



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR WORKING PAPER WP 2010-017
FEBRUARY 2010

Real and synthetic household populations and their analysis: an example of early historical micro-census data (Rostock, 1819)

Siegfried Gruber (gruber@demogr.mpg.de)
Rembrandt D. Scholz (scholz@demogr.mpg.de)
Mikołaj Szoltysek (szoltysek@demogr.mpg.de)

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

Real and synthetic household populations and their analysis: an example of early historical micro-census data (Rostock, 1819)

Siegfried Gruber, Rembrandt D. Scholz, Mikolaj Szoltysek*

Abstract

The main purpose of this paper is to report on an initial validation of methods for dealing with micro-census data with no delineated households. After describing the 1819 census of Rostock we test the possibilities of using an algorithm that creates households according to a strictly defined set of rules. The census of 1867 will be taken as our reference point for designing such rules of assigning people to household units and for assessing the appropriateness of the algorithm's fit to the census of 1819. In the final part we discuss the outcome of the algorithm for different groups within the urban population and the strengths and weaknesses of this approach.

Keywords: algorithm, census, delineating household

Rostock censuses 1819 and 1867:

The two dates of 1819 and 1867 mark two censuses carried out in Mecklenburg, and they offer valuable early data. The 1819 Census of Mecklenburg is one of the oldest surviving individual-level data population censuses in Germany. For the first time ever, the total population of Mecklenburg-Schwerin was quantitatively and qualitatively recorded. The census was ordered by the Grand Duke of Mecklenburg-Schwerin. The purpose of the census was to determine the exact military contingent within the army of the German Confederation (Deutscher Bund). The lists were to constitute a comprehensive register "of every person living on the day of the census, as young or as old as they may be, of every gender, religion, trade or status" (translation from *Wochenblatt* 1819, 67). Census enumerators were instructed to visit every household in order to list every person living in that household. The lists were due to be submitted by August 25th. However, taking the census took more time than expected, and the closing date was delayed to mid-November, and then again to early December. The last survey questionnaires were completed as late as in February 1820 (Manke 1999, 651-653).

Because of the wealth of information collected in the census of 1819 (sex, first name, last name, day of birth, place of birth, parish of birth, relationship to household head or occupation, property ownership, duration of residence, marital status, religion), this census is

* Laboratory of Historical Demography, Max Planck Institute for Demographic Research, Rostock, Germany. Address all communications to: Gruber@demogr.mpg.de or Szoltysek@demogr.mpg.de.

widely known for being a leading German population census of that time (Tscharnke 1943, 29). It was followed by the 1867 Census, which used modern refined population-counting methods, which were first implemented in Germany in the 1860s. The 1867 census of the Grand Duchy of Mecklenburg-Schwerin was taken on the night of December 2-3, 1867. That night authorised enumerators visited every house in the district assigned to them, listing every person staying there at that moment in time. The head of each household was responsible for the correct completion of the census form. The census was taken based on the rules used by the North German Confederation (Norddeutscher Bund) and the German Customs Union (Deutscher Zollverein), which the Grand Duchy of Mecklenburg-Schwerin became part of in 1867/68. The census was supposed to gain an overview of the taxable and conscriptable population (Manke 2005, 462-467). Both censuses provide a wealth of information and, making certain allowances for the specificity of the 1819 data, can be used to conduct a structural analysis of the family system (see, however, certain reservations discussed below).

In analysing changes in household structures over time using two separate population censuses, a constant effort has to be made to ensure that we compare actually the same unit called “household”. Even slight differences in definitions of a household will yield different results (Schmid 1988, 14). In this regard, despite the relative wealth of material that is available to us, we face certain difficulties. Whereas the census of 1867 features borders between households that were clearly assigned during the process of collecting the census data, the census of 1819 delineated no such household borders in the census manuscripts. The 1819 census is a list of inhabitants, with no clear designation of where one household ends and where a new household starts. There is also no information about the addresses of the people in the census manuscripts.

