

Possibilities and Limitations in Longitudinal Analysis of Homelessness Based on Information From Official Registers

Tobias Børner Stax¹

The Danish National Institute of Social Research

Possibilities and Limitations in Longitudinal Analysis of Homelessness Based on Information From Official Registers	1
<i>Én gang socialt marginaliseret – altid... ?</i> : attempting to identify overall changes	2
The population	2
The information obtained through official registers	5
A unique identifier	6
Access to abundant information and less worries with drop outs.....	6
Where the data came from – and a quick presentation of a few distributions	8
Table 2: Way of death and primary cause.....	9
The research questions	13
Some problems in using register-based data based on the experiences from <i>Én gang socialt marginaliseret – altid... ?</i>	14
All there is, is already there	15
Problems with private ownership of registered information.....	15
Problems concerning time.....	16
Only some information exist.....	17
The construction of register information	19
A project under way: attempting to reconstruct trajectories	23
The Sample and the data	24
The research questions	25
References	26

This paper reflects upon the use of information from official registers in – primary longitudinal – research on homelessness. The aim is to discuss the strengths and weaknesses in methods using such data. I shall mostly reflect on this based on experiences with a concrete research projects: *Én gang socialt marginaliseret – altid ... ?* by Stax (1999; 2003). In the end of the paper I shall turn briefly to a second project currently underway at The Danish National Institute of Social Research

¹ Herluf Trolles gade 11, DK-1052 Copenhagen, Denmark; tob@sfi.dk

which – also based on data from official registers – tries to look more in depth at trajectories over time than it was done in *Én gang socialt marginaliseret – altid ...?*

Én gang socialt marginaliseret – altid...?: attempting to identify overall changes

The study I shall dwell upon mostly is conducted by myself and published at The Danish National Institute of Social Research in 1999. The title will in English be something like “Once Socially Marginalized – always...?”. The overall aim of the research was to look at living conditions over time among people who had used institutions for homeless (e.g. a shelter or a hostel) or centres for battered women in 1988-89. The analysis focused on the time *after* they had stayed at such a place in 1988-89. As such it was a study of what happens to people already having experienced homelessness and not a study of the causes leading to homelessness. The study drew upon two methods: analysis of data gathered from official registers and analysis of data gathered through qualitative interviews.

The population

Én gang socialt marginaliseret – altid ...? is based upon a registration of 1.185 people who used institutions for homeless people and/or centres for battered women² between the summer 1988 and the summer 1989. Before 1998 the people who used shelters, hostels, centres for battered women, or other institutions³ established under the section 105 in The Social Assistance Act⁴ were not registered on a national level in Denmark. However these 1.185 people were registered for another research project conducted by P. Brandt (1992).

Intentionally the 1.185 people should include every person who

² The reason that both types of institutions were included was that they there both provided in accordance with the same section in the social legislation.

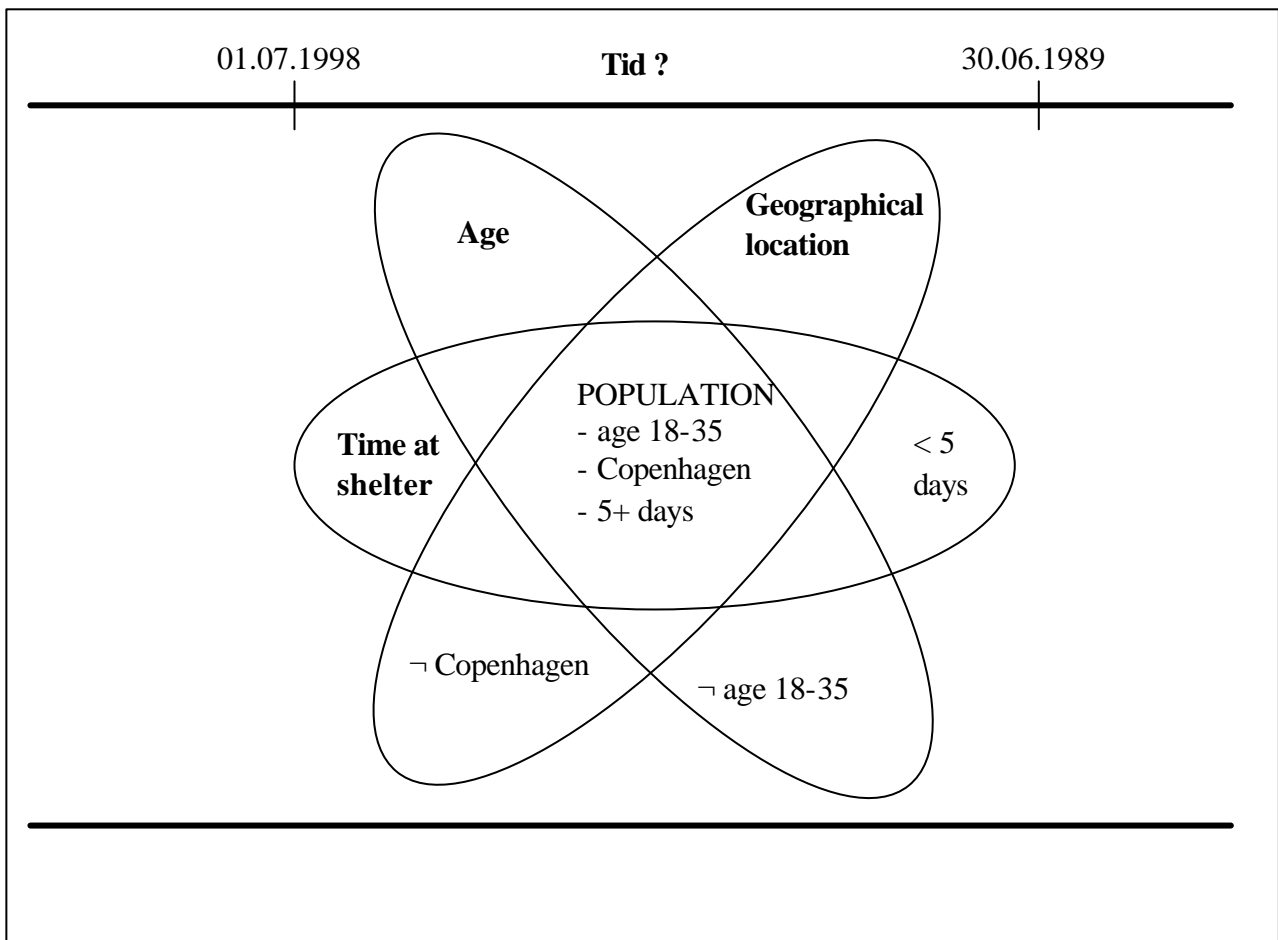
³ I shall mostly refer to all the different institutions as shelters throughout the rest of the paper. Sometimes, however, I shall distinguish between centers for battered women and the rest.

