WORKING PAPER

ESTIMATING THE HOMELESS POPULATION

SAMPLING STRATEGIES

Wouter Schepers & Ides Nicaise (HIVA-KU Leuven)

February 2017

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

KU LEUVEN HIVA RESEARCH INSTITUTE FOR WORK AND SOCIETY
KU LEUVEN LUCAS CARE RESEARCH & CONSULTANCY
Université de Liège
la Strada
Centre d'appui au secteur bruxellois d'aide aux sans-abri
Steunpunt thuislozenzorg Brussel

ESTIMATING THE HOMELESS POPULATION
Sampling strategies

Wouter Schepers & Ides Nicaise (HIVA-KU Leuven)

Belgian Science Policy Office
belspo

KU Leuven
HIVA RESEARCH INSTITUTE FOR WORK AND SOCIETY
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium
hiva@kuleuven.be
www.hiva.be

© 2017 HIVA KU Leuven
No part of this publication may be reproduced in any form, by mimeograph, film or any other means, without permission in writing from the author.

FOREWORD

This paper is part of the MEHOBEL project (Measuring Homelessness in Belgium), a research project funded by the Belgian Federal Service for Scientific Policy BELSPO and co-ordinated by Prof. Koen Hermans (LUCAS – KU Leuven). The research team consists of LUCAS and HIVA, (both interdisciplinary research institutes from the University of Leuven), Centre d'Etudes de l'Opinion (University of Liège) and La Strada (a resource centre for homelessness service providers in the Brussels Region) The Mehobel project has a policy-oriented and a scientific goal. On the one hand, we want to develop methodologies to measure and monitor homelessness in Belgium. On the other hand, we aim to contribute to the evidence base concerning homelessness, and more specifically, to develop innovative approaches to measure the dark number of homeless people, to produce stock and flow data and to integrate the lived experience of homeless people into the monitoring strategy.

One of the objectives of MEHOBEL is to explore the possibilities of the capture-recapture technique, which can be used at relatively low cost to estimate the overall incidence of homelessness, based on existing and partly overlapping datasets. The present paper focuses on this particular objective, while also discussing some alternative techniques. Unfortunately, the datasets that are currently available in Belgium are fragmented and present some major design flaws. We put forward concrete suggestions for the improvement of data collection in the future, hoping to pave the way for feasible statistical series on homelessness.

The authors are grateful to BELSPO for the funding of the project; to all members of the steering committee (including representatives of people with a personal experience of homelessness); to Koen Hermans and other colleagues for useful comments and suggestions.

Contents

1 | INTRODUCTION: HOMELESS PEOPLE ARE HARD-TO-REACH

Some specific populations can be harder to survey than the general population. Homeless people are frequently mentioned as hard-to-survey groups by the literature (Kalton, 2014). The reasons explaining why groups as the homeless people are harder to survey seem to vary. Tourangeau (2014) has classified sources of difficulty, taking into account the different, sequential steps of the survey process. This results in a typology that distinguishes between populations (or members of these populations) which are *hard to sample*, *hard to identify*, *hard to find or to contact*, *hard to persuade* to participate and *hard-to-interview*. In this section, we present the typology of Tourangeau (2014) from the perspective of homeless people with a focus on the steps of the survey process that might be used for counting or estimating the number of homeless people: To which extent are homeless people hard to survey? Which characteristics of the specific context in which homeless people live, make them hard to survey? How can hard-to-survey groups of homeless people nevertheless be reached or contacted? Although the amount of necessary steps of the survey process varies according the used method, the following overview will clarify why counting or estimating the number of homeless people often poses a great challenge:

1.1 Hard-to-sample

Sampling groups can be used to estimate the size of the target group and its prevalence in the total population. Most cited examples of hard-to-sample populations are rare populations and hidden populations. Rare populations represent only a small fraction of the frame population. Given that the proportion of homeless people is considerably small compared to the rest of the population, they are definitely considered to be rare. Next to rare populations, hidden populations pose challenges for sampling. Cepeda and Valdez (2010) define hidden populations 'as a subset of the general population whose membership is not readily distinguishable or enumerated on the basis of existing knowledge and/or sampling capabilities' (Cepeda & Valdez, 2010). Taking into account that a significant proportion of homeless people are hard to track down and do not appear in registers, they can be considered as a hidden population. Consequently, homeless people are hard to sample because of their small number and/or their unknown character.

1.2 Hard-to-identify

When surveying specific groups, wider samples of respondents first need to be screened in order to ensure that subgroups can be identified as belonging to the target group. To meet this condition, it is necessary that the respondents or informants are 'both willing and able to answer the screening questions accurately' (Tourangeau, 2014). However, many homeless people are not eager to be identified, due to stigma, sensitivity or motivated misreporting. Moreover, informants are not always willing to answer the screening questions correctly. In that respect, Glasser et al. (2014) referred to the underreporting of homeless people in the US census. The main reason was that family or friends, who are hosting homeless relatives for a short or longer period, but who are tenants themselves, 'would not want to risk jeopardizing their housing by admitting that additional people were staying with them.' The fear of losing welfare benefits or incurring some other penalty may thus lead to deliberately omitting of some household members (Glasser et al., 2014). Other

reasons of misreporting are suspicion or distrust in outside researchers and a lack of interest in being researched.

Next to reducing misreporting by the hard-to-reach groups, the accurate classification of individuals into a specific group by the researcher is critical. To that end, clear definitions of group characteristics are needed. According to Williams (2010), defining the homeless population poses a great challenge. The difficulties of definitions are compounded by the very mobile nature of the homeless population. For definitional purposes homelessness must be operationalised as measurable locations in a given geographical area. However, homeless people are elusive due to (1) their geographic mobility, (2) their limited visibility (no bank accounts, no medical registration, not on voting lists, etc.) and (3) their itinerant lifestyle (substance users, mentally ill, etc.) (Williams, 2010).

1.3 Hard-to-find or hard-to-contact

Another source of difficulty in surveys is that their members are hard to find (i.e. they include relatively high numbers that are never contacted), or hard to contact (i.e. they require high numbers of contact attempts). According to Tourangeau (2014), in general, 'populations that are only loosely attached to a specific home or place are difficult to find', causing problems for population surveys and for longitudinal, or panel, surveys in particular. Besides the problems of high mobility and distrust towards outsiders among the homeless people, the difficulty in finding or contacting this subgroup may also be caused by a lack of knowledge on behalf of the researcher on how, who and where to contact certain groups or individuals. By doing ethnographic research on homeless people, Glasser et al. (2014), for example, found out that the U.S. Census had previously overlooked low cost hotels and motels as possible sites where homeless people could be found.