How can we cope with these data insufficiencies? Do they rule out the achievement of our research goals? Is it possible to invent a realistic and meaningful way to delineate households in the 1819 census? A household has been defined by Laslett as a co-resident domestic group, as a “series of names of individuals in blocks, with clear indications of where one block ended and the next began” (Laslett 1972a, 24). This co-resident domestic group has three basic characteristics: The members of the group sleep under the same roof, share a number of activities, and are related to each other by blood or by marriage. In addition, the group may occasionally include non-related persons—like servants, visitors, boarders, or lodgers—as members of a household (Laslett 1972a, 25). Later definitions presented by Wall concentrate on the first two characteristics (Wall 2001). Households are similarly defined in contemporary demographic discourse (Schmid 1988, 14-15; Ermisch 1988, 23-26). The

apparent straightforwardness of the definitional approach notwithstanding, extracting household data from historical micro-census counts can be a complicated and confusing task. The extreme difficulties we face are similar to the difficulties faced by other scholars working in the field (Berkner 1972, 1975; Hammel 1984; Hammel & Wachter 1996a, 1996b; Sovic 2008).

Delimiting households:

The city's population is represented by households in a published version of the Rostock census of 1819, based on a database which has been created for historical research (Manke 2005a). These "households" were "created" during the data entry of the census based on the characteristics of a person's last name, marital status, sex, property, relationship to household head and occupation (Manke 1997, 131). A household defined in this way would contain the nuclear family of parents and unmarried or economically dependent children, co-resident relatives, personnel (domestic servants, employees) and other persons (e.g., boarders and lodgers) (Manke 2005a, 18). The basic criteria for being a head of a new household were the following:

- being adult,
- having no relationship to a member of the previous household or family, and
- being no immediate dependent employee.

Another decisive factor for heading a household was having an income while not being a live-in employee like a servant. Manke states that this was relatively unproblematic (Manke 1997, 131, endnote 12), but later he is more sceptical about the effect of the absence of household borders (Manke 1999, 651, 2005b, 458). In a book about the city of Rostock between 1750 and 1850, he again downplays this problem (Manke 2000, 346). A possible inclusion as mother-in-law or father-in-law, boarder or lodger would have been arbitrary and not according to the source or the current state of research. Therefore there is a high number of one-person-households, very often lone elderly people (Manke 2005a, 19).

Schlumbohm found a similar census without defined households of 1811/1812 for his research in Northwestern Germany. He concludes that the order of the people allows to create groups of parents, children and servants. Problems arise with people like widowed persons or older couples. Therefore he used this census only as a supplementary source (Schlumbohm 1994, 37f., 195).

Although the rules might present themselves as relatively straightforward, they actually led to an overcounting of the number of one-person households, very often lone elderly people (Table 2; 1819 File A). This fact was recognised by Manke, but was not considered a problem (Manke 2000, 19, 346). As a consequence, people who lived in a poorhouse or soldiers were treated as individual households in this file, and also in the published edition of the census. This is why a second file was created, which allowed for the creation of institutional households and the inclusion of more people into the previous households. The factor of having an income was considered in a more restricted way, and therefore fewer persons qualified for heading an independent household (see 1819 File B in Table 2).

An algorithm for delimiting households

In the next stage, we decided to test the quality of both files with an algorithm that creates households according to a strictly defined set of rules. The census of 1867, which was not affected by the definitional problems discussed so far, was taken as our reference point for designing such rules of assigning people to household units and for assessing the appropriateness of the algorithm's fit to the "real" data structures. Artificial household structures simulated along a common set of rules for both the 1819 and 1867 censuses have yet another advantage. We hope to make these two data sets more comparative by imposing standard scenarios of household membership on undifferentiated groups of individuals in both enumerations.

Our algorithm was developed in a two-fold process, which went back and forth between designing additional rules and the thorough comparison of real and simulated structures. We conducted a number of experiments using various scenarios for assigning individuals to domestic groups, with the goal of obtaining the most satisfactory match with the 1867 census, and, if possible, with the 1819 census as well. The first assumption for dealing with this problem is, that the members of a household were registered consecutively. This procedure is confirmed by other scholars, although the German Customs Union (Deutscher Zollverein) obliged its member states to count the population according to households only in 1843 (Manke 2005b, 457). We have actually found only a few households, where some household heads were registered first and then all the members of their households together. But their number is too small to affect the outcome of the algorithm. The order of the persons within a household was not always like a sequence household

head/wife/children/servants (Manke 2005b, 458), and therefore the algorithm could not be build upon such a sequence.