⁴ Today the same type is organized in accordance with section 94 in the Bill on Social Service (for a discussion of the changes in the Danish social legislation and its implication for homelessness see Stax & Kæmpe, 1999).

- used an institution for homeless people or a centre for battered women provided in accordance with section 105 in The Social Assistance Act
- in the Copenhagen Municipality
- between July, 1, 1988 and June, 30, 1989
- as long as that person was between 18 and 35 years of age
- and used the institution for more than four days in a row.

In addition the registration included some people who had used the institution for four days or less as the registration at the institutions was to be carried out by the fifth day of enrolment the latest. Thus, there should thus be no differences between the population and the sample in this study: no one has been selected for the sample from among the population, as the whole population is – at least intentionally – included.

Those included in the population can be illustrated with figure 1



On the people fitting within these criteria the following information was gathered:

- A person CPR-number, i.e. an unique personal identifier (see below).
- Which section 105-institutions in the Municipality of Copenhagen that were used by each client during the registration period, i.e. between the summer of 1988 and the summer of 1989.
- How many enrolments did each client have during the registration period.
- How many days did the client stay at an institution during each enrolment in the registration period.⁵

These information were the ones gathered in 1988-89, and they were the ones handed over to us. I could add no further information on the clients' use of shelters during the period of registration should I have wanted to.

During the project the sample declined from the initial 1.185 people to 1.016 people: 96 people were taken out of the population due to what turned out to be an invalid CPR-number (see below), 68 people had not spend five days or more at the institutions,⁶ four people did not fit the limitation in age or information on the use of institutions were missing. One person was left out due to a discrepancy between the use of institutions and official date of death.

Table 1: Some information on the population (N=1.016).

		Gender		Citizenship			Age		
		?	?	Danish	The West*	Others	18-23	24-29	30-36
Gender	?	336		289	14	33	109	130	97
	?		680	600	25	55	125	247	308
Citizenship	Danish			889			202	336	351
	The West				39		5	10	24
Age	Others					88	27	31	30
	18-23						234		
	24-29							377	
	30-36								405

* That is Nordic, EU and North American (excluding Mexico) countries.

** Age per mid 1989

⁵ Regarding people who were staying at an institution at the time where the registration period began, these were counted as if they enrolled at the specific shelter on the first day of the registration period. In regard to people who were enrolled in the registration period, and stayed beyond the June, 30th 1989 the number of days counted for the research project were only those within the registration period.

⁶ As I expected that there were quite a lot of people who had stayed for shorter time than five days and who had not been registered I decided to leave out this group all together. It might here be added that the five-day limitation on the registration procedure might also exclude some people who are very mobile – drifting between institutions frequently. Theoretically it is possible that a person has spend a good deal more than five days at a shelter during a year, but none of his or her stays are more than a couple of days and the person is thus not included in the sample.

The population in this research project is somewhat special due to its inflexibility. The Danish National Institute of Social Research were handed unique identifiers on a specific group of people where the unique identifiers were gathered about eight year prior to our involvement in the project. At the time we decided to attempt to identify the changes in living conditions of the people included in the study there were no chances of changing the population, e.g. no further individuals could be included should I for example find the age or the geographical limitation of the population strange. I had to work with the population that was handed over to us.

Furthermore I had no chances of making an own assessment of the reliability of the study concerning the relationship between the intended population of the study versus the actual population of the study. To illustrate the point: I could not make an own assessment of how the procedure of having the different shelters report information on their clients worked: did they for example do it in a similar fashion? How were they informed about the project? Might there have been differences between the procedures at the different institutions reporting? Etc. This became an more severe problem, as the issue of how the data were generated was only touched upon very superficially by the researcher who initially gathered the data, and no documentation of the procedure – except a brief presentation in a book where the results where published – existed as I embarked upon the project.

With these limitations on the possibilities for manipulating the population and for including additional variables concerning the use of shelters back in 1988-89 I can treat these information as equivalent to those coming from an official register – the information are gathered by others and in accordance with rationales that I – the researcher – have not been involved in constructing or discussing.

The information obtained through official registers

In 1997 The Danish National Institute of Social Research was handed the registrations of the 1.185 people and I initiated the project. I intended to reconstruct the living conditions of the homeless people after their use of shelters in 1988-89 through the collection of information already existing in official registers.

A unique identifier

One central condition that made the study possible needs to be mentioned here. In Denmark life is generally assumed to begin with the birth of the child and then follows a continual development of identity until death – which might not be different from most other places. However, what might be different is, that one of the first components in this continuous construction of identity is a ten-digit number – the CPR-number.⁷ This number follows a person during life – and stays after death – and it does not undergo a change like the rest of our identity. The number is unique in the sense that there should not exist two alike. The number includes date of birth (ddmmyy) and four personal digits where the last indicates sex (equal for female, unequal for men). In almost any encounter with public institutions – and in many encounters with private institutions as well – the number is provided and registered, e.g. when visiting doctors, dentists, taxation offices, educational institutions, trade unions, employers, libraries, sports clubs, banks, video rental stores... Ideally – from a researchers point of view – the number thus provides a key for tracking down interactions between a concrete person and a large number of institutions. Especially when the registrations of encounters are collected and kept centrally – which is often the case in Denmark, at least in regard to encounters with public institutions.⁸ The CPR-number was the key that made *Én gang socialt marginaliseret – altid ... ?* possible. With the CPR-number I could get information on each individual in the population from other sources.