Based on the framework of Tourangeau (2014), it becomes clear that several difficulties may arise when trying to count the group of homeless people. Due to these difficulties related to simple headcounts, some studies or enumerations have moved beyond simple headcounts of rough sleepers. These may combine simple headcounts with data from hostels and other temporary accommodation on the enumeration point of time. However, many quantitative studies of homeless people have not attempted to enumerate, but use alternative sampling strategies to estimate the number of homeless people. As such, a distinction can be made between methods that directly attempt to enumerate the populations using census type methods (such as headcounts) and methods that provide samples which can be inductively inferred to emulate the population.

In what follows, we will discuss the most important sampling methods used to estimate the number of homeless. Although the focus will be on the potential of the capture-recapture method to estimate the size of the homeless population, the possibilities of location sampling and network-based sampling will also be discussed. The last section includes a feasibility study of a capture-recapture approach based on linked administrative sources in Belgium.

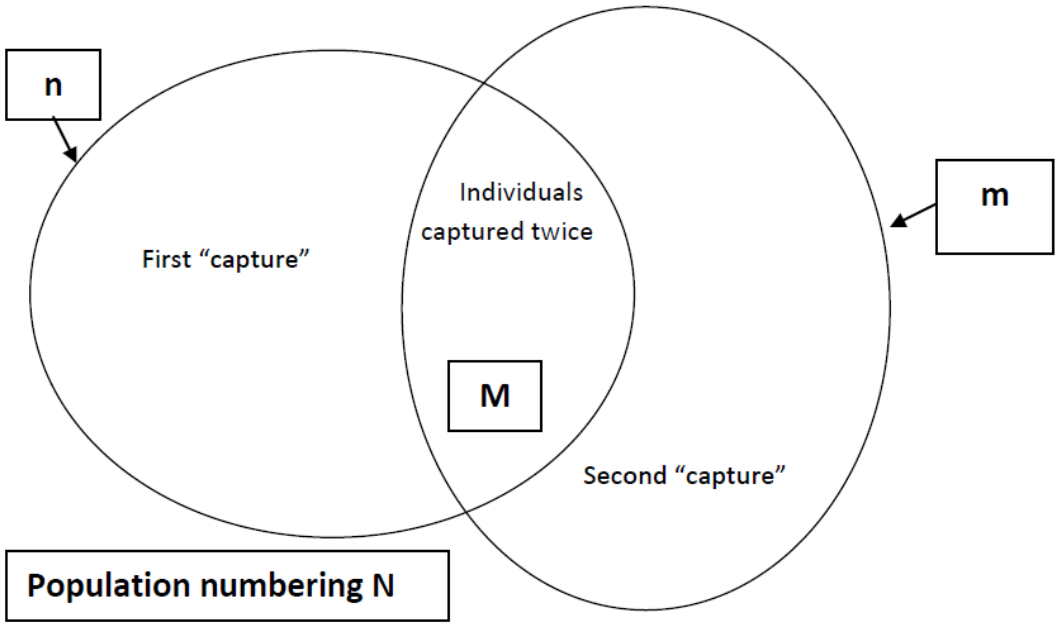
This report is part of the broader MEHOBEL-research project (Measuring Homelessness in Belgium) which aims at developing a methodology to measure and monitor homelessness in Belgium.

2 | THE CAPTURE-RECAPTURE METHOD

2.1 Method

The capture-recapture technique is based on at least two independent observations (or sources) of the target population. In order to estimate the size N of the target population, the number of persons in the populations observed the first time (n), the number of persons observed the second time (m) and the number of persons observed on both occasions (M) need to be known. N is then estimated by calculating (n*m)/M. The persons have to be identified in an identical way in both samples (by whatever identifier) in order to measure the intersection M (for more details about the deduction of the formula, see Brittain & Böhning (2009)).

Figure 2.1 The capture-recapture method



Source Marpsat & Razafindratsima (2010)

In the field of homelessness, there are two main variants of capture-recapture applications:

1. Using two or more independent samples of the study population in their own environment. In the case of homeless persons, the observations are made, for instance, where homeless make use of services. They then fill out a brief questionnaire at each visit to one of the centres observed. The questionnaires are anonymous but an identifier is used that allows the persons re-contacted to be identified.
2. Using incomplete administrative lists with members of the study population (Trochim, 2006).

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 (Coumans et al., 2015). While this method's

underlying concept is simple, the hypotheses that must be met in order for the model to remain valid are fairly restrictive:

- first hypothesis: the observation must be collected in such a way that all individuals in the population have the same chance of being selected in each period: In the case of the homeless, the population of interest may be extremely heterogeneous in its characteristics and also its behaviour. For instance, if the frame for a survey of the homeless omits certain sites, then the homeless linked only to those sites are likely to be missed in both the initial and recapture survey. This may induce bias if *either* the profile of visitors *or* their probability of recapture are different from those visiting the selected locations;
- second hypothesis: the observations on two different occasions are independent: Several elements lead to the assumption that the same persons (those living on the sidelines of the assistance services, avoiding contact with interviewer and managing to hide particularly effectively on the streets) are the ones who elude observation. As such, lists from shelters or medical records can be problematic because the homeless closer to service providers are more likely to know of and be referred to such facilities. When members of the rare population systematically vary in their elusiveness, this variation will produce correlation bias;
- third hypothesis: the reference population does not change between two contacts or captures, i.e. the population remains fixed. This hypothesis can only be relaxed if the number of 'captures' increases. Where there are changes in the reference population, the use of the simplest model also leads to biases. For mobile populations such as the homeless, short time intervals between observations help ensure a closed population (Berry, 2007).

2.2 Capture-recapture in practice

Taking into account the specific characteristics of the hard-to-reach population of homeless people, it is extremely difficult to comply with the stringent assumptions of the capture-recapture approach for estimating the homeless population in practice. However, certain advanced applications of the approach can circumvent those assumptions. In the following section, we will first discuss the implementation of rather traditional capture-recapture methods based on observations in Canada and the United Kingdom. The two other cases are implementations of the more advanced approach of the capture-recapture method using incomplete lists in the Netherlands and Australia.

2.2.1 The walking and vehicle approach of capture-recapture in Toronto, Canada (Berry, 2007).

Those providing services to homeless regularly identify insufficient data on the homeless population as a major problem, particularly information about the street homeless who tend to not use shelters. An accurate estimate of the street homeless population is useful for projections of service needs such as emergency response during cold winter months. However, most large cities, including Toronto, primarily relied on indirect estimates of the homeless population, in particular annual shelter-use statistics. Unfortunately, these annual cumulative counts poorly gauge the need for street-oriented homeless services in the city at any given point in time. In order to overcome this problem, Berry (2007) tested a novel capture-recapture estimation approach that models individuals' intermittent daytime visibility to enumerate the street homeless population.