After several modifications, our household membership rules came to consist of the following principles:

a person belongs to the previous household if –

- the family name is the same,
- the occupational title belongs to a list of occupational titles indicating co-residence (e.g., servant, apprentice, journeyman) and the person is unmarried,
- the relationship to the household head indicates this,
- the occupational title is the same as the previous one and the person is not married,
- the person is absent at the time of the census,
- there are indications that the person belongs to an institutional household (e.g., poorhouse, home for the mentally ill)
- the person is unmarried and below age 20,
- the person is an unmarried woman below age 25,
- there is no information about the relationship to the household head and no occupational information,
- the person is a married woman and there is no information about the relationship of the person to the household head, or
- the person is an unmarried woman between the ages of 20 and 39 and there is no occupational information about her.

The first version of the algorithm consisted only of the four rules at the beginning of the list presented above. We improved the algorithm by refining the second rule and adding additional rules and could reduce the amount of mismatches of the first version by about 50%. Table 1 presents the results of the algorithm, which are quite fine for the file of the 1867 census for the first two measures, but are much less satisfactory for the overall measure of complete matches of households. The 1819 files have results which reflect the number of households: The larger number of households in File A leads to a higher number of household heads who are not found by the algorithm, and a lower number of additional household heads created by the algorithm. The overall fit is better for File B. Better still, we can be sure that

more than 80% of all households in all files are defined correctly, and thus for most of the results there are only minor effects of using different household limits.

Table 1: Results of the algorithm, version 7

Census file	1819 file A	1819 file B	1867
Number of households	4,098	3,601	6,826
Number of households according to algorithm	3,832	3,832	6,694
Household heads not found	421	195	494
Percentage	10.3 %	5.4 %	7.2 %
Household heads found in both files	3,677	3,406	6,332
Perce	89.7 %	94.6 %	92.8 %
Additional household heads created by the algorithm	155	426	362
Percentage	3.8 %	11.8 %	5.3 %
Complete matches of households: not matched	843	700	1,146
Percentage	20.6 %	19.4 %	16.8 %

Household structure

In a second step, we focus on assessing the effects of the algorithm on household structures, and compare household types of all three files with the typology of the households simulated by the algorithm. There is almost no difference between real and simulated household structures for the 1867 census, as can be seen from Table 2. The only minor difference is a slight increase in the number of solitaries at the expense of simple family households.

The algorithm for the census of 1819 resembles File B to a much greater extent, with the exception of solitaries. Almost similar proportions of simple and extended family households, along with almost absolute sameness of results for other types of domestic groups, seem to be critical here. Nonetheless, the observed mismatch of data for solitary households between simulated files and groupings from File B for the 1819 census has two implications. First, since we know from the file of 1867 that the algorithm generally overestimates the number of solitaries and underestimates the number of simple family households by approximately 2%, we may think of the discrepancy between the two files as being a little bit smaller than is suggested. The different versions of the algorithm yield quite similar results, there is only a slight tendency towards more simple and extended family households at the expense of solitaries during this process.

On the other hand, our knowledge of the literature and the existing data sets lead us to believe that the proportion of solitaries in File A for 1819 is too high. Litchfield's study of

19th-century cities found only one case in which the share of households of people living alone or with non-kin only was 18%. All the others had lower percentages (Litchfield 1988; see also Reher 1987). The highest proportion of solitary households ever registered comes from the 1802 census of Reims, where they constituted 19% of all domestic groups (Fauve-Chamoux 1983, 481; see also Duben & Behar 1991, 41). The high proportion of solitaries in File A would exceed all these examples by far.

Table 2: Comparison of percentages of household types

	1819 file A	1819 file B	1819 algorithm, version 7	1867 file	1867 algorithm, version 7
1	27.4	17.4	24.1	17.7	19.7
2	5.4	2.0	2.3	2.7	2.7
3	53.4	75.3	69.1	70.5	68.6
4	13.5	5.1	4.4	8.2	8.7
5	0.1	0.1	0.0	0.1	0.0
6	0.2	0.2	0.2	0.8	0.2
N	4,098	3,601	3,832	6,826	6,694

Note: The household typology is the one suggested by Peter Laslett (Laslett 1972, 31). Type 6 includes institutional households.

