimination of the three-factor interaction [SGA]. The resulting model 4 has more parameters than model 3, but is to be preferred due to the lower AIC. It yields a slightly higher population size estimate of 17,767. Its AIC is also very close to that of the rejected model 2 (285.6 vs. 282.8), which seems to support our suggestion that the interaction term [WL] in model 3 might be due to sampling error.

Table 3 presents the selected models, the estimates and their associated confidence intervals for reference dates of January 1st 2010–2013 using the same method as well as the same updated registers. These results show that the models of the subsequent years are very similar to the finally selected model of the year 2009 (model 4, Table 2). The main differences are that in 2010 and 2011 the interaction term [GAP]—indicating the interaction between gender, age and place of living—was included, and in 2012 and 2013 the interaction between origin and place of living—as expressed by [OP]—was adopted in the

Table 2 Model selection in 2009

Model	Deviance	df	AIC	Pop. est.
1. [S][W][L][G][A][O][P]	1653	242	1672.7	12,567
2. [WGA][SGA][LAO][SWP][WAP][WOP][SO][LP][WL][GO]	175	198	282.8	31,415
3. [SWP][SGA][SLO][WOP][WAP][WGA][LAO][LP][GO]	180	196	291.6	17,584
4. [SWGP][SA][SLO][WOGP][WGAP][LGA][LAO][LP][GAO]	134	176	285.6	17,767

Table 3 Selected models in 2010–2013

Year	Model	Pop.est	CI
2010	[SWGP][SAP][SO][SL][WAP][WOP][LGA][LAO][LP][GAO][GAP]	23,344	20,519–26,078
2011	[SWGP][SAP][SO][SL][WAP][WOP][LGA][LAO][LP][GAO][GAP]	24,421	21,723–27,504
2012	[SWGP][SL][SAP][SOP][WO][WAP][LGA][LAO][LP][GAO][GAP][OP]	27,256	23,856–31,927
2013	[SWGP][SL][SAP][SOP][WO][WAP][LGA][LAO][LP][GAO][GAP][OP]	24,760	22,146–28,749

model. For the years 2010 and 2011 and for the years 2012 and 2013, the models are identical. This implies that the model for this population estimation is fairly robust, which is an important indicator for its reliability. This also supports the trustworthiness of the resulting estimations.

The estimated homeless populations has increased from about 18,000 to 27,000 in 2012. Although there seems to be a considerable increase over the years, the accompanying confidence intervals overlap for consecutive years and the only significant difference was found between 2009 and 2012.

5.2 Findings

We describe model 4, as selected for reference date of January 1 2009, in some more detail. Based on 5169 registered homeless, we estimate that there are 12,589 homeless people who are not included in any of the three registers. Together, this yields the estimate of 17,767 homeless in the Netherlands, with a 95 % confidence interval ranging from 15,601 to 21,225. In this model, two pair-wise interactions between the registers are included, i.e. between Shelter (S) and WWB (W), and between Shelter (S) and Ladis (L). The interdependence between Shelter and WWB is reflected by the four-factor interaction term [SWGP], which also includes the covariates with gender (G), and location in or outside one of the four largest cities (P). The corresponding parameter estimates (see Table 4 of appendix) reveal a positive interdependence, which is strongest for homeless women living in one of the four largest Dutch cities, and weakest for men living in one of these cities. In other words, the probability of being registered in both Shelter and WWB is highest for women in the large cities and the lowest for men in the large cities. Note that this interaction term is also consistently included in all models of the subsequent years 2010–2013.

The interdependence between Shelter (S) and Ladis (L) is reflected by the term [SLO], which also includes the covariate O, denoting a person's origin. Interpretation of the corresponding parameter estimates shows a negative dependence, which is strongest for western immigrants and weakest for non-western immigrants. In the models of 2010–2013 the pairwise interdependence of being on Ladis with being on the Shelter list as expressed by [SLO] is absent.

As log-linear parameter estimates are sometimes difficult to interpret, this section also explains some of the relevant margins of the fitted values of the table of registers by covariates, based on the data and model of 2009.

As an illustration, Fig. 1 shows the estimated distribution of Dutch homeless people recorded on one of the three lists by the covariates used in the model, i.e. gender, age, origin and place of living. The distribution shows that most of the homeless are registered in the WWB and fewest are registered in Ladis. It seems plausible that many homeless people receive income support, and only a limited percentage receives help for alcohol and drugs addiction.