In the study, two variants of a street-centered capture-recapture approach were piloted in order to estimate the daytime outdoor homeless population in winter in the downtown core of Toronto. As such, the repeated daytime street observations identify an important subset of the homeless - those who regularly engage in 'street living' in the public spaces visible from and adjoining the street environment. The observations were carried out in downtown Toronto. The boundaries of the study area circumscribe 90% of the homeless-service agencies and shelters in the central downtown area.

The researchers carried out two variants of visual capture-recapture- a multiteam walking approach and a single-team vehicle approach:

- the multiteam approach consists of 42 data-collection teams (two or three members per team) which visited 42 randomly selected intersections twice on a specific day (10 March 2004). Each team was instructed to walk a predetermined one-block periphery on the sidewalks and public spaces of their assigned intersection area. To minimise double counting, the data-collection periods were simultaneous for all teams;
- the single-team approach uses a four-member team making observations from a specially equipped vehicle (one driver, three observers). For the pilot project, the team drove a random continuous predetermined route three times on a specific day (9 March 2005).

In both approaches, the teams relied on visual identification of the homeless population in the street environment using a recognisable set of behavioural and visible markers. Where possible, each observer recorded gender, ethnicity, age, and specific notes on physical characteristics, behavior, and location. Observers were strictly instructed not to interfere with the people they were observing. Despite its limitations, direct visual identification of homeless on the street has several benefits over incomplete lists acquired within institutional settings: As such, visual identification has been used successfully in prior counts in which there is appropriate training, it can feasibly be completed within a single day to minimise bias caused by population mobility, and it is unlikely to cause the individual to avoid follow-up contact as interviews might. Street exposure also gets closer to the immediate collective risks the homeless face.

Using the numbers of homeless individuals observed in multiple samples (i.e., the repeats) and the numbers observed in just one sample, capture-recapture methods estimate the number not caught in either sample to infer total population size. The estimates scaled up to represent the entire study area suggest a daytime outdoor homeless population of about 900 to 1,100 unique individuals on the observation days.

The researchers concluded that the use of the capture-recapture method entailed several limitations:

1. Capture-recapture estimators are sensitive to deviations from their assumptions. Assumptions about the distribution of sighting probabilities lead to significantly different estimates in the alternative models.
2. Ambiguity in defining who is homeless based on visual observation may create some unfortunate assumptions, and inevitably, some inaccuracies.
3. Street-based capture-recapture methods are not well suited for all urban environments. The approach is most useful in urban environments in which the homeless have a fairly high probability of being sighted (i.e. 10% to 20%). The researchers believed that the downtown core of Toronto fits this criterion because there are few marginal spaces away from the streets.

Despite these limitations, the findings show promise. The further refinement of systematic observations of street environments can provide useful and reasonably narrow confidence intervals of vulnerable street populations. Unlike administrative lists, the street data are independent of forces of selective and self-selective referral of the homeless to one place or another. According to the researchers, capture-recapture methods may be most useful in conjunction with other approaches. One possible use of capture recapture-based street estimates would be to combine them with surveys of service users (shelters, soup kitchens, clinics) that gauge general time-use patterns and what proportion engage in daytime street life. These two pieces of information may be combined to provide an alternate projection of the overall homeless population that can be compared to conventional estimates. Another possibility would be to conduct carefully timed daytime counts of individuals in shelters, soup kitchens, and clinics within a given study area while street observations are being conducted.

2.2.2 Repeated simple capture-recaptures in Torbay and Plymouth, United Kingdom (Williams, 2010).

In order to estimate the number of homeless persons in the UK, Williams (2010) used the capture-recapture method. The researchers applied the method in Torbay and Plymouth, where the number of homeless people was estimated on the basis of their visits to various services intended for them. In these studies, the geographical factor was beneficial for the research targets. Both urban areas were surrounded by countryside with a low population density and the sea, thus making movement in and out of the areas into adjoining neighbourhoods somewhat easier to control for.

In capture-recapture individuals should be reliably 'tagged'. In these studies, this was achieved through four identifiers: sex, date of birth, where staying at present, and length of time (converted into days) spent in either study location (Plymouth or Torbay). In most agencies these data could be taken straight from agency records, but for the rest a standard monitoring form was completed. It was inevitable to fully exclude the presence of false positives and false negatives, i.e. some people will be wrongly identified as homeless and others wrongly identified as not being homeless. In addition, assumptions about dependency cannot be entirely identified. Nevertheless, the data on location enumeration permitted a sampling of individuals to check for causal links between locations.

Although the homeless are a target group that is ill-defined and very mobile, which makes it even more difficult to enumerate, the Plymouth and Torbay research was considered to be a success. It used a hybrid model that combined a simple two-capture approach, but repeated 3 times over a one year period (providing a total of 6 captures in each location). The longitudinal element provided important data on process, also because the results of each set of captures in each location varied little from the mean, thus indicating some level of measurement reliability. If there had been much variance between the models, the conclusion would have been one of important measurement error or environmental changes.

2.2.3 Capture-recapture of linked administrative sources in the Netherlands (Coumans et al., 2015).

This study presents an advanced application of the capture-recapture approach 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 preeminently 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. The study also illustrates how the capture-recapture approach can be applied to assess the homeless population in terms of background characteristics. By using covariates, they were able describe the estimated population by the levels of these covariates, giving insight into characteristics of individuals not included in any registers.

The focus is on roofless people (between 18 and 65 years old) in accordance with the ETHOS-typology, whom are defined as individuals who had no permanent accommodation on the reference date. Following categories were distinguished by the researchers:

- 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 this study, the capture-recapture of linked administrative sources was used. 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.
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
3. A list of individuals registered as being homeless in the National Alcohol and Drugs Information System (LADIS)

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, the researchers were also able to link for each homeless person the officially registered background characteristics according to the GBA. The capture-recapture technique which was used for estimating the size of a population was 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. 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 X 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. The population size estimate was 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 was obtained by fitting a log-linear model to the incomplete contingency table.

For the reference date of 1 January 2009, the researchers estimated that there are 12,589 homeless people who are not included in any of the three registers based on 5,169 registered homeless. Together, this yields the estimate of 17,767 homeless in the Netherlands, with a 95% confidence interval ranging from 15,601 to 21,225. The results show that the models of the subsequent years are very similar to the finally selected model of the year 2009. The estimates and their associated confidence intervals for reference dates of 1 January 2010-2013 were based on the same method as well as the same updated registers.