Different population groups and the algorithm

We have seen that the overall match of the algorithm with the census data files is quite good, but we have to be aware that there are considerable differences within the population. Therefore we should know for which subpopulations the algorithm works better than for other ones. The algorithm matches male household heads much better than female household heads. Married household heads have the highest rates of matches, while widowed household heads have also low rates of household heads not found by the algorithm, but much higher rates of additional household heads created by the algorithm for the data files of 1819. Unmarried household heads have very high rates of mismatches, either as household heads not found by the algorithm or as additionally created household heads by the algorithm. This pattern of matches by marital status is also reflected by the proportion of matches according to the age group of the household heads. Young household heads have the highest rates of mismatches, while middle-aged people have the lowest rates of mismatches. These patterns are very similar for all three data files, only the extremes are more pronounced for the data files of 1819.

Table 3: Matches of household heads of the algorithm by personal characteristics (version 7)

	Complete match %	Household heads not found %	Additional household heads created by the algorithm %	Number of household heads in the data file
1819 File A:				
Male	91.8	8.2	2.0	3,098
Female	83.4	16.6	9.2	1,000
Unmarried	52.6	47.4	7.3	715
Married	97.5	2.5	0.9	2,472
Widowed	97.8	2.2	8.7	911
20-29 years	63.0	37.0	6.8	468
30-39 years	94.6	5.4	3.5	896
40-49 years	95.6	4.4	2.0	858
50-59 years	95.1	4.9	2.0	812
60-69 years	92.2	7.8	4.6	562
70+ years	89.6	10.4	7.8	385
Overall	89.7	10.3	3.8	4,098
1819 File B:				
Male	96.6	3.4	4.6	2,873
Female	86.8	13.2	40.2	728
Unmarried	70.9	29.1	27.7	437
Married	97.6	2.4	1.0	2,466
Widowed	98.7	1.3	40.3	698
20-29 years	82.9	17.1	17.1	328
30-39 years	96.5	3.5	6.4	854
40-49 years	96.7	3.3	4.7	824
50-59 years	96.9	3.1	7.6	754
60-69 years	96.1	3.9	20.9	465
70+ years	91.9	8.1	40.1	284
Overall	94.6	5.4	11.8	3,601
1867:				
Male	94.6	5.4	5.7	5,146
Female	87.1	12.9	4.0	1,680
Unmarried	68.8	31.2	27.4	919
Married	96.8	3.2	1.6	4,513
Widowed	95.7	4.3	2.2	1,335
20-29 years	72.4	27.6	29.7	602
30-39 years	95.0	5.0	4.4	1,771
40-49 years	95.7	4.3	2.5	1,715
50-59 years	95.3	4.7	2.4	1,226
60-69 years	95.3	4.7	1.7	969
70+ years	93.7	6.3	2.8	493

Overall	92.8	7.2	5.3	6,826
---------	------	-----	-----	-------

Note: The number of household heads may not sum up to the overall number because of missing information about gender, marital status, or age.

We shall look now at complete matches of households, i.e. the first person of the respective household and the first person of the next household match in the data file with the algorithm. The fit according to household size shows a U-shape: high proportions of mismatches of one-person-households, low proportions of mismatches for smaller households, and increasing rates of mismatches for larger households. The fit of the algorithm according to the household typology points to the fact of complexity. The best fit can be achieved for simple family households, followed by extended family households. Solitaries and no families have worse rates of matches than these two types. There are only very few multiple family households, households of undetermined type, or institutional households. Overall we can conclude that the algorithm is quite good in detecting simple family households or extended family households. Households without a couple or a parent-child relationship have much worse results for matching the data file with the algorithm. Servants do not increase the rates of mismatches significantly, while other non-kin (lodgers and others) have such an effect.