Access to abundant information and less worries with drop outs

Using register based information seems to present some advantages in comparison to e.g. conducting a panel study, or attempting to conduct a survey on people having at any given previous time been doing something of interest: First one gains access to a host of information. In *Én gang socialt marginaliseret – altid ... ?* I requested information on about 150 variables, which (almost all of them except those concerning criminality – see below) was ready for statistical analysis right away. In the ongoing study currently underway at The Danish National Institute of Social Research

⁷ Should one not be born in Denmark, the number is provided as soon as one seeks permission for staying longer than one is allowed as a tourist.

⁸ I of course turn my blind eye to those aspects that might question the intentions in the system with the CPR-number: I do not reflect here on the possibilities of people living illegally in Denmark, I do not reflect upon the possibilities of

information on 417 variables from 20 different registers have been requested, but not yet obtained (see the final part of the paper).

Second, one does not have to worry about recapturing the population that once used shelters. This could otherwise have posed a significant problem as should become clear by presenting some findings from *Én gang socialt marginaliseret – altid ... ?* In the study I attempted to find out the number of people not living where they were thought to be living according to information on addresses in official registers. By visiting the officially registered address for each person who formally had such in the Copenhagen Municipality⁹ and then to look for signs of the supposed inhabitant I found:

- That 80 percent did show signs of living on the address (either their name was located on the front door, on the mail box, or on a display board in a staircase).
- That at 15 percent of the addresses only other people's names were present.
- That at 5 percent of the addresses there were no names at all.

Thus, of the 424 people who's addresses I looked at there were some uncertainty as to whether they actually lived there in 20 percent of the cases. A significant portion of this group I would expect not to be able to reach should I attempt to conduct e.g. a survey. Additionally one cannot include the people who has died or disappeared (see for example table 4 below) in a survey, whereas 'their' experiences can be included in the register-based research.

A further indication on the difficulties of re-capturing people that once lived at shelters might as well be found in the data generation process of *Én gang socialt marginaliseret – altid ... ?* In my attempt at finding people willing to participate in qualitative interviews a total of 90 letters were sent to selected people from the population. Only on a few of the letters did I get a positive response.

- 24 letters were returned with the note "addressee unknown" attached.
- 41 letters I never saw again nor did I hear from those who should have received them.
- 25 people responded on the letters, but five of these said they were not interested in participating in an interview.

people not informing the public of an eventual birth. I do not reflect upon the possibilities of people having avoided a number when the system was introduced in the '60'ies. These are valid questions but I cannot evaluate their impact.

⁹ 424 addresses

Of course there might have been a different response rates had I send out a questionnaire, or maybe formulated the letter of introduction differently, but I do think that the low response rate also indicates the existence of a significant problem should one attempt to conduct a survey or a panel study with people who had in the past used institutions like a shelter.

In comparison with a survey the problem with non-response can be assumed to be much smaller in register based research. As the registers are not dependent upon whether a concrete citizen can be located and / or decides to participate, but are dependent on the public authorities' documentation of their decisions as professional practitioners and doing so being part of their job, one could expect fewer missing information. It might be expected that the public authorities know, e.g., which dwellings that exist – especially in a country with only very few squatting areas – or who would have used treatment centres for people using illegal substances – especially in a country where such places for treatment rely heavily on public financing which is again based on registration of enrolments (by CPR) – or who would have had encounters with the police, or who have received unemployment benefit, etc.¹⁰ Thus drawing upon information not necessitating a subjects own, voluntary participation the problem of non-response seems to be irrelevant.

However, in conducting a study based upon registers other problems surface:

- There is a limitation on the information available that is beyond the control of the person conducting the study.
- A conversion of information continually takes place and it can be rather difficult to evaluate the effect of these conversions.

I shall return to these two issues later, but first I will outline the actual information that was included from registers in *Én gang socialt marginaliseret – altid ...?*

Where the data came from – and a quick presentation of a few distributions

¹⁰ Such expectations are given by, e.g., Statistics Denmark – where I have obtained many of the register based information. An example: as an argument for a high validity of the information in the National Register they state that as the interests of those providing the information is to have valid information such can be expected. Or they refer to "a general recognition that the quality [which I take to be validity] is high compared to other sources" (Statistics Denmark, 2000, my translation).

The information that was handed over from a previous research project (Brandt, 1992) consisted (as written above) of peoples CPR-numbers as well as registrations of use of institutions in 1988-89. With the CPR-numbers it was possible to gather a whole variety of information pertaining to each person in the population. The information that I chose to include in the project *Én gang socialt marginaliseret – altid ...?* included:

1) Each persons **criminal record** did such exist. Here I obtained information on previous criminal charges that led to conviction and ongoing criminal charges not yet settled. The information was obtained from the police. The information was passed on to us as 7.000 sheets of matrix-printed paper that we had to register electronically. The register included any encounter leading to a criminal sentence independent of date. I thus obtained information pertaining to the life prior to the registration period. However, two years after death a person was removed from the register, and therefore I only received information pertaining to the people having been alive in 1996 as we got the information in 1998. Among the 877 people having used either a shelter or a centre for battered women and been alive in 1996 642 (or 74%) had a criminal record. Among the men the rate was 77% among the women it was 23%. The rate for the people who had used only shelters during the period of registration was 85%.

2) Information on an eventual **death**, and the registered causes for such. These information was obtained electronically from The National Board of Health. The information received included an assessment of the cause of death – i.e. was it for example caused by alcohol, drugs, AIDS or simply unknown - and way of death – i.e. for example natural, accident, or suicide.

A few information is provided in table 2.

Table 2: Way of death and primary cause

	Natural	Accident	Suicide	No registration*	Total
No cause given	2	0	0	14	16
Alcohol	7	1	0	0	8
Drugs	8	37	6	16	67
AIDS	13	0	0	0	13
Other	11	10	10	4	35
Total	41	48	16	34	139

* Including two homicides

Comparing the mortality rate in the sample with the general population given age and gender I found that for each death in the general population 13 people in our sample died. But it also became clear that the sub-group in the sample consisting of women having used shelter (i.e. if one distinguished between the women having used shelters versus women having used centres for battered women) had a mortality rate of 24 to 1, given age, i.e. for each time a woman in the general population died 24 of women in the sample who had used shelters could be expected to die.