The use of the capture-recapture method 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. Another major advantage is that this approach deals with incomplete lists, which is often an evident problem using registers of this population. However, capture-recapture 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, 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. In this case, the researchers assumed that the three-factor interaction between the registers was absent. Another important assumption for the capture-recapture method 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 the researchers used one reference date for each of the registers, this assumption was met. Another necessity is the possibility to link the individuals in the different registers to each other. For this, 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.

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. 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.

The researchers concluded 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.

2.2.4 Capture-recapture based on homeless registers in Adelaide, Australia (D'Onise et al., 2007).

Australia has a significant number of homeless people, with the 2001 census counting 99,900 homeless people, and evidence suggesting that homelessness is increasing. This 2001 Australian Bureau of Statistics (ABS) census is the most widely accepted and used count in Australia, which informs government planning and policy-making. The ABS has used special techniques for counting the homeless since 1996, with the technique changing subtly from one census to the next. The count of primary homeless in 2001 used information from local homeless workers to identify areas where homeless people gather in advance of the census. However, the census count is thought to have problems with over- and under-estimation of certain subgroups in homeless populations. The census was undertaken in winter in South Australia, which reduced the chance of finding homeless people who avoided the bad weather. Moreover, no attempt was made to count people who may have not used the designated gathering places.

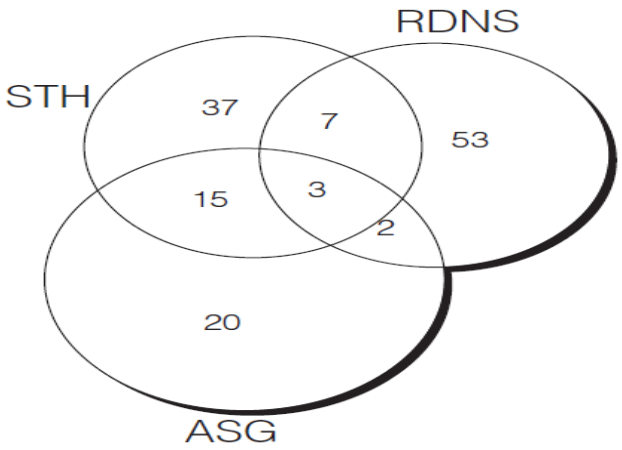
The 2001 census found 104 primary homeless people were living in the Adelaide city council area. This was in contrast to the opinion of people who work with homeless people, who consider the numbers to be far greater. Due to the limitations of the ABS census count, the capture-recapture approach was applied to estimate the size of the homeless population in Adelaide by using overlapping and presumably incomplete but intersecting sets of data about that population. As it is designed to estimate the undercount from existing data sources, the capture-recapture method has the potential to estimate the degree of undercounting by the census of primary homeless people in Adelaide.

The study used the Chamberlain and McKenzie definition of primary homelessness, as used in the ABS census, in order to improve the comparability of the results with the census. Primary homelessness was defined as 'people without conventional accommodation'. As such, all people who were known by the services to live on the streets, in deserted buildings, improvised dwellings, under bridges or in parks or other public places were included as a primary homeless person. The services selected to participate in this study all work with homeless people as their key business, in particular adult homeless people. They were asked to construct a list from their records of all people who met the inclusion criteria. List 1 was provided by Royal District Nursing Service (RDNS) from the clinics they provide in two of Adelaide's homeless day centres. List 2 was provided by the 'Mobile Assistance Patrol' (MAP) service of the Aboriginal Sobriety Group (ASG), a transport service that takes people from unsafe environments to safe environments. The third list

was provided by Street To Home (STH), which is an assertive outreach housing service for homeless people. The lists consisted of the name, date of birth, gender and ethnicity (Aboriginal/ Torres Strait Islander or not) of people who met the inclusion criteria to identify duplicate cases.

Figure 2.2. displays the numbers of primary homeless from the three agencies. There were 137 primary homeless people in total, with 65 on the RDNS list, 40 on the ASG list and 62 on the STH list. A small proportion of primary homeless people were registered with more than one agency (19.7%) and only 2.2% were registered with all three agencies simultaneously.

Figure 2.2 Number of primary homeless people registered at three agencies



Source D'Onise et al. (2007).

This capture-recapture analysis suggests a markedly different result to the census of four years prior, with a much larger population of primary homeless counted in Adelaide in 2005 (137) than was counted by the census in 2001 (104). This difference could indicate a marked increase in primary homeless since 2001 or a systematic under-enumeration of primary homeless in Adelaide city council area by the census. Key informants in the homeless service sector consider the capture-recapture analysis result to be more representative of the numbers of people they work with. Importantly, capture-recapture demonstrates an alternative method for counting homeless populations that is available to homeless services using locally available information.

Nevertheless, the small total numbers involved- in particular, the small proportion of people captured on at least two lists - has resulted in a low accuracy capture-recapture estimation of the true primary homeless population size. Two different statistical modelling techniques were applied to incorporate the dependencies among the samples and overcome the problem of a small data set. Consequently, the results from both statistical techniques used should be considered together:

1. The log-linear analysis gave an estimate of 455 (95% confidence interval of 299 to 762) for the homeless population with a standard error 133.
2. The sample coverage approach estimated the overall homeless population to be 311 (95% confidence interval of 229 to 466) with a standard error 58. In addition, the sample coverage approach detected a net positive dependence among the data, which has been shown to lead to an underestimate of the true population size

Given these findings, the sample coverage result of 311 can be taken as a lower bound estimate for the underlying population estimate. Considering this in conjunction with the log-linear modelling result of 455 (299, 762) indicates a primary homeless population in Adelaide city council area in 2005 of at least three times the 2001 census, but possibly much higher.

Besides the small data set problem, potential violations of the main assumptions necessary for the capture-recapture approach to be valid have to be taken into account (see section 2.1). As such, it

was found that ethnicity and sex may cause dependence among the sampled individuals, violating the assumption that individuals should be homogeneous with regard to capture probability. One solution to overcome heterogeneity is to stratify the sample by both ethnicity and sex, then perform separate log-linear modelling for each stratum. In this study, due to a small sample, stratification resulted in sparse strata, increasing the uncertainty of the estimate for each stratum. Another important limitation to the capture-recapture technique is the relative violation of the third major assumption that the population being studied is closed. According to the researchers, it is likely that this assumption is violated in all population groups to a certain extent, and so this assumption can only be satisfied to a reasonable degree. To this end, the study period was limited to three months.