Furthermore, the findings show that more men than women are registered in Shelter and WWB, which means that those lists register males better than females. All three lists seem to register the 30–49 age category better than older and younger categories. Homeless living in one of the big four cities are better registered by WWB than those living else-where. This is true to a lesser extent for Ladis, while Shelter shows the opposite picture: this list registers better outside 'the big four'. Lastly, compared to native Dutch homeless and western immigrants, non-western immigrants are best registered by the WWB. For the other two lists there are no or only slight differences in the extent to which the three groups are registered.
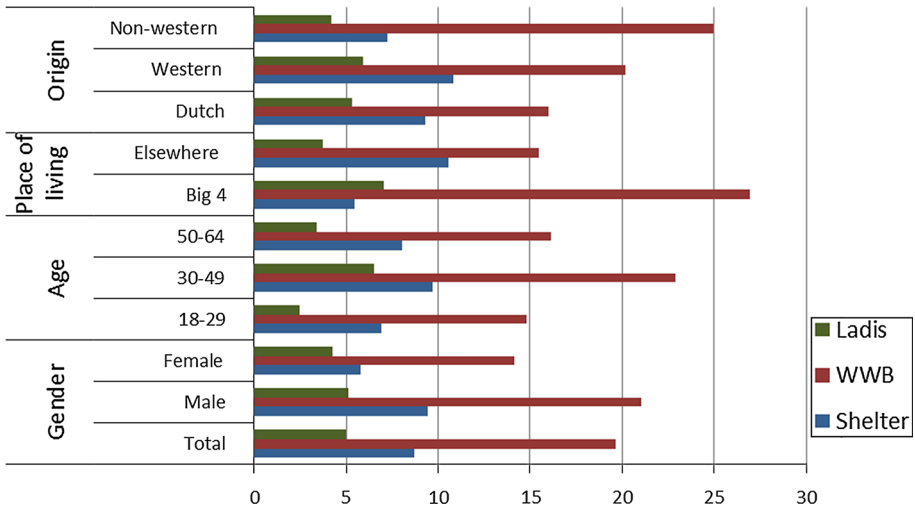
Fig. 1 Estimated inclusion probabilities on one of the three list by gender, age, origin and place of living, 2009 (CBS/SIVZ). Note that these values are fitted values using the log-linear modelling estimates
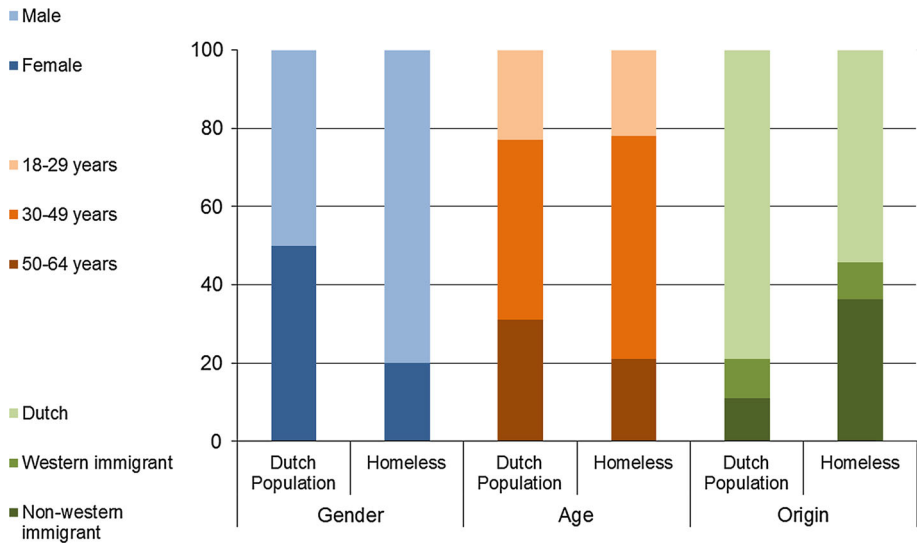

Fig. 2 Estimated homeless 18–64 aged population by gender, age, and country of origin, 2009 (CBS/SIVZ). Note that these values are fitted values using the log-linear modelling estimates

The homeless population can be characterised by the covariates included in our model. Figure 2 shows the percentages of the homeless population compared to those of the general Dutch population at the same reference date of January 1.
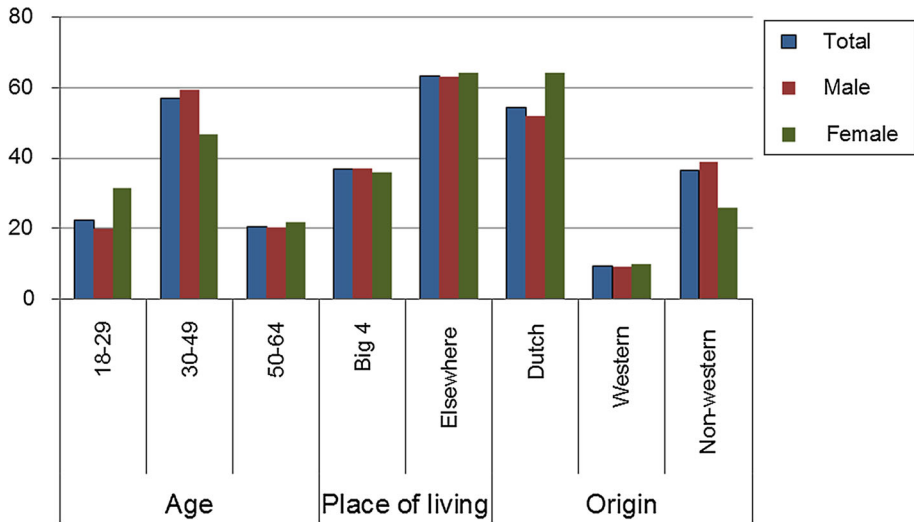
Fig. 3 Estimated homeless population, gender by age, place of living and origin, 2009 (CBS/SIVZ). Note that these values are fitted values using the log-linear modelling estimates

Figure 2 clearly shows that, compared to the total Dutch population, males are largely overrepresented. Also people aged 30–49 years are overrepresented at the expense of the youngest age category. The category with a non-western background is over three times as large as that of the total Dutch population.

If we compare the background characteristics of homeless males and females, the results show that about 30 % of the women are between 18 and 29 years old, compared with one-fifth of homeless men (Fig. 3). Relatively many homeless men are aged between 30 and 49 years: 59 %, compared with 47 % of homeless women. Homeless women are more likely to be native Dutch than homeless men (39 % against 26 %); men account for most of the non-western homeless. There are no remarkable differences between male and female homeless in terms of their place of living: of both sexes 37 % lives in the big four cities.

Comparing the homeless living in one of the four large cities with those who live elsewhere, no differences between gender and almost none between age categories reveal (Fig. 4).

The place of living strongly differentiates for ethnic background of the homeless: 58 % of homeless people in the four large cities have a non-western foreign background, while this is only around a quarter for homeless people living elsewhere.

The distributions for the subpopulations in 2010–2013 are quite similar to the distribution found in 2009 (see also Statline: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=80799ned&D1=0-1&D2=a&D3=a&VW=T).