3) Information on whether a person had received any **treatment for the use of illegal substances**. These information was also obtained electronically from The National Board of Health. The period covered by this register was two years. When an additional year of information was added the oldest was deleted, thus no more than a two year time span could be included in the research project (I included information on 1996-97). Through this register I obtained information on a persons eventual enrolment in drug treatment, I obtained information on what illegal drugs the person told he or she had used within the month prior to the enrolment in treatment, and what type of drug a client saw as his or her main drug. A few information from this register is provided in table 3.

Table 3. The relative share of people using drugs by selected background characteristics (N=1,016)

	Percentages estimated as using illegal drugs	N
Gender		
Women	19	336
Men	41	680
Citizenship		
Danish	36	889
Nordic, EU, or North America	10	39
Other	15	88
Age*		
18-23 years of age	29	234
24-29 years of age	37	377
30-36 years of age	32	405
Type of §105-institution used in 1988-89**		
Institutions for homeless	40	789
Centres for battered women	9	224
Number of days enrolled in 1988-89		
5-30 days of enrolment	34	286
31-90 days of enrolment	35	289
91-180 days of enrolment	33	193
181-360 days of enrolment	31	248
Number of enrolments in 1988-89		

1 enrolment	27	677
2 enrolments	38	179
3-4 enrolments	52	104
5+ enrolments	66	56

* Age mid 1989

** For three people the institution used in 1988-89 was not registered.

(Table 3 is taken from Stax, 2003:451)

4) Information on the person's **current address, type of dwelling** where this address was registered and the **household composition**. These information was provided through two sources: the *National Register* supposedly containing every citizens formal address and provided to us by the National Register and through various different registers kept at Statistics Denmark.¹¹ The information used from Statistics Denmark covered the years 1990 and 1996. The information from the National Registers covered the situation on a single day in 1996. Some of the information pertaining to the household composition is presented in table 4 and 5.

Table 4: Official registration in the National Register, October, 1996.

	N	%
Registered address	760	75
Registered address similar to a shelter's address	52	
Registered in Greenland	11	1
Special Registration	64	6
Prison, Long term patient, or other 24H institution	17	
Unknown address	47	
Disappeared	7	1
Emigrated	35	3
Dead	139	14
Total	1.016	100

Table 5: Two categorizations of households where a person in the sample formally lives (percentages)

	Centre for battered women		Shelter	
	1990	1996	1990	1996
Household categorized due to assessed civil relationship between co-tenants				
Exactly on single person	55	48	27	45
Exactly one couple	21	37	13	17
More single people	12	9	38	24
More couples or one couple and at least one single	12	6	22	14

¹¹ "Statistics Denmark is Denmark's central statistical office and is responsible for the co-ordination of all official statistics concerning Denmark and Danish society. These statistics cover a broad spectrum of topics within the areas of population, business, industry, the environment and the economy." <http://www.dst.dk/HomeUK/About/PresentSD.aspx> See www.dst.dk for a presentation of the institute and the information that are available for free.

Total	221	216	739	622
Household categorized due to registration of purpose of dwelling				
Address registered as en institution	1	1	21	8
Address registered as residential property	88	97	56	76
Address registered as something else	11	2	23	16
Total	221	207	728	556

5) Information on the types of **incomes received** – for example whether a person did receive unemployment benefit, social welfare, or early retirement pension – the **connection to the labour market** – for example the degree of unemployment – and information on **civil status** and eventual **children**. Also these information was obtained from the Statistics Denmark. Selected information is presented in table 6 - 8.

Table 6: Civil status and eventual kids (percentages)

	Centre for battered women		Shelter	
	1990	1996	1990	1996
Civil Status				
Single	77	61	88	82
Part of a couple	23	39	12	18
Total	221	216	739	622
Children?				
No children	23	27	72	78
Children	77	73	28	22
Total	221	207	728	556

Table 7: Type of transfer income received (percentages)

	Centre for battered women		Shelter	
	1990	1996	1990	1996
Temporary transfer income like unemployment benefit or social assistance	98	88	72	51
Permanent transfer income like early retirement pension	2	12	28	49
Total	200	196	697	575

Table 8: Yearly unemployment rate (percentages)

	Centre for battered women		Shelter	
	1990	1996	1990	1996
Unemployment rate = 0.100	15	15	11	10
0.100 < unemployment rate = 0.500	34	28	24	12
0.500 < unemployment rate = 0.800	13	7	10	7

0.800 < unemployment rate	11	10	6	9
Unemployment rate = 0.000 and no occupation in November	29	41	49	62
Total	200	196	697	575

The research questions

These were the information with which I hoped to answer the question: what happens to whom after they have used a shelter?

From the beginning of the project I were inspired by a passage in a Danish novel dealing with an escapee from a prison. During the first part of the book after the prisoner has run of he generally slide downward and when he is about to hit rock bottom one can find the following quotation from the book:

“The only possibility left is to enlist among the excluded. Here he will receive food and sheltering. But has he first reported here, has he first placed himself among the down-and-outs, he shall never, in any way, be able to return, to find a place, not even the most humble place, within society. Slowly he approached the shelter...” (*Madsen, Tugt og utugt i mellemtiden* [own translation])

My initial question might be described as an attempt at finding out whether there was any truth in the quotation? Was it the case that exclusion and marginalisation tended to embrace its subject so completely and possessively that they would never leave?

Basically, in *Én gang socialt marginaliseret – altid ...?* I would have liked to have information on a few very general questions:

- Does a given respondent have a more or less secure and normal position in society a certain amount of time after having stayed at a shelter?
- Does a given respondent have a better, a worse, or an unchanged life a certain amount of time after having stayed at a shelter?

And I would have liked to know something more specifically about the respondents' current living situation, like whether:

- A given respondent have a satisfying place to stay according to her or him self or according to some politically accepted standards?

- Or a given respondent have the finances necessary for sustaining acceptable living conditions according to him or her self or according to politically acceptable standards?