Due to the limitations, the researchers conclude that the results of capture-recapture studies should be considered to be only an estimate of the size of the population, which in some population groups such as the homeless is an advance on current census techniques. Counts such as the census have value in indicating a trend over time for homeless populations, with relatively similar methods being used over time, particularly on a regional to national scale. It is possible, however, on a small area level for the direct use of services, that the results may not accurately reflect the local population of primary homeless. Capture-recapture techniques have the benefit of being readily applied to the local area level. They are designed to estimate the undercount - the most significant problem -and they allow for qualitative worker involvement in the enumeration, while minimising the involvement of homeless people themselves.

3 | LOCATION SAMPLING

3.1 Method

In the case of mobile or 'elusive ' populations that are not easily linked to any one place (such as the homeless), a potential promising strategy to estimate their number often involves 'indirect sampling', i.e. sampling places or services where the members of the elusive population are likely to be encountered (Kalton, 2014). The general condition for location sampling is that the population of interest visits a certain number of places that, conversely, are not often visited by the rest of the population. In case of the homeless, these places can be shelters, services providing meals, accommodation, clean clothes, etc. These places can be sampled after an exhaustive list has been made of them, and then a sample can be taken of the persons visiting those places. However, attention must be paid to the number of times when the target population visits the premises (Trochim, 2006).

Location sampling has been used to sample a variety of hard-to-sample populations. The general requirement is that a set of locations can be identified such that a high proportion of the target population will visit one or more of these locations during the data collection period. With the visitor as the unit of analysis, the unequal probability of visits across individuals during the data collection period must be taken into account. Moreover, given that sampling is likely to continue for some period of time on the one hand and that such populations are mobile on the other hand, members may also have multiple chances for selection (Kalton, 2014). The Generalised Weight Share Method (GWSM) allows an unbiased estimator to be established by taking into account the different probabilities of inclusion in the sample. At the design stage, the objective must be to create a combination of the most parsimonious list and the shortest possible time interval that has high coverage since that will help to control for multiple chances of selection (Trochim, 2006).

3.2 Location sampling in practice: the ENFAMS survey in the Paris Region (Vandentorren et al., 2015).

In 2013, the Observatoire du Samusocial de Paris conducted the ENFAMS survey (Enfants et familles sans logement) among homeless children and families. The objectives of this study were to estimate the size of this population in the Paris region, to describe their socio-demographic characteristics and to evaluate their health status and needs. The survey was based on a sample of 801 families provided with accommodation in emergency centres, long-term rehabilitation centres, social hotels and centres for asylum seekers.

The sample for the ENFAMS survey was collected based on a time-location sampling design. In a first stage, an exhaustive list of all services accommodating homeless families in the Paris region was created. Data on the number of the sheltered families and the number of minor children within each homeless service was collected through a telephone survey. The type of the service, its distance from Paris and its distance to a train station were taken into account in stratifying the sampling frame into 36 strata. Farthest transport zones and emergency shelters were oversampled, whereas asylum centres and social hotels were undersampled. In the second stage, families were selected using simple random sampling in each selected service. During the third and final stage, one child was randomly chosen in each family among all those younger than 13 years.

The researchers registered 796 accommodation services in Paris region of which social hotels represented 73%, while long-term rehabilitation centres, emergency shelters and asylum centres

represented, respectively, 16%, 6% and 5%. Among the 796 services, 251 were sampled, 237 contacted and 193 finally visited. The estimated number of families in the Ile-de-France region was obtained by summing the sampling weights (inverse of the inclusion probability) of all the participants. Chi-square tests were used to study the factors associated with participation. In the Paris region, the number of sheltered families was estimated at 10,280 (95% confidence interval: 9,507-11,053), which corresponds to an estimated 35,000 individuals with 17,660 children younger than 12 years.

It is important to stress that one of the goals of this study was to estimate the homeless family population and not the entire homeless population. Whereas most homeless families with children will make use of services such as emergency shelters, this is certainly not always the case for homeless people who are single. Given that they are a substantial part of the homeless population, it will be likely that the use of the time-location sampling method will not take them sufficiently into account.

4 | NETWORK-BASED SAMPLING

4.1 Method

Network-based sampling is an approach enabling the researcher to become an active and trusted participant in the community under study and the members of the community to become active participants in the research. Because of this process, recruitment methods have evolved that trace social links or relationships in the community, which allows the researcher to penetrate more deeply and become more substantially integrated with the community. The purpose of these methods is to estimate demographic information about hard-to-reach groups by reaching members of these groups through their social network. Some network-based approaches, such as respondent-driven sampling recruit respondents directly from other respondents' networks making the sampling mechanism similar to a stochastic process on the social network.

Snowball sampling, and its more recent outgrowth respondent-driven sampling (RDS), are the most popular forms of network sampling intended to reduce the problems of identifying members of rare or stigmatised populations such as the homeless. Snowball sampling begins with a convenience sample of initial subjects. These subjects serve as 'seeds', through which wave 1 subjects are recruited. Wave 1 subjects in turn recruit wave two subjects and the sample subsequently expands wave by wave like a snowball growing in size as it rolls down a hill. This process continues until the desired sample size is obtained or the survey period is over. Respondent-driven sampling (RDS) is a sampling design in which members of the hidden population are enlisted to do the actual sample recruitment based on their social connections. It can be seen as an advanced version of snowball sampling. However, unlike snowball sampling, RDS allows unbiased estimators to be produced under certain conditions.

In order to get a reliable estimate on the size of any hidden population, the following conditions must apply:

1. In theory, the respondents should be a random sample from the population, but this is impossible to obtain in practice for almost any hidden population. To approximate a random sample the respondents should be obtained, as much as possible, from several independent sources (e.g. social meeting points). Whether a given population member is included as a respondent in the sample should be independent of whether the nominees mentioned by this respondent also are themselves included as respondents.
2. To obtain an estimate with a reasonable precision, the initial sample size should not be much smaller than the square root of the population size (if on average the respondents mention at least about 10 nominees; for a smaller average number of nominees, the initial sample should be larger) (David & Snijders, 2002).

4.2 Network-based sampling in practice: Snowball sampling in Budapest, Hungary (David & Snijders, 2002).

Although the network-based sampling techniques are mostly known for studying target groups such as drug users, HIV-patients and homosexuals, the method has also been used in poverty related research. As such, David and Snijders (2002) used the snowball sampling method to estimate the size of the homeless population in Budapest, Hungary. Although homelessness was a

relatively new issue in Hungary in 2002, the problems that social policy makers had to face were the same as anywhere else. In order to make any decisions they wanted to know numbers. One of the methods used to estimate the size of the homeless population in Budapest was snowball sampling.