6 Discussion

As recent studies have shown the Netherlands to be a high-trust country with high levels of social capital (Schmeets and Te Riele 2014), 18 in every 10 thousand inhabitants being completely deprived from any societal rights and services could be a rather high figure. Therefore, the question may arise whether this figure is plausible for a relatively small country such as the Netherlands.
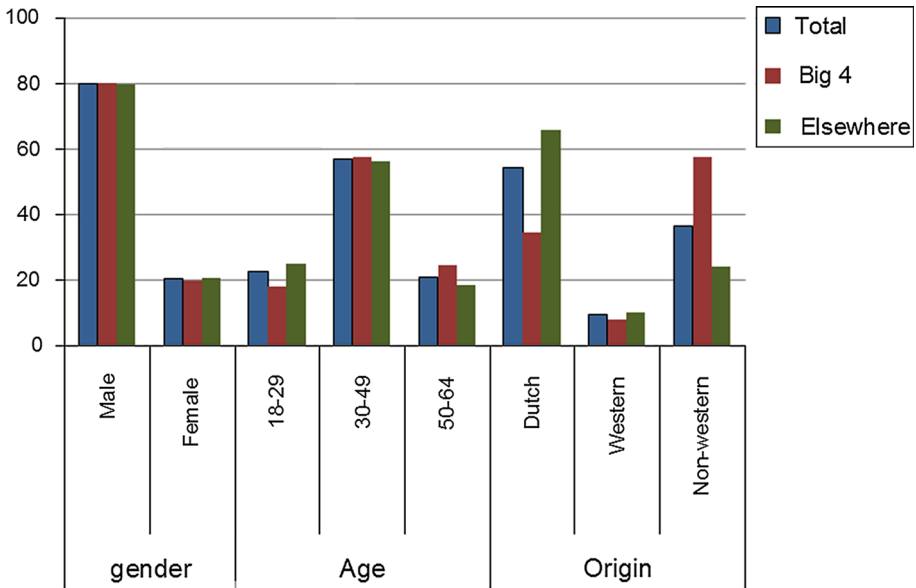
Fig. 4 Estimated homeless population, location by gender, age, and origin, 2009 (CBS/SIVZ). Note that these values are fitted values using the log-linear modelling estimates

Taking into account a confidence interval (CI) ranging from 15,601 to 21,225 our estimate does not seem to deviate substantially from de study of De Bruin et al. (2003), who estimated the homeless population at about 15,200 people, with a CI from 12,500 to 18,000. On the other hand, our estimate of 17,767 homeless people is lower compared to the estimate of Heydendael and Brouwers (1989), which was 30,000 and ranged from 26,000 to 34,000 roofless and homeless. To a lesser extent the same can be said of the estimate of Van der Zwet et al. (1990), which was 20,000, ranging from 17,500 to 34,000. However, both studies were conducted some 30 years ago, a completely different time frame. Furthermore, there are other explanations for why the figures do not match. The first is that each study used different definitions and focused on different categories of homeless people. Our estimation also includes people who stay with friends and relatives, while the study of De Bruin et al. (2003) did not cover this homeless category. The other way around, our study did not estimate the residentially homeless, while theirs did include some residential homeless categories. The studies by Heydendael (1989) and Van der Zwet et al. (1990) also covered some residentially homeless groups. The second reason for the different estimates is the methods used: extrapolation, in particular, may lead to overestimation, especially for smaller municipalities. Van der Zwet et al. (1990) acknowledged this problem in their research report. Overall, these differences in definitions as well as methods seriously complicate a direct comparison.

There seems to be a considerable increase of the homeless population over the years. However, as noted before, the accompanying confidence intervals overlap for consecutive years and the only significant difference found is between the 2009 and 2012 results. Furthermore, it is important to note that the estimate for 2009 is not in all respects comparable to those for 2010–2013, mainly because the quality of all three registrations has increased in the course of the years, especially of the Shelter register and Ladis. The Shelter list has been improved by yearly updating the original list of institutions and the quality of Ladis by including more people who were assigned a personal identification number and could therefore be linked

to the GBA and thus be included in our estimation. This differences in quality complicate the interpretation of the apparent rise of this population since 2009. However, in view of the financial crisis an increase in this population would not be unrealistic. Experts from the largest relevant organisations confirm that for a few years now, they have also observed an increase in the number of certain homeless people, such as homeless people who visit night shelters or crisis centres. Apart from that, they also mention the emergence of a new category of homeless people, the "new poor": people who, despite their low income, were just able to make ends meet, but got into difficulty when they lost their jobs or their social benefits, and consequently became homeless. Although it is hard to quantify the causes of the increase of the homeless population into exact percentages, we suspect that most part of the increase is due to the crisis and that a smaller share is caused by the improved data quality. The fact that we used the same data sources as well as the same methodology over the years, and that this yielded similar or even identical statistical models, supports our confidence that the increase was not an artefact.