Table 4: Complete match of households: percentage not matched by household size and household type (algorithm, version 7)

	1819 file A		1819 file B		1867	
	%	N	%	N	%	N
Household size:						
1	42.7	910	28.3	396	30.5	894
2	13.9	669	12.7	623	13.2	1057
3	12.1	636	14.6	609	12.0	1219
4	12.7	561	18.8	580	12.1	1093
5	14.5	441	18.1	442	13.6	814
6	16.1	317	20.7	324	14.6	595
7	16.3	208	25.3	237	17.8	370
8+	18.0	356	26.7	390	24.5	784
Household type:						
1	38.8	1,123	28.8	626	29.0	1,208
2	19.9	221	29.2	72	25.8	186
3	12.2	2,190	16.0	2,710	12.7	4,811
4	15.8	552	33.5	182	15.5	561
5	40.0	5	20.0	5	25.0	4
6	100.0	7	50.0	6	91.1	56
Overall	20.6	4098	19.4	3601	16.8	6826

Note: Household type 6 includes also institutional households.

Conclusions

We see that there is quite a considerable level of mismatches of complete households according to table 4, while the differences for different household types are much smaller according to table 2. The results at an aggregate level are quite reliable, while the results for single households are much more questionable. Therefore these results can only be used for analysis on an aggregate level and not for investigating individual persons or households.

The results are better for the census of 1867 compared to the one for 1819. The main factor for the different results of the algorithm for both censuses is clearly that less information is available in the census of 1819 (see table 5). The more information is available, the better we can design rules for an algorithm and fewer cases fall into the category of “unknown”, which is the category causing more troubles than the other ones.

Table 5: Available information in the census

	1819		1867	
	men	women	men	Women
Relationship to household head	41.2 %	77.7 %	92.9 %	93.3 %
Occupational title	66.2 %	26.6 %	79.2 %	40.1 %

Later versions of the algorithm improved the fit of the algorithm to the census data files, but the improvements became smaller and smaller. Additional problems arise in the fact that an improvement in matching household heads in the data files leads generally to an increase in creating additional household heads by the algorithm. We are somehow trapped in between increasing the matches for household heads (plus creating additional “wrong” household heads) and decreasing the number of additional “wrong” household heads (plus decreasing matches of “real” household heads). In addition some improvements for the 1867 data file turned out to have negative effects on the fit for the 1819 data files.

Further improvements will require either a high investment of time in designing additional rules or refining existing ones. This could partly be done on a more thorough use of the occupational information. Another possibility would be the creation of new kinds of rules, like the inclusion of a random factor for groups of people for whom we cannot define clear rules for being household head or not being household head, but something like “about half of them are household heads”.

The algorithm works best for the “standard” household in Rostock: a simple family household with a married or widowed male household head. The matches are worse for all other kinds of households and household heads. This can on the one hand be caused by a (not intended) cultural bias of the persons designing the algorithm, but the reason is more probably that couples and parent-child relationships are more easily caught by any kind of algorithm compared to other kin or non-kin relationships in a given household.

Despite all shortcomings the use of such an algorithm can help decide about using one or the other census data file applying different rules for delimiting households by individual scholars. A census data file with existing household borders in the source material can serve as a reference point in testing an algorithm and then the algorithm can help making a decision about which data file or rules for delimiting households to use.

Notes:

We thank Josh Goldstein (MPIDR) for directing our attention to this possibility for overcoming the deficiencies of the 1819 census.

An earlier version of this article was presented as a part of a paper at the SSHA conference in 2009 and later published as a working paper (Szotlysek et al. 2009).

References:

- Berkner, L. K. 1972. Rural family organization in Europe: a problem in comparative history. *Peasant Studies Newsletter* 1:145–156.
- Duben, A., and C. Behar. 1991. *Istanbul households: marriage, family, and fertility, 1880-1940*. Cambridge: Cambridge University Press.
- Ermisch, J. 1988. An economic perspective on household modelling. In *Modelling household formation and dissolution*, edited by N. Keilman, A. Kuijsten, and A. Vossen, 23-40. Oxford: Clarendon Press.
- Fauve-Chamoux, A. 1983. The importance of women in an urban environment: the example of the Rheims household at the beginning of the industrial Revolution. In *Family Forms in Historic Europe*, edited by R. Wall, J. Robin, and P. Laslett, 475-492. Cambridge: Cambridge University Press.
- Hammel, E. A. 1984. On the *** of studying household form and function. In *Households: comparative and historical studies of the domestic group*, edited by R. McC. Netting, R. R. Wilk, and E. J. Arnould, 29-43. Berkeley: University of California Press.
- Hammel, E. A., and K. W. Wachter. 1996a. Evaluating the Slavonian Census of 1698: Part I: Structure and Meaning. *European Journal of Population* 12:145-166.
- Hammel, E. A., and K. W. Wachter. 1996b. Evaluating the Slavonian Census of 1698: Part II: A Microsimulation Test and Extension of the Evidence. *European Journal of Population* 12:295-326.
- Laslett, P. 1972. Introduction: The history of the family. In *Household and family in past time: comparative studies in the size and structure of the domestic group over the last three centuries in England, France, Serbia, Japan and colonial North America, with further materials from Western Europe*, edited by P. Laslett and R. Wall, 1-89. Cambridge: Cambridge University Press.

- Litchfield, R. B. 1988. Single people in the nineteenth-century city: a comparative perspective on occupations and living situations. In *Continuity and Change* 3,1:83-100.
- Manke, M. 1997. Möglichkeiten und Grenzen der soziostrukturellen Auswertung der Volkszählungslisten 1819 für Rostock. In *Mecklenburg und seine ostelbischen Nachbarn: historisch-geographische und soziale Strukturen im regionalen Vergleich*, edited by I. Buchsteiner, 122-133. Schwerin: Stock & Stein.
- Manke, M. 1999. Die historische Bevölkerungsstatistik in Mecklenburg-Schwerin bis zur ersten Hälfte des 19. Jahrhunderts. In *Genealogie* 24,9-10:641-658.
- Manke, M. 2000. *Rostock zwischen Revolution und Biedermeier – Alltag und Sozialstruktur* (Rostocker Studien zur Regionalgeschichte 1). Rostock: Neuer Hochschulschriftenverlag.
- Manke, M., ed. 2005a. »... dass alle Welt geschätzt würde.« *Die Einwohner der Stadt Rostock nach der Volkszählung von 1819* (Kleine Schriftenreihe des Archivs der Hansestadt Rostock 15). Rostock: Ingo Koch.
- Manke, M. 2005b. Die Volkszählungen des Großherzogtums Mecklenburg-Schwerin in den Jahren 1819 und 1867 unter Mitberücksichtigung der Volkszählung des Jahres 1866. In *Genealogie* 27,3-4:449-468.
- Reher, D. S. 1987. Old issues and new perspectives: household and family within an urban context in nineteenth century Spain. In *Continuity and Change*, II,1:103-143.
- Schlumbohm, J. 1994. *Lebensläufe, Familien, Höfe. Die Bauern und Heuerleute des Osnabrückischen Kirchspiels Belm in proto-industrieller Zeit, 1650-1860* (Veröffentlichungen des Max-Planck-Instituts für Geschichte 110). Göttingen: Vandenhoeck & Ruprecht.
- Schmid, J. 1988. Principles emerging from sociology for definitions and typologies of household structures. In *Modelling household formation and dissolution*, edited by N. Keilman, A. Kuijsten, and A. Vossen, 13-22. Oxford: Clarendon Press.
- Sovic, S. 2008. Definitions and Documents in Family History: Towards an Agenda for Comparative Research, in *Social Behaviour and Family Strategies in the Balkans (16 th – 20 th Centuries)*. In: *Proceedings of a Regional Symposium, New Europe College Bucharest , June 9-10, 2006*, 137-158. Bucharest: New Europe College.
- Szolysek, M., et al. 2009. *Social change and family change in a Central European urban context: Rostock 1819-1867* (MPIDR Working Paper WP 2009-039). Rostock: Max Planck Institute for Demographic Research.
- Tscharnke, J. 1943 *Anlässe und Gebiete der Volkszählungen in beiden Mecklenburg*. LLD diss., Rostock University.
- Wall, R. 2001. The Household. In *Encyclopedia of European Social History* 4, edited by P. N. Stearns, 109-124. Detroit: Scribner.
- Wochenblatt, Großherzoglich Mecklenburg=Schwerinsches offizielles. 1819. Achtzehntes Stück.