Of course I would also like to know about causes leading to the better, the worse, and to status quo – but first and foremost I was interested in getting to know something about the development in the actual living conditions of the – at least previous – homeless people. This also turned out to be enough for this one project – but today we are at The Danish National Institute of Social Research working on with the questions concerning causality (see the final part of the paper).

I tried to answer these general questions by looking for changes in the population's position in regard to some of the phenomena that are associated with homelessness:

- Could I find indication that they were capable of finding permanent housing, thus finding a place in society on a housing dimension?
- Could I find indication that they were capable of finding a job, thus finding a place in society on a working dimension?
- Could I find indication that they were capable of becoming self-supporting, thus becoming independent of public support?
- Could I find indication that they were capable of establishing a family, thus finding a place in society on a civil dimension?
- Could I find indication that they were capable of leaving behind the social problems associated with a marginal position, e.g. a use of illegal drugs and criminal activities?

These were some of the questions that I attempted to answer with the information gathered.

Some problems in using register-based data based on the experiences from *En gang socialt marginaliseret – altid...?*

I shall now attempt to outline the problems that I encountered in the attempt at answering the above questions through the use of register-based information. I have already outlined what I find to be the two overall types of problem:

- That there is a limitation on the information available that is beyond the control of the person conducting the study leading to problems with reliability.
- And that a conversion of information continually takes place making it difficult to evaluate validity of the information.

All there is, is already there

The first type of problems concerns a defining characteristic of register based research: that it is working with information created by others than those doing the research and working with information created prior to and independent of the specific research project. At least this is the case if I – as it is done in the study by Stax, 1999 – define register based research as “[r]esearch drawing data from registers created for administrative purpose rather than created for research. The research thus draws upon data from sources established prior to the actual research project. A register is defined as a set of standardized quantitative data that can be re-associated with the concrete object that the data is intended to inform about, where the object can be an enterprise, a person, or a similar unit.” (Johnson, 2003:763).

Problems with private ownership of registered information

Initially I expected to be able to combine all the different information in one big bulk and then through statistical analysis of theoretical expectations attempt to assess the influence on future life situation of the various variables included in the data material. But this expectation had to be modified. As the information from different registers came at different point in time, and as one of the places from which I obtained information – Statistics Denmark – would not allow these to be transferred out of their internal data system, and as bringing the rest of the data into their system was neither cost-free, nor possible due to the different points in time at which I obtained the information from the different registers, I decided to work with the information in two bulks. In the one group I combined the information on contact with the police, on treatment for the use of illegal drugs, and information about deaths with the information about the use of institutions in 1988-89. In the other group I combined socio-economical information – source of income, housing situation, family, and contact to the labour market – with the information about the use of institutions in 1988-89. The two bulks of information were not combined. Had I primarily been looking for trajectories this design would not have worked, but given the time and finances available no alternative existed. This downgraded the questions of causes even further.

This brings me back to one of the previously mentioned problems in register-based research: the availability of information. At least two different problems on availability exist: existence and access. Access I thus encountered as a potential problem in regard to 1) keeping costs at bay and 2) in regard to different institutions policies. In Denmark the owner of a register decides part of the rules under which the information can be shared.¹² That this might raise some problems as can be illustrated with my attempt at obtaining information on the population's criminal record. As stated above I obtained these information from the police, and I received them as matrix-printouts that we had to transform into an electronic format. But the information actually did exist electronically at Statistics Denmark. However, they would not provide us with exactly these information as they found that they were too sensible, given the population. As an alternative I thus tried to obtain the information from the source also delivering the information to Statistics Denmark, and they had a different understanding – and provided us with the information.¹³

Problems concerning time

In *Én gang socialt marginaliseret – altid...?* I initially hoped to obtain information from the different registers covering the period from 1988 on onward. Thus obtaining information on the whole period after the people stayed at a shelter and up till the conducting of the study. This, however, was not possible – as it has been outlined above in our presentation of the data used. The different registers covered different periods of time. For example, the information available on treatment for the use of illegal substances covered two years only. Every year another period was added and the oldest deleted. The information on criminal record covered the time from a first encounter and up till present. The information on income, connection to the labour market, types of residence, and family relations – information obtained from Statistics Denmark – were available for the period from 1990 to 1996 – at least if the information should stay payable. The information from Statistics Denmark would have been available for a longer period, but not within our budget. This question of financing is an issue that I shall not touch upon any further, suffice it to say that the prices I paid did vary significantly and were of a size that necessitated considerations.

¹² But does so within a large set of general limits ensuring a high degree of safety regarding the keeping of and passing on of information.

¹³ It can be added that the police as well had a different price – a price that cooperate the economic assumption that actors being close to having market control set their prices to the disadvantage of the consumer.

Only some information exist

When working with register-based data one has to accept that what information exists is something beyond the actual research project. The registrations are results of practices established prior to the actual project, of practices that have been in existence over time, and (often) of practices that have been created for different purposes than that of the research project. Consequently the questions asked by the researcher might focus upon something that differs somewhat from what is registered. This can be illustrated by pairing some of the research questions posed above:

Questions asked

Does a given respondent have a satisfying place to stay according to her or him self or according to some politically accepted standards?

Does a given respondent have the finances necessary for sustaining acceptable living conditions according to him or her self or according to politically acceptable standards?

Some of the obtained information from registers

- Formal address
- Which other people formally living on a subject's address
- Official type of housing that the formal address is registered as

- Eventual transfer income
- Gross annual income
- Formal connection to the labour market

If I first look at the question regarding the place for staying one of the key distinctions seem to me to be *formally versus actual*. Of course it is interesting to know whether a person formally have a place useable for permanent residency, but in evaluating the changes in living conditions it seems more of interest to know whether a person have *actual* and *secured* access to a permanent place to live. In regard to the question of financial possibilities I encounter similar problem but here they might be captured by a distinction between *official versus actual*. Again, what I am interested in is information about the actual economic situation of the people in the sample, but what I can find information about is the official financial situation of the people, which might be significantly different from the actual situation. Or in other words: I am more interested in the actual lived life than in the officially registered life – and the registers tell us only about the latter.¹⁴

Thus, a significant problem is that the information that I wanted was not there to be found in registers. Had one conducting a survey one could decide to ask the house such questions that the researcher could theoretically argue would satisfy the need for information necessary for answering

such questions,¹⁵ or at least ask questions that would bring us as close as one would theoretically think it possible – but that cannot be done when doing register-based research. When doing register based research the researcher is in a sort of take-it-or-leave-it situation, and this raises significant questions concerning the reliability of the register based research.