To explore the plausibility of our estimate further, we should compare this figure with the findings of similar European countries. However, such a comparison is not feasible. Edgar (2009) wrote an extensive review of homelessness in Europe in 2009, including an "Appendix" with the numbers of homeless people according to the ETHOS typology. For some European countries, the number of homeless—and more specifically roofless—people is not available, or at least not available for all the categories included in this study, i.e. people sleeping rough, people in emergency accommodation, and people temporarily living with family/friends. And for countries that do provide this information, it is still questionable whether their figures are comparable with those for the Netherlands. Several countries, for example, provide figures only for some main cities (e.g. Belgium, France, Hungary, Czech Republic, Portugal, Italy). For some countries the figures were based on available bed spaces (England and Scotland), others were based on survey data (Spain), period prevalence data (Germany and Latvia); and for some countries stock data were used for the estimations. Apart from this, it is important to note that the present study also includes people without a home who are temporarily staying with friends or relatives in the roofless population, while this is not the case in the ETHOS typology. In addition to the review by Edgar (2009), Benjaminsen and Dyb (2008) estimated the homeless population in Denmark, Norway and Sweden using surveys of services and organisations in the field of homelessness. Their estimates indicate homeless rates of 1.3 per 1000 inhabitants in Sweden, 1.2 in Norway and in 1.0 in Denmark. However, these estimates are based on the whole population, and taking into account that among youngsters as well as among elderly the homeless rates might be substantially lower, such rates will very probably be higher for the 18–65 aged. Consequently, compared to these rates found in the Scandinavian studies, the rate of 1.8 per 1000 inhabitants in the Netherlands for the 18–65 years old group is fairly high, but does not seem to be out of range. Apart from the plausibility of the estimate itself, it is also relevant to compare the characteristics of the estimated population with those of other studies. Findings show that four out of five homeless people are men, over half are aged between 30 and 50 years, and more than one-third (37 %) are of non-western origin. This profile is consistent over the years and also in line with that of other studies about Dutch homeless populations. De Bruin et al. (2003) found that 87 % of the people they met during their street survey were men, most of them were between 25 and 55 years old and 41 % were not born in the Netherlands. Also, recent local studies in several Dutch municipalities show largely the same profile (Hulschbosch et al. 2003, 2005, 2006; Reinking et al. 2001; Van Zwieten et al. 2008; Vocks et al. 2008). Furthermore, 37 % of the homeless population are estimated to live in one of the largest cities. This matches exactly the proportion found by De Bruin et al. (2003). The absolute figures are also similar: De Bruin et al. estimated that 5667 of homeless people lived in one of the four largest cities, while our findings show 5818.

One last remark we should make about the plausibility of the outcome is that our study focused on the population of literally homeless people aged between 18 and 65 years. As mentioned before, the Dutch homeless policy has a separate focus on homeless young people; most older people have left the streets and live in nursing homes or other accommodation for older formerly homeless people. Instead of roofless, older people often become residential homeless. In 2003, De Bruin et al. reported that in their street sample survey they met almost no homeless people younger than 18 years and the number of homeless aged 65 or older was also almost zero. However, as the younger and older homeless populations are particularly relevant for policy, future research should also attempt to estimate the size of these populations. The same is true for other subgroups this study did not take into account, such as people staying in hostels, temporary or transitional sheltered accommodation or women's shelters, people in accommodation for immigrants, people due to be released from penal institutions and people receiving long-term support.

Since the plausibility of the results depends also on the applied methodology, a discussion to the chosen approach is indispensable. Over the last decades the CRC methodology has been used and further developed for application in a wide range of population size estimations. A recent approach proposed by Böhning et al. (2013) has extended the Chao-estimator in order to take covariate information into account. The Chao estimator uses data stemming from one single source in which people can appear multiple times. For example, in the context of the homeless, this could be a sleeping address for homeless, that registers over a limited time period how many homeless appear once at this address, how many appear twice and so on. The estimator then calculates the number of homeless that appeared zero times at this address and by adding this estimate to the number of homeless that did appear at this address, a population size estimate of the homeless population is obtained. Although this approach is fruitful for obtaining local estimates of homeless, for example in a city or in a region, for a country it is difficult to obtain this type of data.

Therefore, in this study a different approach was used, which provides some important insights into how to estimate a hidden population at the national level. First, it yielded an recent estimate at the national level, while this has not been available for years. Second, periodically applying the same methodology to the same registers made it possible to show trends and developments over time. Third, the CRC as applied to registers is efficient in terms of both costs and time. Fourthly, this approach provides valuable insights into characteristics of the homeless population in the Netherlands, including the part of that population missed by the three registers. Lastly, the unique opportunity to use the Dutch official municipal population registration (GBA) in combination with the availability of three national registers of fairly high to very high quality increased the accuracy of the findings.

Compared to estimations based on two registers, the current estimation makes use of assumptions that are not very restrictive. First, we assume that the three-factor interaction between the registers is absent. This is difficult to verify, yet it is much less restrictive than assuming that inclusion probabilities of pairs of registers are independent. Second, it is unlikely that the covariates in the model explain all the heterogeneity of the inclusion probabilities. However, it is uncertain whether adding covariates would lead to a more satisfactory solution, as this would increase the number of parameters exponentially, together with the risk of unstable and non-informative solutions. Therefore, we do not expect the probability of being included in the municipal population registration (GBA) for the people living in "mismatched" shelters to deviate substantially from that for the people living in the successfully matched locations.

Another assumption is that the population is closed, i.e. that the population did not change during the period of observation. One way to comply with this assumption is to keep the period of

observation as short as possible. As we used one reference date in this study for each of the three registers (1 January of each year), this first assumption is met. Note that the chosen reference date itself is not relevant for the estimate as long as it is the same for all registers as well as for the date of their linkage to the municipal population registration. However, strictly speaking, the size of the population estimated consequently refers only to this date. This type of estimate, which is based on stock data, has been criticised because it does not take into account duration of homelessness or transitions into and out of homelessness and therefore does not give any information about the persistence and complexity of the problem (Hulchanski 2000). Also, using stock data may beg the question of whether all people on the lists are really roofless. Acknowledging the fact that homelessness is a dynamic process, this study intends to estimate the size of the homeless population at one moment in time, and the results do not give any information about routes into or out of homelessness. However, taking into account the conditions and criteria to be included on the WWB and the Shelter list, the probability that someone is registered erroneously or for a very short period of time is small. Inclusion on Shelter or WWB almost certainly indicates rooflessness, and thus deprivation in several areas, for a longer period of time. For example, a person will not be entitled to income support for only a few days, and will not use the address of a shelter as an official postal address unnecessarily or incidentally. For the Shelter register, we checked the influence of different selections of institutions on the estimates by including and excluding certain locations according to the potentially degree of bias. Therefore, we repeated all estimates with different selections of locations, i.e. (1) including all low threshold services; (2) Including low threshold services who according to a closer investigation, offer shelter to the literal homeless; (3) Excluding all locations who offer shelter to more than one single target population, regardless whether they are literally homeless or not. This check showed that the effect of using other selections appeared limited. As noted before, Ladis may be biased by the inclusion of people who are no longer homeless on the reference date or became homeless between the moment of registration and the reference date. In order to quantify the effect of this potential bias to our estimates, an extra effort should be made to ascertain the exact size of this shortcoming. However, this is only feasible in close cooperation with the institution that gathered the data and provided this list, the Stichting Informatie Voorziening Zorg (SIVZ).

