WP. 24
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Madrid, Spain, 20-22 October 2003)

Topic (ii): Developments related to new methods and techniques

THE APPLICATION OF NIM FOR HOUSEHOLD AND BUILDING DATA IN THE SWISS CENSUS 2000

Supporting Paper

Submitted by SFSO, Switzerland[1]

I. INTRODUCTION

1. The Swiss Census 2000 consists of the population census and of the building census which includes the dwelling census. The population census and the building census are linked together by the definition of households and dwelling units. The NIM (New Imputation Methodology) methodology and programs developed by Statistics Canada have been applied along with other programs to edit and impute the data. We will lay emphasis on the characteristics of the input data for NIM and show the impact of the design of the questionnaires on the output of NIM by examples of the mentioned two parts of the Swiss census. NIM processed the demographic variables of the population census and the building characteristics of the building census. Further we will shortly show the limits of applying NIM to other fields than the population census by the example of the Swiss dwelling census 2000.

2. Section II describes the Swiss census 2000 in general. Section III shows the preliminary evaluations we executed on the household data of the Swiss census 1990. In Section IV follows the application of NIM in the Swiss population census 2000. Section V describes the application of NIM in the Swiss building census 2000. In Section VI we discuss NIM in the Swiss dwelling census. Finally, some concluding remarks are given in Section VII.

II. THE SWISS CENSUS 2000

A. General description of the Swiss Census 2000

3. For many decades the Swiss census has been consisting of a population and a building part. A person questionnaire is sent to each resident, each household gets a household questionnaire and each owner of a building with at least one dwelling gets a building questionnaire with questions about the building characteristics and questions about each dwelling unit of the building. In the former censuses the questionnaires were filled in with the help of a census agent. The main part of the edit and imputation process was executed by the municipalities (except the deterministic imputations in the population census) in the former censuses. The census 2000 played the role of an intermediate census in the sense that the census 2010 will be essentially based on administrative data and the whole census process will be

[1] Prepared by Daniel Kilchmann, Statistical Methods Unit. Daniel.Kilchmann@bfs.admin.ch.

executed at SFSO. The treatment of the questionnaires was executed by SFSO in a subcontracted service centre with the support of the municipalities in the census 2000. In addition, some of the data were based for the first time on administrative data. Hence, nearly the whole process of the Swiss census 2000 had to be invented and developed. Census day was 5 December 2000.

B. Special tasks of the Swiss census 2000

4. The population census has the main task to provide data on an individual level for each resident of Switzerland according to economic and civil residency. Further, the data were used to harmonise the administrative data of the municipalities in view of the census 2010. The main task of the second part of the Swiss census is to provide data for each building having at least one dwelling unit. The building and dwelling data were also used to build up the Swiss register of buildings and dwelling units. This register will be used essentially for the census 2010 and will be continuously updated by the municipalities. Generally we can say that the Swiss census 2000 had a macro and a micro task at the same time: produce high quality statistical information on different geographical levels and gather data for each record to be used in registers. It is not possible to get for all variables the values of the reality on record level which is asked for the registers. Succeeding in this second task would mean that the first task would also be succeeded in. This impossibility led us to the conclusion that it is necessary to split the processing according to the two different goals before using NIM.

C. Households and dwellings in the Swiss census 2000

5. The households were first defined on the base of administrative data. The definition of households was then adapted during the census by means of the answers of the person, household and building questionnaires. The households define the dwelling units. This requires the link between households and buildings for the use of the census. The link of households with the dwelling units was necessary for the use of the Swiss register of buildings and for the analyses of household data crossed with building and dwelling data.

6. In the process of defining households and their links to the buildings and dwellings we were confronted to the problems of erroneous administrative data, erroneous lecture of the scanner and erroneous answers on the questionnaires. Deadlines for the publication forced us to abandon this process when not all people were assigned to a household or the households formed by some of these people could not be assigned to an existing building. Hence, a part of the households which were linked to a building were not complete when NIM processed the person data. These households were completed and some households were created after NIM processed the person and household data. All the data of these households had to be processed manually after their definitive definition.