It can be added, that there were some aspects that would have been very interesting for us to look closer at, but about which no information existed. Thus there was no registration of the use of shelter among the people included in the population for any period other than the one in 1988-89 where the subjects were initially identified. That means that I could not know whether a person had used a shelter during any of the following years, and perhaps done so while officially living at an ordinary address.

As I could not always obtain information directly on the questions that I where interested in I instead tried to combine information from different registers in an attempt at fitting the information to the questions asked. An example can be found in the attempt at figuring out the type of dwelling in which a person formally lived at any given point in time, thereby attempting to estimate how settled into a regular permanent dwelling the people in the population were. For this I combined information on the composition of the housing unit in which a person lived (the number of adults and children) with information on the purpose of the housing unit (was it officially for business, regular living, or an institution (see table 5 above)). By combining these information I hoped to be able to isolate the types of dwellings with, e.g., institution-like features from the core-family-housing unit.

Another example can be found in the attempt at establishing the degree of contact to the labour market. Here I had information about the degree of unemployment: during any year the degree of unemployment should be between 0000 and 1000 with 0000 being equal to no unemployment. However, a person who would not have been part of the labour force at all during a given year had not been unemployed either, and was thus also registered with the value 0000. Trying to separate the people not having experienced unemployment but who were part of the labour force from people who had not been unemployed but who had also withdrawn from the labour force I

¹⁴ Below I shall enhance the problems of this distinction be touching upon the long and uncertain path that information travels before they end up in registers used for research.

¹⁵ There are of course many methodological considerations that complicate this picture – but I shall leave such untouched here.

combined the registered degree of unemployment with a registration of the job position held during one week in November. The logic was that did one not possess a job during this week one could not have been employed throughout the year, thus the 0000 would indicate a withdrawal from the labour force (see table 8 above).

Both these examples illustrate that for getting the information I found interesting I had to mingle through the available register-based information. In itself this should not necessarily pose a problem – after all the logic about the connection to the labour market in itself seems without flaws, and in regard to the type of dwelling in which one lives I am simply talking about operationalizations which should be needed anywhere – right? However, it might not be as simple: perhaps the logic is flawless, but the existence of an uncertainty concerning the relationship between a phenomenon and the registration of a phenomenon might make the logical neatness somewhat redundant.

The construction of register information

In analyses based on register information the methodological argumentation often concerns the mathematical logic used to approximate distributions in the material: one often finds reflections on the application of proper statistical methods given the *registration of* conditions. It is a question of whether the *registration* of a phenomenon fits the criteria for a given calculation, e.g. not too few or too skewed distribution of observation. It is not my intention here to argue against or around this mathematical logic providing the foundation for the statistical calculations. Rather, I shall emphasize a point that no one will probably disagree about, but still a point that surprisingly little attention has been directed towards – at least in a Danish context. I shall emphasize that the methodological reflections on the clear and beautiful logic of the math is only the superstructure on a basis of subjectivity and categorizations (e.g. Bourdieu, 1984; Ronkainen, 1998; Stax, 1999). But, it is a superstructure that, as an ideological hegemony, hides the constructed-ness, and subjective categorization of this basis. The issue of validity is reduced to calculations of models, probabilities and confidence intervals – or reduced to an issue answerable within a discourse of statistics. The question of the relationship between a phenomenon and the transformation and categorization of the phenomenon is, on the other hand, frequently left unreflected upon.¹⁶

¹⁶ I base this conclusion on a preliminary reading of publications using register-based information from The Danish National Institute of Social research. In none of these I find reflections upon the issue of how the data are generated.

The issue of the construction of information becomes especially significant due to the fact that the information drawn from the registers are constructed in many different ways. Some information is registered as part of the general job of, e.g., a social worker. This is, for example, the case when a financial transfer of some kind is provided to a client. Here the social worker either writes a cheque to a client or makes a transfer to an account. In both cases the registration of the sum of money is a product of the function that the social worker fulfils as part of her job. This information then goes to the taxation offices that use them in the calculation of the taxes levied on the client. Such information might be believed to be rather accurate: the amount transferred to the client probably matches the amount reportedly transferred to the client (at least when the systems are coordinated so that only one entry of the sum transferred by the social worker is necessary) – which has nothing to do with whether the amount of transfer income that the client has received is more or less than he or she is entitled to.

However, other types of information are not registered as a part of the actual job functions carried out by the social worker or others. Sometimes registers are constructed on information passed on from a local level to a central level in the form of filled out questionnaires. In a previous study on special housing units for socially excluded Stax tried to use information from Statistics Denmark on the number of shared dwellings established in accordance with the – now changed – section 68b in the old social legislation. This information was gathered by Statistics Denmark through a questionnaire sent to the municipalities in Denmark. Similarly, a couple of years ago Statistics Denmark provided information on the number of shelters, the number of beds, and the number of people using these during one week each January. These figures were created through reports submitted by the shelters themselves and then in the end collected at Statistics Denmark. The creating of such information was not a necessary part of the daily job functions of people working in the shelter, but rather an extra reporting being requested during that one week. And this way of constructing registers raises a lot of questions concerning validity and reliability. For example all the kind of questions that one would pose to a questionnaire – or one should pose to a questionnaire

Such reports might state which registers and which variables they included in their analysis. But the information is then treated as if they were the actual events, not registrations of such.