The last assumption made to the CRC approach is that the linkage of the registers is perfect. In this study we used identification numbers based on the numbers in the municipal population registration, which aims to register all inhabitants in the Netherlands. An unique identification number is assigned to every citizen at birth or in the case of permanent residence. These numbers were used for all three sources, and as they are unique for each inhabitant of the Netherlands, data linkage can be assumed to be nearly perfect. We could not link all addresses of low threshold shelters to the addresses in the GBA. This meant that the location of the service was not registered as an address in the GBA register and thus not officially used for a place of residence. As the inclusion probability of people living at successfully matched addresses may differ from that of people at mismatched addresses, we checked the effect of removing some shelters from the address list. These analyses showed that the removal of these shelters or agencies from the list did not substantially affect the estimation.

7 Conclusion

For years, no accurate and reliable estimate of the size of the Dutch homeless population at the national level has been available. Even the Homelessness Action Plan that the Dutch cabinet and the four largest cities have implemented in recent years was based on

guesstimates of local homeless populations. This study has shown that with relatively little extra effort, an accurate, reliable and replicable estimate of the size of the homeless population can be calculated, as well as a description of its characteristics.

According to the estimation presented in this study, 17.8 thousand homeless people aged 18–64 years old lived in the Netherlands on 1 January 2009. This corresponds to almost 18 homeless people per 10 thousand inhabitants in this age category at that date. To put this in some perspective, 4.2 % of the Dutch population are socially excluded, i.e. deprived of social contacts, social values, and basic rights and materially deprived (Coumans and Schmeets 2014; Schmeets and Coumans 2013). This figure is based on a general population survey, in which—as already mentioned—hardly any homeless people participated. It could be argued, therefore, that this figure should be raised to 4.4 % by adding the roofless population. We discussed the plausibility of our results and provided some methodological remarks. We also highlighted the potential value of the findings for policy.

Al in all, the discussion justifies the two-folded conclusion that (1) applying the CRC approach to three registers and covariates has proven to be a worthwhile venture, and (2) that the resulting estimates are plausible both in the Dutch context and in the actual timeframe. This can serve as an indispensable basis for informed policy-making. These statistics not only provide insight into which specific services and interventions are necessary to meet the basic needs of the current homeless population, but also shed light on systematic longitudinal measures necessary to prevent circumstances that increase the risk of homelessness: poverty, social exclusion and non-participation (unemployment, social isolation etc.), and lack of affordable housing. The relevance of information on the characteristics of homeless people may not be underestimated, as these statistics also point to factors at the individual level, such as non-western background, that may increase the risk for social exclusion and homelessness (Coumans and Schmeets 2014).

Appendix

See Tables 4 and 5.

Table 4 Parameter estimates model 4

	Estimate	SE	Pr(> |z|)		Estimate	SE	Pr(> |z|)
(Intercept)	−2.05	0.35	0.00	G2:H2	0.79	0.61	0.20
G2	2.93	0.25	0.00	G2:H3	−0.09	0.31	0.77
W2	0.33	0.25	0.18	W2:S2	0.32	0.49	0.51
Z2	3.23	0.29	0.00	Z2:S2	0.22	0.36	0.54
S2	−1.06	0.46	0.02	S2:B2	−0.70	0.37	0.06
L2	2.92	0.29	0.00	W2:H2:B2	−0.25	0.24	0.30
L3	1.09	0.35	0.00	W2:H3:B2	−0.62	0.16	0.00
H2	−1.47	0.69	0.03	W2:L2:B2	−0.68	0.22	0.00