For the research, homeless persons were asked whether they knew other homeless people by name. The original sample consisted of 1,404 people who mentioned 426 names in the first wave. The type of relation was also asked for. Since this population is quite segregated not only from the non-homeless population but internally as well, the rate of those who could mention any other homeless person was low: only 29% of the respondents could give full names of the persons nominated by them. Most people listed by the respondents were from the same place where they themselves usually appear or spend the night. Although the sample size seemed much bigger than necessary, the researchers were aware of a potential underestimation of the total size given that there can be gathering places which were not surveyed by them and which are totally unconnected to the places that were surveyed. Based on an estimator, the researchers estimated the homeless population to reach 3,444 persons. Because of the special structure of the homeless relations, this number will probably be an underestimate. To test this hypothesis, they modified the original data set by restructuring it along the 11 different sites of the sampling. In this case, they obtained the estimate to be 4,097 which may still be an underestimate.

In order to get a more reliable estimate by using a snowball method, the researchers suggest that the sites of the survey should be chosen even more carefully. It is not only the number of the different (quasi-independent) sites that is important but rather the type of that site. For instance, future research using this method should sample places such as hospitals where the chance for homeless people to be independent of each other is higher than in a shelter or in a soup kitchen.

5 | EVALUATION OF THE SAMPLING STRATEGIES

The overview of sampling strategies used to estimate the homeless population shows that each method entails important limitations. Nevertheless, the extent to which sampling strategies are suitable for the estimation of the homeless population differs:

Regarding **location sampling**, one of the main downsides is that establishing and updating the list of the places visited and the times at which they are visited is often time-consuming and costly. Where a high proportion of the population of interest does not visit these places or does so only very rarely, it can lead to a coverage bias. This is certainly the case among the homeless population given that not all homeless people frequently visit shelters or other selected locations. Moreover, there may be data collection problems associated with the places due to refusal by managers/owners, rapid departure of users/clients, etc. In addition, calculating the weights for individuals (e.g. using the weight share method) requires the inclusion of specific questions about attendance of these places. This relies on the individual's memory and may induce errors. Moreover, if multiple use is not duly accounted for, once again the estimators will be biased.

Though **network-based sampling** strategies can be advantageous, financial and logistic challenges often prevent researchers from employing these methods, especially on a large scale. Persons with a poor social network have a lower probability of being reached. Moreover, it is necessary to verify whether each person recruited is a member of the population of interest. This is not easy for all populations and may be intrusive. With network-based sampling, the informant is asked to provide screening data for other persons linked to the informant in a clearly defined way living elsewhere. A key requirement is that every member of the linkage must know and be willing to report on the target population membership of all those linked to them, a factor that influences the choice of linkages. In some cases, the original informant is able to provide the survey data for each of the members of the target population that he or she had identified. However, the informant must know and be willing to report the data, which can be problematic in terms of accuracy and/or sensitivity. Other obstacles might be the difficulty to estimate the refusal rate given that refusals are made to the recruiter, and the complexity of the specific data-collection organisation.

We have seen that the **capture-recapture method** entails fairly restrictive assumptions in order to be valid. Taken into account that, due to financial and logistic constraints, the capture-recapture methods based on visual observations are limited to no more than a few observations, deviations from the assumptions may lead to significantly different estimates. Moreover, the ambiguity in defining who is homeless based on visual observation inevitably leads to inaccuracies which jeopardise the estimations. However, the more advanced capture-recapture technique based on linked administrative sources allows researchers to comply with less stringent assumptions and is less dependent on limited observations. The discussion of the cases in the Netherlands and Australia showed that the potential of this method is highly dependent on the number, size and quality of the lists: Whereas the application of the method in the Netherlands (using high-quality national lists) led to very rewarding results, the use of the same method in Adelaide (Australia) involved substantial difficulties due to the limited sample size.

An important statistical issue which has to be taken into account relates to the dynamics of homelessness. The different sampling methods we have discussed all have in common that they are based on stock data (at a given point in time). The important point here is that one single number will not be enough to understand homelessness and to develop and monitor adequate policies to

tackle it. As the profiles of short- and long-term homeless probably differ dramatically from each other, the statistical picture crucially depends on whether 'stocks' or 'flows' of homeless people are measured. Short-term homeless people are unavoidably 'underrepresented' in stock samples, compared to the long-term homeless. At the very least it is important to develop measures that provide not just the number and profile of homeless people at a given point in time (the stock figure) but also the number of people who have become homeless, or ceased to be homeless, over a given time period (the flow figure, which can be divided into 'inflow' and 'outflow'). A qualified picture therefore requires a combination of longitudinal and point-in-time statistics.

It is unlikely that any census technique, despite the effort involved, is complete and so they should not be solely relied upon for the estimation of the number of homeless people. Based on the overview of sampling strategies to estimate the number of homeless people, we therefore conclude that the capture-recapture method based on administrative sources seems highly recommendable as it saves time, reduces costs and yields 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. Whether this will be possible in Belgium will be discussed in the case study below.

6 | FEASIBILITY OF THE CAPTURE-RECAPTURE METHOD IN BELGIUM

The overview of sampling strategies to estimate the number of homeless people showed that the capture-recapture method of linked registers could be very promising. However, the reliability of this method is highly dependent on the quality of the used registers. In this case study, the possibilities and challenges of applying the capture-recapture approach based on linked administrative sources in Belgium will be discussed. For practical reasons, we will particularly try to 'capture' homeless persons that belong to ETHOS light-categories.[1]

An exploration of available national administration registers resulted in the selection of following data sources with the highest potential for applying the capture-recapture approach in Belgium:

1. Persons with a **reference address at Public Centres for Social Welfare** (PCSW) of the National Register: Reference addresses at PCSWs are reserved for homeless people.
2. The **Novaprima** database of the Public Programming Service Social Integration (PPS SI): The database primarily contains information about ex-homeless who received an installation premium (a one-off grant to cover the costs of moving into a new accommodation) and/or followed an individualised social integration contract (ISIP).
3. The European **Treatment Demand Indicator** (**TDI**) dataset, collected by the Scientific Institute for Public Health (WIV/ISP): The TDI data are collected for every drug user reporting for treatment in the rehabilitation facilities recognised by the Federal Institute for Illness and Disability Insurance. Besides information on drug usage and treatment, the TDI also includes information on the housing situation. Although obviously only a fraction of the homeless population has an addiction problem, this information can be useful provided it can be combined with other datasets.