When I have presented this point of criticism during my working with the project I have been countered with statements like 'that I am aware of, and I all know the relationship is just an assumption...' I do however still consider my point to pose problems: first because I have not found the assumption explicated in publications, and second because assuming

– did one conduct a survey. But when doing register-based research these questions are often left unasked – perhaps because the data, by the time the researcher becomes involved, are already neatly organized into seemingly unambiguous variables, and the methods used for creating the information are something that one will have to dig for – if one is interested and has the resources needed. With a very limited focus upon the construction of information in registers it seems that the register-based research is relying on an understanding of an evident and unproblematic correspondence between a phenomenon and the registration hereof. Such understanding must be accepted for limiting issues of validity and reliability to questions concerning the statistical logic used in the treatment of the registrations of the phenomenon. (e.g. Kvale, 1997; Hahn, 1990; Kovar & Royston, 1990; Andersen & Gamdrup, 1990).¹⁷

In *Én gang socialt marginaliseret – altid ...?* I encountered some uncertainty with how the included information was created. Recalling that “one should – when taking over already existing data – ask for the purpose of the original collection of data, ask for the previous use of the data, and ask for the conditions under which the data was created” (Andersen & Gamdrup, 1990)¹⁸ I tried to find out how the information analysed was created and transformed prior to our involvement. But this was rather difficult – if not impossible – to find out. Some of the variables included in the study had travelled far and wide, and it was not always possible to track down how the information had been gathered, transferred, and transformed before ending up as my data. Sometimes a person had left for another job leaving the one who had taken over without knowledge of the procedures used, sometimes a transformation of information had been made by a different department and it was uncertain who had done what, sometimes those organising the registers centrally had only a limited idea of how the information were gathered locally. This is not to be understood as a criticism of the concrete practice at the institutions supplying the statistical information – rather as pointing out a condition for register-based information. It might of course be possible to change procedures to avoid some of these aspects – e.g., higher salaries might prevent people from seeking new jobs, etc. – but intervening in such partial manner does not relate to the central aspects of the criticism. The central point concerns the idea of unambiguous gathering, transferring, transformation, and use of information. The critique springs from scepticism towards the understanding of correspondence

is not enough. I do find that an assumption should be sustained and strengthened by a theoretically grounded argument for its necessity or its use.

¹⁷ It is interesting to note that the texts discussing the issue are somewhat critical hereof. But in research using register based data I cannot find reflections, or references to reflections, on the correspondence theory.

¹⁸ My translation.

theory – from scepticism towards the idea of unambiguous categorizations, language, and interpretation of information. It is scepticism grounded more in a philosophical understanding of science (inspired by, e.g., Garfinkel, 1967 and Bourdieu, 1977; 1990) rather than in an understanding of the problems being consequences of the specific organization of relevant agencies.

In the process of trying to interpret and understand the registered information I stumbled over outcomes on variables that should not have been possible: In the study I was interested in finding out the number of adults (Var. A) living in household with at least one person from my population. I had information on the number of persons (Var. P) and information on the number of children (Var. C) at each household, and by subtracting the latter from the former I should obtain the wanted information – at least logically ($\text{Var. P} - \text{Var. C} = \text{Var. A}$). I did the calculation for the seven years from which I had information and a funny result appeared: The number of households with a number of kids equal to or larger than the number of persons ($\text{Var. C} = \text{Var. P}$) increase throughout the period. Should one find it worth testing the development in the quality of the registers and measure quality with the number of household with and erroneous composition of kids and persons between 1990 and 1996 one would, with my data, find what the quality was significantly ($p < 0,001$) decreasing over time.

I furthermore stumbled over suspicious changes on variables: On a variable that should register whether a person was undergoing education I had a value that no one could tell me what meant. This value appeared with a very unequal frequency: from about 15 percent to 0 percent to 95 percent of the population in different years.

I also stumbled over systematic changes in distributions: In regard to income a group of about 20 percent was registered with a gross income of 0 D.kr. per year for the first four of the seven years looked at. But then there was a sudden decrease in the ratio of people with no income, and in the last three years this group constitutes about four percent.

But the interesting aspect – besides opening my eyes – was how answers within the discourse of correspondence theory were absent. For each question an answer that referred to some unique – and for the process of register-based research external – circumstances was provided. Sometimes the explanation was a programming error, sometimes it was organisational changes, but in the very end it was my population: After all, my population consisted of weird people, I was told. I. Lakatos (1970) once wrote a brief story that I here find it worth quoting in length:

“The story is about an imaginary case of planetary misbehavior. A physicist of the pre Einsteinian era takes Newton’s law of gravitation, N , the accepted initial conditions, I , and calculates, with their help, the path of a newly discovered small planet, p . But the planet deviates from the calculated path. Does our Newtonian physicist consider that the deviation was forbidden by Newton’s theory and therefore that, once established, it refutes the theory N ? No. He suggests that there must be a hitherto unknown planet $P1$, which perturbs the path of p . He calculates the mass, orbit, etc. of this hypothetical planet and then asks an experimental astronomer to test his hypothesis. The planet $P1$ is so small that even the biggest available telescope cannot possibly observe it; the experimental astronomer applies for a research grant to build yet a bigger one. In three years time, the new telescope is ready. Were the unknown planet $P1$ to be discovered, it would be hailed as a new victory of Newtonian science. But it is not. Does our scientist abandon Newton’s theory and his idea of the perturbing planet? No. He suggests that a cloud of cosmic dust hides the planet from us. He calculates the location and properties of this cloud and ask for a research grant to send up a satellite to test his calculations. Were the satellite's instrument to record the existence of the conjectural cloud, the result would be hailed as an outstanding victory for Newtonian science. But the cloud is not found. Does our scientist abandon Newton's theory, together with the idea of the perturbing planet and the idea of the cloud which hides it? No. He suggests that there is some magnetic field in that region of the universe which disturbed the instruments of the satellite. A new satellite is sent up. Were the magnetic field to be found, Newtonians would celebrate a sensational victory. But it is not. Is this regarded as a refutation of Newtonian science? No. Either yet another ingenious auxiliary hypothesis is proposed or ... the whole story is buried in the dusty volumes of periodicals and the story never mentioned again.” (p.100-101)

This defensive strategy, illustrated by Lakatos, do highlight some of the problems that I have encountered in my working with register-based information. But it is a problem that can be dealt with first and foremost through recognition.