Table 4 continued

	Estimate	SE	Pr(> |z|)		Estimate	SE	Pr(> |z|)
H3	0.70	0.39	0.08	W2:L3:B2	−1.10	0.26	0.00
B2	1.08	0.21	0.00	G2:W2:B2	0.71	0.22	0.00
H2:B2	−0.40	0.14	0.00	Z2:L2:H2	0.89	0.37	0.02
H3:B2	−1.18	0.08	0.00	Z2:L3:H2	0.51	0.45	0.26
G2:B2	−1.22	0.12	0.00	Z2:L2:H3	−0.11	0.27	0.67
W2:H2	0.04	0.23	0.87	Z2:L3:H3	−0.53	0.33	0.11
W2:H3	0.01	0.15	0.96	G2:Z2:H2	−0.98	0.63	0.12
G2:W2	0.40	0.16	0.01	G2:Z2:H3	0.27	0.33	0.40
S2:H2	−0.37	0.41	0.37	W2:S2:L2	−0.72	0.43	0.09
S2:H3	−0.83	0.23	0.00	W2:S2:L3	−1.08	0.60	0.07
L2:H2	−0.70	0.34	0.04	Z2:S2:L2	−0.19	0.33	0.56
L3:H2	−0.44	0.42	0.29	Z2:S2:L3	0.84	0.49	0.08
L2:H3	−0.46	0.26	0.08	G2:W2:S2	1.08	0.43	0.01
L3:H3	−0.57	0.32	0.07	G2:S2:B2	0.49	0.31	0.12
Z2:L2	−0.92	0.21	0.00	W2:S2:B2	0.04	0.56	0.94
Z2:L3	−0.12	0.26	0.66	W2:S2:H2	−1.61	0.59	0.01
L2:B2	−0.19	0.11	0.07	W2:S2:H3	−0.23	0.30	0.45
L3:B2	−0.50	0.13	0.00	S2:H2:B2	−0.69	0.32	0.03
S2:L2	−0.58	0.39	0.13	S2:H3:B2	−0.18	0.22	0.42
S2:L3	−1.70	0.55	0.00	S2:L2:B2	0.38	0.26	0.14
W2:L2	0.05	0.22	0.81	S2:L3:B2	0.39	0.33	0.25
W2:L3	0.66	0.26	0.01	S2:L2:H2	0.83	0.40	0.04
G2:S2	0.00	0.25	1.00	S2:L3:H2	0.90	0.46	0.05
G2:Z2	−0.36	0.21	0.08	S2:L2:H3	0.19	0.23	0.40
Z2:H2	0.19	0.69	0.78	S2:L3:H3	0.44	0.32	0.16
Z2:H3	0.47	0.39	0.23	G2:W2:S2:B2	−1.15	0.46	0.01
Z2:B2	0.82	0.15	0.00	W2:S2:H2:B2	2.20	0.67	0.00
W2:B2	1.01	0.26	0.00	W2:S2:H3:B2	0.26	0.44	0.56
G2:L2	−0.27	0.13	0.03	W2:S2:L2:B2	0.42	0.47	0.38
G2:L3	−0.17	0.16	0.30	W2:S2:L3:B2	1.47	0.66	0.03

Table 5 Coding scheme of the variables used

Label	Variable	Categories and codes
S	Shelter	1 = yes, 2 = no
W	WWB	1 = yes, 2 = no
L	Ladis	1 = yes, 2 = no
G	Gender	1 = yes, 2 = no
A	Age	1 = 18-29, 2 = 30-49, 3 = 50-64
P	Place of living	1 = one of the four largest Dutch cities, 2 = other
O	Origin	1 = native Dutch, 2 = western foreign background, 3 non-western foreign background

References

Amore, K., Baker, M., & Howden-Chapman, P. (2011). The ETHOS definition and classification of homelessness: An analysis. *European Journal of Homelessness, 5*(2), 19–37.

Baker, R. (1990). Pobability estimation and information principles. *Structural Safety, 9*(2), 97–116.

Beata, D., & Snijders, T. A. B. (2002). Estimating the size of the homeless population in Budapest, Hungary. *Quality and Quantity, 36*, 291–303.

Benjaminsen, L., & Dyb, E. (2008). The effectiveness of homeless policies—Variations among Scandinavian Countries. *European Journal of Homelessness, 2*, 45–67.

Berry, B. (2007). A repeated observation approach for estimating the street homeless population. *Evaluation Review, 31*(2), 166–199.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: The MIT Press.

Böhning, D., Vidal-Diez, A., Lerdsuwansri, R., Viwatwongkasem, C., & Arnold, M. (2013). A generalization of Chao's estimator for covariate information. *Biometrics, 69*(4), 1033–1042.

Brent, B. (2007). A repeated observation approach for estimating the street homeless population. *Evaluation Review, 31*(2), 166–199.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed., Vol. XXVI). New York: Springer.

Busch-Geertsema, V. (2013). *Housing first Europe: Final report*. Brussels: Feantsa.

Chao, A., Tsay, P. K., Lin, S.-H., Shau, W.-Y., & Chao, D.-Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine, 20*, 3123–3157.

Collins, D. (2010). Homelessness in Canada and New Zealand: A comparative perspective on numbers and policy responses. *Urban Geography, 31*(7), 932–952.

Cormack, R. (1989). Log-linear models for capture-recapture. *Biometrics, 45*(2), 395–413.

Coumans, M., & Schmeets, H. (2014). The socially excluded in the Netherlands: The development of an overall index. *Social Indicators Research*. doi:10.1007/s11205-014-0707-6.

Coumans, A. M., & Spreen, M. (2003). Drug use and the role of homelessness in the process of marginalization. *Substance Use and Misuse, 38*(3–6), 311–338.

Cousineau, M. R., & Ward, T. W. (1992). An evaluation of the S-night street enumeration of the homeless in Los Angeles. *Evaluation Review, 16*(4), 389–399.

Cowan, C. D., Breakey, W. R., & Fischer, P. J. (1986). The methodology of counting the homeless. In *Proceedings of the American statistical association* (pp. 170–175).

De Bruin, D., Meijerman, C., Verbraeck, H., Braam, R., Leenders, F. & Van de Wijngaart, G. (2003). *Zwerven in de 21ste eeuw: een exploratief onderzoek naar geestelijke gezondheidsproblematiek en overlast van dak-en thuislozen in Nederland. [Dwelling in the 21st century: An explorative study for mental illness among and nuisance from roof- and homeless people in the Netherlands]*. Utrecht: CVO, Centre for Drug Research.

Devine, J. A., & Wright, J. D. (1992). Counting the homeless—S-night in New Orleans. *Evaluation Review, 16*(4), 409–417.

D'Onise, K., Wang, Y., & McDermott, R. (2007). The importance of numbers: Using capture-recapture to make the homeless count in Adelaide. *Australian Journal of Primary Health, 13*(1), 89–96.

Edgar, B. (2009). *European review of statistics on homelessness. European housing research Ltd*. Brussels: European Federation of National Organisations Working with the Homeless.

Edin, K. (1992). Counting Chicago homeless—An assesment of the census bureaus street and shelter night. *Evaluation Review, 16*(4), 365–375.