A successful capture-recapture of linked administrative sources assumes that the used registers are of high quality. As was mentioned in Chapter 2, the reliability of the estimates depends on underlying assumptions including:

- **identical definition of target groups:** the registers need to include (at least partly) the same categories of homeless people. Therefore, a clear definition and delimitation of the population are necessary for each register;
- **a closed population**: the population does not change during the period of observation. This is obviously problematic given the high turnover in homelessness. One way of complying with the 'closed population' assumption is to keep the period of observation as short as possible by using the same reference date for each of the used registers;
- **a unique identification number**: it is assumed that individuals are uniquely identifiable and can be perfectly matched across the different data sources (so that there are no false negatives/positives in matching individuals).

[1] More information about ETHOS light can be found in the report of the broader MEHOBEL-research project.

In what follows, we will verify to what extent the aforementioned federal data sources comply with these necessary assumptions for a successful implementation of the capture-recapture approach in Belgium.

6.1 Assumption 1: the identical definition of target groups

The identical target groups condition implies that the registers include the same categories of homeless people. This implies that only relevant individuals should be selected before analysing the datasets. In the context of the MEHOBEL research project, this means that at least one (hopefully more) ETHOS light category can be matched across the different datasets. In order to verify whether the databases contain the right individuals, a clear definition and delimitation of each database should be possible:

A) PERSONS WITH A REFERENCE ADDRESS AT PCSWS:

The royal decree of 21 February 1997 stipulates the conditions for having a reference address at a PCSW. In this legislative document, the homeless population is defined as 'persons that do not have a residence due to insufficient means'. In this sense, the group of homeless that can have a reference address at a PCSW falls within the ETHOS light-categories. However, it has to be taken into account that important numbers of homeless will be missing in the database:

- the homeless person who applies for social assistance has the duty to sign up at the PCSW once every three months;
- not all homeless persons apply for a reference address at the PCSW. In this sense, homeless persons can have a reference address among private persons. Moreover, a significant number of the homeless persons remain under the radar and do not have a reference address;
- interviews with social workers of PCSW made clear that policy guidelines are ambiguous and willingness to grant a reference address differ (and notably declined during the past years), as was also stated by the inter-federal Combat Poverty Service.

It is obvious from the above that a list including persons with a reference address at PCSW does not include all relevant homeless persons. Taking into account that the capture-recapture method deals with incomplete lists, this should not be a problem as long as the explanatory variables for the inclusion of certain individuals are incorporated in the model. However, at this moment, reference address holders cannot be subdivided according to the reason for applying (such as homelessness, itinerant life or work abroad, prisoners) and the modality of the reference address (private person/PCSW/ organisation) is not specified.

B) THE NOVAPRIMA DATABASE:

The Novaprima database includes information on **ex**-homeless who received an installation premium and/or followed an ISIP. The definition of these homeless persons is based on the legislation of article 1 of the Royal Decree of 21 September 2004. In this sense, homeless persons are defined as 'persons not having a residence, not having the means to provide for such a residence or residing temporarily in accommodation for the homeless' (POD MI 2005). Although the definition of homeless persons falls within the ETHOS Light categories, it has to be taken into account that this part of the Novaprima database deals with the outflow from homelessness and does not contain information about the individuals who do not exit homelessness (in other words, those who are and remain homeless).

C) THE TREATMENT DEMAND INDICATOR:

Contrary to the relatively clear definitions of homeless persons among the individuals with a reference address at PCSWs and the Novaprima database, the categorisation for the Treatment Demand Indicator (TDI) is based on a self-assessment of the individual who fills in the questionnaire. Concretely, one of the questions of the TDI-questionnaire asks about the 'living status in the past 30 days'. The relevant response categories in the light of this research were 'in the street', 'in an institution' and 'different places'. According to the TDI-guidelines, the 'different places' option refers to clients who have lived in different places (friends' home, street, shelters, etc.), moving from one place to another in the period prior to treatment entry (EMCDDA, 2012). In this sense, it can be expected that a self-assessment of the living situation in the past 30 days might lead to certain ambiguities in the responses.

As to the compliance of the three national databases with the identical target group assumption, the following can be concluded: Whereas the designation of homeless people with reference addresses at PCSWs and ex-homeless in the Novaprima database are based on definitions in legislative documents (royal decrees), the delimitation in the TDI-questionnaire is based on a self-assessment of the individual's living status in the past 30 days. Moreover, the relevant categories in the TDI-questionnaire are rather vague ('street' and 'changing residence'). Ambiguity in defining who is homeless inevitably leads to inaccuracies which jeopardise the estimation. Another obstacle is that the status of the reference address holders (homeless/itinerant/working abroad/imprisoned) is not specified making it impossible to select the homeless individuals from the entire group. The selection of those holding a reference address at PCSW by delivering a PCSW address list to the National Register could be part of the solution. Another problem is that many municipalities still refuse to assign reference addresses to claimants (or find ways to get around the law). Moreover, not all homeless persons applying for social assistance take their reference address at their PCSW. Different practices result in (a) systematic under-counting of the homeless population through reference addresses and (b) potential bias in its geographical distribution.

6.2 Assumption 2: a closed population

A vital assumption for the capture-recapture method to be successfully used is that the population remains stable over the observation period, i.e. that there will not be any new members or departures of old ones. This closed population assumption can be fulfilled by using the same reference date for the three registers. Note that the chosen reference date itself is not relevant for the estimate as long as it is the same for all registers. Strictly speaking, the size of the population estimated consequently refers only to this date.

A) PERSONS WITH A REFERENCE ADDRESS AT PCSWS

The list of individuals in the National Register can be consulted on a specific date. Taking into account that the individuals with a reference address are included in the register, a specific reference date can be easily selected. Consequently, the list of individuals in the National Register can be perceived as stock data (a number of persons at a given point in time).

B) THE NOVAPRIMA DATABASE

The Novaprima database includes a specific date which indicates from when the personal information on each individual's form is valid. Consequently, it is possible to compile a list based on Novaprima on a specific date. However, as was mentioned before (6.1), this list consists of

individuals who *were* homeless in the past. Moreover, it is not possible to verify at which moment these individuals were homeless. In this sense, the Novaprima database only contains (out)flow data of ex-homeless. In other words, people who were homeless in the period of measurement but who did not exit homelessness are *not* registered at all.

C) THE TREATMENT DEMAND INDICATOR

The use of a specific reference date is particularly problematic for the TDI-database since it only registers the start of each treatment. Concretely, each patient is registered every time he/she starts a treatment episode. If the patient moves to another treatment centre to start another episode, he/she is registered again. However, the end of each treatment is not registered. This implies that the TDI-register over-estimates the number of homeless patients on a specific date, because it includes a 'redundant' number of patients who had left homelessness by that time. In other words, the TDI registers the incidence (a number of new cases encountered in a given time period) and not the prevalence (the number of cases at a particular time). Consequently, it is not possible to select a sample for a single reference date. In this sense, the TDI-database can be considered as a flow sample including people who became homeless over a certain period (inflow). It might be possible to reconstruct a stock sample provided that the dates of exit from homelessness could be registered.