A project under way: attempting to reconstruct trajectories¹⁹

Currently The Danish National Institute of Social Research is carrying out a second register based research project on a homeless population. This project is still in the phase of collecting data and thus no results are available, nor are many concrete experiences concerning the actual development of the project. I shall hope to be able to include some of the experiences from this project in this paper before I finalize it for the 4th workshop. I will, however, here outline the project as it tries to take the use of register based research further than it was done in the above discussed study, and as

¹⁹ I will like to thank I. Koch-Nielsen for help with this part of the paper.

the sample of the study is constructed with this in mind, i.e. is made significantly larger than it was the case with the study *En gang socialt marginaliseret – altid ...?*

The Sample and the data

This study is a follow-up study, where persons who had been registered as users of shelters²⁰ in at least one of six²¹ Danish regional counties in 1997 will be included in a register created at The Danish National Institute of Social Research. Intentionally everybody having used a shelter within any of the six counties during 1997 should be included – independent of the length of stay. The counties were chosen so that they would include the biggest cities in Denmark (Copenhagen, Aarhus, Aalborg and Odense). The register containing the sampled people was created by visiting each shelter in the six counties, and writing down the CPR-numbers of clients from 1997 from the shelters own internal file systems. These files were of different quality, but almost everywhere did they exist in 1997 – one exception was the largest shelter in Copenhagen, where we could only obtain information for half the year of 1997. All together approximately 4.000 CPR-numbers were registered. These approximately 4.000 numbers were run through The National Register, which resulted in a sample of a little more than 3.000 persons who were alive and whose addresses could be traced. From among the 3.000 about half was then selected for interviewing – an additional dimension of this study where we try to obtain peoples own assessments of the various social services they have encountered during their life on the margin. The sample in the study is outlined in table 9.

Table 9. The sample

	Males		Females		All	
	Number	%	Number	%	Number	%
Own housing	1.843	52,9	793	67,3	2.636	56,5
Actual shelters or housing arrangement targeted on homeless people	309	8,9	39	3,3	348	7,5
Dead	528	15,2	77	6,5	605	13,0
The rest	803	23,1	270	22,9	1073	23,0

²⁰ Here excluding centers for battered women.

²¹ Out of a total on 14 Counties plus the cities of Copenhagen and Frederiksberg, which have a special status as they are not actually part of a county, but instead are municipalities having taken on the obligations of the Regional Counties as well.

I alt	3.483	100,0	1.179	100,0	4.662	100,0
-------	-------	-------	-------	-------	-------	-------

[I am currently confused of these figures? Were no invalid CPR-numbers found? Or are they ‘rest’? How many are we trying to obtain information from registers about? All 4.662? I shall make this clear at a later time and prior to the finalization of the paper.]

About the people included in the sample we shall then draw upon information from Statistics Denmark on altogether 417 variables from 20 registers, e.g. a register on mobility, a register on housing-units, a register on activation, a register on hospitalisations, and a register on transfer incomes. The information will be obtained for as long a period back in time as possible. In regard to some of the information this is back until 1980. In addition we, by now, have a continual registration of users of shelters, and this registration includes the client’s CPR-number. This registration has been in place since 2000. Thus it is now possible find out whether any of the people included in the sample have been using a shelter in 2000 or later.

The research questions

Through the project we hope to be able to answer the following questions: How many of the former users of shelters can be said to have achieved a more stable life (or to have exited from homelessness), and how many are still or again experiencing such spell? This question is somewhat similar to some of the questions asked by Stax. But with this project we hope to be able to say more about the different trajectories lived by the sampled people. For example, is it possible to identify typical trajectories leading to more or less marginal positions? For answering such questions - when we are dealing with complex issues where many variables are needed together for obtaining a reasonable picture – we are in need of a large sample. In *Én gang socialt marginaliseret – altid ...?* it was not really possible to identify trajectories over time – even though we, in regard to some of the information gathered, had information on a whole period between 1990-1996. Too few people could actually be followed. The many had missing information on one or more of the variables that were used for one or more of the years that we would have liked to look upon. With this study we hope that the sample has been made large enough for such analysis, and that the quality of the register-based information has been improved. But it is yet to early to know.

References

- Andersen, V & P. Gamdrup (1990): Om problemformulering og projektarbejde. In Andersen, H. *Videnskabsteori og metodelære*. Frederiksberg, DK: Samfundslitteratur.
- Bourdieu, P. (1977): *Outline of a Theory of Practice*. Cambridge, UK: Cambridge University Press.
- Bourdieu, P. (1984): *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.
- Bourdieu, P. (1990): *The Logic of Practice*. Cambridge, UK: Polity Press.
- Brandt, P. (1992): *Yngre hjemløse i København*. Copenhagen, DK: FADL's forlag.
- Garfinkel, H. (1967): *Studies in Ethnomethodology*. Cambridge, UK: Polity Press.
- Hahn, R.A.(1990): Comment on Suchman & Jordan (1990). *Journal of the American Statistical Association*. Vol.85:409:244-246.
- Johnson, T.P. (2003): Glossary. *Substance Use & Misuse*. Vol. 38:3-6.
- Kovar & Royston, (1990): Comment on Suchman & Jordan (1990). *Journal of the American Statistical Association*. Vol.85:409:246-247.
- Kvale, S. (1997): *InterView*. Copenhagen, DK: Hans Reitzels Forlag.
- Lakatos, I. (1970) Falsification and the Methodology of Scientific Research Programmes. In Lakatos, I. & A. Musgrave (eds.): *Criticism and the Growth of Knowledge*. London, UK: Cambridge University Press.
- Ronkainen. S. (1998): Impossibility of Reality and Possibility of the Survey as a Construction. *Nordiske Udkast*. #2:51-61.
- Statistics Denmark, (2000): *Varedeklaration for Befolkningen: Pr. 1. januar, befolkningens bevægelser, befolkningen i amter, kommuner*. www.dst.dk.
- Stax T.B. (1999): *Én gang socialt marginaliseret – altid ... ?* Copenhagen, DK: Socialforskningsinstituttet (99:21).
- Stax, T.B. (2003): Estimating the use of illegal drugs among homeless people using shelters in Denmark. *Substance Use & Misuse*. Vol. 38:3-6.
- Stax, T.B. & J. Kæmpe (1999): *Support and Housing - two complementary aspects of a social policy: National Report for FEANTSA, 1998*. Brüssel, B: FEANTSA