Félix-Medina, M. H., & Thompson, S. K. (2004). Combining link-tracing sampling and cluster sampling to estimate the size of hidden populations. *Journal of Official Statistics, 20*(1), 19–38.

Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika, 59*(3), 591–603.

Fisher, N., Turner, S. W., Pugh, R., & Taylor, C. (1994). Estimating numbers of homeless mentally ill people in north east Westminster by using capture-recapture analysis. *Britisch Medical Journal, 308*, 27–30.

Fitzpatrick, S., Pawson, H., Bramley, G., Wilcox, S., & Watts, B. (2012). *The homelessness monitor: England 2012*. London: Institute for Housing, Urban and Real Estate Research, Heriot-Watt University and Centre for Housing Policy, University of York.

Hastie, T. J., & Pregibon, D. (1992). Generalized linear models, Chapter 6. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S*. Edmonton, AM: Wadsworth & Brooks/Cole.

Heydendael, P. H. J. M. (2000). *Naar huis. Over territoria van mensen die niet in de pas lopen. [Going home. About territoria of people who fall out of line]*. Nijmegen: UMC St. Radboud.

Heydendael, P. H. J. M., & Brouwers, H. G. (1989). Mensen in de marge in soorten en maten. [People on the fringes in all kinds and measures]. *Tijdschrift voor Sociale Gezondheidszorg, 8,* 4–8.

Heydendael, P. H. J. M., & Brouwers, H. G. (1990). *PLOTT: Prognostisch landelijk onderzoek thuislozen en thuislozenzorg: de opbouw van het mentaal systeem. [Prognostic National Study of Homelessness: The construction of the mental system].* Nijmegen: UMC St. Radboud.

Hombs, M. E. (2011). *Modern homelessness: A reference handbook.* Santa Barbara: ABC-CLIO.

Hopper, K. (1992). Counting the homeless—S-night in New York. *Evaluation Review, 16*(4), 376–388.

Hopper, K., Shinn, M., Laska, E., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health, 98*(8), 1438.

Hudson, C. G., & Vissing, Y. M. (2010). The geography of adult homelessness in the US: Validation of state and county estimates. *Health & Place, 16*(5), 828–837.

Hulchanski, J. D. (2000). *A new Canadian Pastime? Counting homeless people.* Toronto: University of Toronto.

Hulschbosch, L., Nicholas, S., Smit, F., & Wolf, J. (2003). *Dakloos in Alkmaar: onderzoek naar omvang en kenmerken van de daklozenpopulatie. [Homeless in Alkmaar: Study to estimate the size and characteristics of the homeless population].* Utrecht: Trimbos-Institute.

Hulschbosch, L., Nicholas, S., & Wolf, J. (2005). *Dakloos in Leiden. Onderzoek naar omvang en kenmerken van de daklozenpopulatie. [Homeless in Leiden. Study of size and characteristics of the homeless population].* Utrecht: Trimbos-Institute.

Hulschbosch, L., Nicholas, S., & Wolf, J. (2006). *Omvang van de daklozenpopulatie in de stad Utrecht. Resultaten van een omvangschatting in het voorjaar van 2005. [Size of the homeless population in the city of Utrecht. Results of a size estimation in the spring of 2005].* Utrecht: Trimbos-Institute.

International Working Group For Disease Monitoring and Forecasting, IWGDMF. (1995). Capture-recapture and multiple record systems estimation. Part I. History and theoretical development. *American Journal of Epidemiology, 142*(10), 1059–1068.

Koegel, P., Burnam, M. A., & Morton, S. C. (1996). Enumerating homeless people alternative strategies and their consequences. *Evaluation Review, 20*(4), 378–403.

Kriegler, B., & Berk, R. (2007). *Small area estimation of the homeless in Los Angeles: An application.* Los Angeles: UCLA, department of statistics.

Kriegler, B., & Berk, R. (2010). Small area estimation of the homeless in Los Angeles: An application of cost-sensitive stochastic gradient boosting. *Annals of Applied Statistics, 4*(3), 1234–1255.

Link, B., Phelan, J., Bresnahan, M., Stueve, A., Moore, R., & Susser, E. (1995). Lifetime and 5-year prevalence of homelessness in the United States—New evidence on an old debate. *American Journal of Orthopsychiatry, 65*(3), 347–354.

Link, B. G., Susser, E., Stueve, A., Phelan, J., Moore, R. E., & Struening, E. (1994). Lifetime and five-year prevalence of homelessness in the United States. *American Journal of Public Health, 84*(12), 1907–1912.

Maas, M., Al Shamma, S., Altena, A., Jansen, N., & Wolf, J. (2012). *Discussie Amsterdam: Housing First. Evaluatie van de werkzaamheid. [Discussion Amsterdam: Housing first. Evaluation of its effectiveness].* Nijmegen: Onderzoekscentrum maatschappelijke zorg, UMC St Radboud Nijmegen.

Maas, M., & Planije, M. (2009). *Monitor Plan van Aanpak Maatschappelijke Opvang. Rapportage 2009: Amsterdam, Den Haag, Rotterdam en Utrecht. [The Homelessness Action Plan. Report 2009: Amsterdam, The Hague, Rotterdam and Utrecht].* Utrecht: Trimbos Institute.

Metraux, S., Culhane, D., Raphael, S., White, M., Pearson, C., Hirsch, E., et al. (2001). Assessing homeless population size through the use of emergency and transitional shelter services in 1998: Results from the analysis of administrative data from nine US jurisdictions. *Public Health Reports, 116*(4), 344.

Patterson, M., Moniruzzaman, A., Palepu, A., Zabkiewicz, D., Frankish, C. J., Krausz, M., & Somers, J. M. (2013). Housing First improves subjective quality of life among homeless adults with mental illness: 12-month findings from a randomized controlled trial in Vancouver, British Columbia. *Social Psychiatry and Psychiatric Epidemiology, 48*(8), 1245–1259.