As the profiles of short- and long-term homeless probably differ dramatically from each other, the statistical picture crucially depends on whether 'stocks' or 'flows' of homeless people are measured. Taking into account that the persons with a reference address are based on stock data on the one hand and the Novaprima database (outflow) and the TDI-database (inflow) include flow data, the coupling of the different registers would lead to significant bias.

6.3 Assumption 3: a unique identification number

Methods for estimating the size of a closed population from a capture recapture study require the availability of unique and identical identifiers on each of the used lists. These identifiers are used to match individuals across different lists. When individuals appearing on different lists cannot be identified with certainty from the data, matching between the lists is problematic. In Belgium, the most appropriate identifier to link the different databases is the national register number.

A) PERSONS WITH A REFERENCE ADDRESS AT PCSWS

The National Register includes the reference addresses as a separate information type. Consequently, it is possible to draw a list of all persons who hold a reference address in Belgium. Given that the persons with a reference address at PCSWs are part of the National Register central database, their national register number can be easily used.

B) THE NOVAPRIMA DATABASE

The Novaprima forms require the national register number. A form without this number will not be accepted by the PPS SI. This implies that the national register number of all individuals[2] in the Novaprima database is known (POD MI, 2015).

C) THE TREATMENT DEMAND INDICATOR

The TDI database only registers the national register number with consent of the individual. If an individual refuses, the initials and birth date are used as identifier. Over the period 2011-2016, the

[2] For individuals that do not have a national register number, the Bis- number is mentioned.

national register number was available for approximately 75% of the individuals in the TDI-register. As the first digits of a person's national register number co-incide with the date of birth, it should be possible to match most of the other individuals in the TDI-register with the national register by comparing their date of birth and initials with the numbers and names of a limited set of observations in the national register.

6.4 Overall feasibility assessment

The assessment of the registers' compliance with the assumptions clearly shows that the registration in the discussed databases will have to be thoroughly modified for a successful application of the capture-recapture method:

- the modality of the *reference address* holders (private person/PCSW/organisation) is not specified, making it impossible to select the homeless individuals from the entire group. Moreover, not all homeless persons applying for social assistance take their reference address at their PCSW. Different practices result in (a) systematic under-counting of the homeless population through reference addresses and (b) potential bias in its geographical distribution (cf. Section 4.1.1);
- the section of the *Novaprima* database referred to above deals with ex-homeless and does not contain information about the period when the relevant individuals were still homeless. Consequently, the used form should be extended with supplementary variables in order to be considered for the capture-recapture approach (cf. Section 4.1.3). In addition, the complete Novaprima database only contains information on homeless clients who received an installation premium from PCSWs after rehousing and/or followed an ISIP. Although other homeless people who apply for support (other than the installation premium) are also included in Novaprima, they are not registered as homeless because there is no information about their housing situation. This implies that a large proportion of the homeless population (those who remain homeless or who for some reason are not granted the installation premium) are not 'visible' in the database. In order to fill this information gap, it would be appropriate to focus on a comprehensive and uniform registration of the housing situation of *all* clients among the PCSWs.;
- moreover, a list based on the *Treatment Demand Indicator* does not comply with any of the aforementioned assumptions that are necessary for a successful capture-recapture. Even though the reconstruction of stock data based on the current indicator might be possible in the future, the vague definition and incomplete linking issues remain to be tackled.

It can be concluded that the application of the capture-recapture method on federal registers is not yet possible. Further research needs to verify whether the method could be used on the regional level. In this respect, the registration tool of the Flemish Agency for Social Housing (cf. 4.2.3) seems to be promising for Flanders.

REFERENCES

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

Brittain, S., & Böhning, D. (2009), Estimators in capture-recapture studies with two sources. *AStA Advances in statistical analysis, 93*(1), 23-47.

Cepeda, A., & Valdez, A. (2010). Ethnographic strategies in the tracking and retention of street-recruited community-based samples of substance using hidden populations in longitudinal studies. *Subtance Use and Misuse*, *45*(5), 700–716.

Coumans, A.M., Cruyff, M., Van der Heijden, P.G.M., Wolf, J., & Schmeets, H. (2015). Estimating Homelessness in the Netherlands using a capture-recapture approach. *Social Indicators Research*, 24 p.

David, B. & Snijders, T. (2002). Estimating the size of the homeless population in Budapest, Hungary. *Quality & Quantity, 36*, 291-303.

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.

EMCDDA (2012). *Treatment Demand Indicator (TDI): Standard Protocol 3.0.* Lisbon: European Monitoring Centre for Drugs and Drug Addiction j, 80p.

Glasser, I., Hirsch, E., & Chan, A. (2014). Reaching and enumerating homeless populations. In R. Tourangeau, B. Edwards, T.P. Johnson, K.M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 180-200). Cambridge: Cambridge University Press.

Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology*, 35(2), 125–141.

Marpsat, M. & Razafindratsima, N. (2010). Survey methods for hard-to-reach populations: introduction to the special issue. *Methodological innovations, 5(2)*, 3-16.

POD MI (2005). *Toelichting bij de uitbreiding van de maatregel tot de toekenning van een installatiepremie aan personen die hun hoedanigheid van dakloze verliezen*. Brussel: POD Maatschappelijke Integratie, 9p.

POD MI (2015). *Leidraad bij de te gebruiken formulieren voor de terugbetaling van de kosten voor maatschappelijke integratie in het kader van de wet van 26 mei 2002*. Brussel: POD Maatschappelijke Integratie, 31p.

Tourangeau, R. (2014). Defining hard-to-survey populations. In R. Tourangeau, B. Edwards, T.P. Johnson, K.M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 3-20). Cambridge: Cambridge University Press.

Trochim, W. (2006), *Probability sampling* [http://www.socialresearchmethods.net/kb/sampprob.php]

Vandentorren, S., Le Méner, E., Oppenchaim, N., Arnaud, A., Jangal, C., Caum, C., Vuillermoz, C., Martin-Fernandez, J., Lioret, S., Roze, M., Le Strat, Y., & Guyavarch, E. Characteristics and health of homeless families: the ENFAMS survey in the Paris region, France 2013. *European journal of public health*, 1-6.

Williams, M. (2010). Can we measure homelessness? A critical evaluation of 'Capture-Recapture', *Methodological innovations, 5*(2), 49-59.