Pointing, B., Babacan, H., & Clough, A. (2013). *Homelessness in Douglas: It's much more than not having a roof over your head.* Cairns, Queensland: James Cook University.

Reinking, D., Nicholas, S., Van Leiden, L., Van Bakel, H., Zwikker, M., & Wolf, J. (2001). *Daklozen in Den Haag: onderzoek naar omvang en kenmerken van de daklozenpopulatie. [Homeless in The Hague: Study for size and characteristics of the homeless population].* Utrecht: Trimbos-Institute.

Ringwalt, C. L., Greene, J. M., Robertson, M., & McPheeters, M. (1998). The prevalence of homelessness among adolescents in the United States. *American Journal of Public Health, 88*(9), 1325–1329.

Salo, M. T., & Campanelli, P. C. (1991). Ethnographic methods in the development of census procedures for enumerating the homeless. *Urban Anthropology, 20*(2), 127–140.

Schmeets, H., & Coumans, M. (2013). *The assessment of social cohesion and social exclusion in the Netherlands*. Paper prepared for the UNESCO workshop on measuring and assessing inclusive social policies. Paris: UNESCO.

Schmeets, H., & Te Riele, S. (2014). Declining social cohesion in The Netherlands? *Social Indicators Research, 115*(2), 791–812.

Shaw, I., Bloor, M., Cormack, R., & Williamson, H. (1996). Estimating the prevalence of hard-to-reach populations: The illustration of mark-recapture methods in the study of homelessness. *Social Policy & Administration, 30*(1), 69–85.

Spreen, M. (1992). Rare populations, hidden populations, and link-tracing designs: What and why? *Bulletin de Méthodologie Sociologique, 36*(1), 34–58.

Spreen, M. (1999). *Sampling personal network structures: Statistical inference in ego-graphs*. Ph.D. Thesis, University of Groningen.

Spreen, M., & Coumans, A. M. (2001). A note on network sampling in drug abuse research. *Connections: Official Journal of the International Network for Social Network Analysis, 24*(3), 44–51.

Stark, L. R. (1992). Counting the homeless—An assessment of S-night in Phoenix. *Evaluation Review, 16*(4), 400–408.

Thompson, S. K., & Frank, O. (2000). Model-based estimation with link-tracing sampling designs. *Survey Methodology, 26*(1), 87–98.

Tompsett, C. J., Toro, P. A., Guzicki, M., Manrique, M., & Zatakia, J. (2006). Homelessness in the United States: Assessing changes in prevalence and public opinion, 1993–2001. *American Journal of Community Psychology, 37*(1–2), 47–61.

Toro, P. A., Tompsett, C. J., Lombardo, S., Philippot, P., Nachtergael, H., Galand, B., et al. (2007). Homelessness in Europe and the United States: A comparison of prevalence and public opinion. *Journal of Social Issues, 63*(3), 505–524.

Van der Heijden, P. G. M., Whittaker, J., Cruyff, M., Bakker, B., & Van der Vliet, R. (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, 6*(3), 831–852.

Van der Heijden, P. G. M., Zwane, E., & Hessen, D. (2009). Structurally missing data problems in multiple list capture-recapture data. *Advances in Statistical Analysis, 93*(1), 5–21.

Van der Zwet, G. P., Van der Meijden, R. R., & Burgers, L. (1990). *Dak-en thuislozen. Aantallen, opvang en gemeentelijk beleid; een inventariserend onderzoek. [Roof-and homeless people. Numbers, shelter and municipally policy]*. The Hague: Vereniging voor Nederlandse Gemeenten (VNG; Council for Dutch Municipalities), afdeling Sociaal-Geografisch en Bestuurskundig Onderzoek (SGBO; department of Social-Geographic and Science of Public Administration).

Van Zwieten, M., Biesma, S., & Bieleman, B. (2008). *Monitor daklozen en harddrugsverslaafden Apeldoorn 2008. Metingen 2004–2007. [Monitor homeless and hard drug users Apeldoorn 2008. Measurements 2004–2007]*. Groningen/Rotterdam: Intraval.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Vocks, J., Mensink, C., & Wolf, J. (2008). *Omvang van de daklozenpopulatie in de regio Waterland. Resultaten van een omvangschatting in 2008. [Size of the homeless population in the region of Waterland. Results of a size estimation in 2008]*. Nijmegen: UMC St. Radboud.

Williams, J. C. (2011). "Stand up and be counted": The politics of a homeless enumeration. *Poverty & Public Policy, 3*(3), 1–27.

Williams, M., & Cheal, B. (2002). Can we measure homelessness? A critical evaluation of the method of 'capture-recapture'. *Social Research Methodology, 5*(4), 313–331.

Wolf, J., Zwikker, M., Nicholas, H., Van Bakel, H., Reinking, D., & Van Leiden, I. (2002). *Op achterstand. Een onderzoek naar mensen in de marge van Den Haag. [About deprivation. A study into marginal people in The Hague]*. Utrecht: Trimbos Institute.

Wood, G., Batterham, D., Cigdem, M., & Mallett, S. (2014). *The spatial dynamics of homelessness in Australia 2001–2011*, AHURI Final Report No. 227. Melbourne: Australian Housing and Urban Research Institute. http://www.ahuri.edu.au/publications/projects/p53027.

Wright, J. D., & Devine, J. A. (1992). Counting the homeless—the Census Bureaus S-night in 5 United States Cities. *Evaluation Review, 16*(4), 355–364.

Zwane, E. N., Van der Pal-De Bruin, K., & Van der Heijden, P. G. M. (2004). The multiple record systems estimator when registrations refer to different but overlapping populations. *Statistics in Medicine, 23*(14), 2267–2281.