D. Item non-response and inconsistent data

7. In the population and in the building census item non-response and inconsistent data should be treated. Item non-response is a missing value where a valid answer is asked for. The inconsistent data in the population census were defined by within person edits and between person edits. An example for the first type is a 5 year old widower. An example for the second type of edit rule is a household with only one husband/wife. The inconsistent data in the building census were exclusively based on within building edits. For example, a building built after being renovated is not acceptable. Subject matter specialists developed the edit rules based on the edit rules of the Swiss census 1990. We changed or added some edit rules with the agreement of the subject matter specialists during the evaluation of NIM with the data of the Swiss census 2000.

III. PRELIMINARY EVALUATION OF NIM

8. SFSO received a prototype of the updated version of the NIM 97 software from STATCAN. In the following we will call this software just NIM. Detailed information about NIM can be found in the

references section. We adapted this version of NIM to the characteristics of the Swiss census. And we tested NIM on the person and household data of the 1990 Swiss census. In these tests we compared NIM to the deterministic imputations (CORA) developed for the 1990 census. We performed our tests only on two Swiss cantons (Fribourg and Valais) and the largest city of Switzerland (Zurich). Both cantons are bilingual but with different characteristics. The canton Valais has a more or less sharp division of the German and the French part. Whereas the language structure, German and French, in the canton Fribourg is mixed in some municipalities. Zurich on the other hand is a patchwork of different cultures and languages. We processed the data of about 850'000 people. The final data of the Swiss census 1990 was considered the "truth" for this evaluation. The item non-response and the inconsistent data were randomly generated with probability equal to the observed item non-response rate and rate of inconsistent data in the Swiss census 1990 for the processed variables. But the 'inconsistent data' we generated were not always really inconsistent because we just exchanged the existing value with another valid value. For the evaluation NIM and CORA processed the variables marital status, mother tongue and the 'situation in the household'[2]. In addition, NIM processed the demographic variables, nationality, education, activity and the mode of transport variables.

9. The edit rules of the 1990 census were used for both methods. We stratified the population data by canton and the number of household members for the tests with NIM. We used the following evaluation criteria: *number of changes, plausibility of the processed data, number of unresolved inconsistent data, preservation of the distribution, restitution of the true values*.

10. NIM did perform better for the first two criteria. The number of unresolved inconsistent data was slightly higher with NIM than with the deterministic imputation method. This is due to the fact that we limited the acceptable number of imputations because we did not want to copy the whole set of variables from the donor with NIM. In addition, in some strata made of households with more than 9 members all households failed at least one edit rule; hence NIM could not find a donor. Further, NIM processed more variables which results in more detected inconsistent households and hence in more unresolved inconsistent data.

11. We used the extension of the McNemar's statistic, cf. Chambers (2000), to test the preservation of the distribution. We had to reject the hypothesis of equal distribution for all variables and both methods. Nevertheless, we observed that the empirical value of the test statistic was much lower for the variables marital status and 'situation in the household' processed by NIM. On the other hand the deterministic imputation method did better for the variable mother tongue. Most of these corrections were based on the principal language per municipality file, which was not available for NIM. This result showed however that although the "truth" was obtained by applying the deterministic imputation method in the Swiss census 1990, it was not possible to reproduce these results with the same deterministic imputations applied to the same data with randomly generated item non-response and inconsistent data. This kind of imputation method depends on the processing order of the variables and the rules. Deterministic imputations on other variables not processed in this test, influence also the processed variables. Further, the randomly generated inconsistent data was not always really inconsistent as we mentioned in 8. And thus, it is not surprising that the corrected results do not absolutely correspond to the "truth". The above-mentioned advantages for NIM for the preservation of the distribution could also be observed by comparing the univariate distributions in the following way. We used a dummy variable for indicating for each category of the variables which one of the two methods is closer to the frequency of the "truth". We simply added up the dummy variables over all categories of a variable to get a very basic measure of the preservation of the distribution. A test of goodness of fit did unfortunately not make any sense because the randomly generated inconsistent data was not necessarily inconsistent.

12. The results for the restitution of the true values were similar to those observed for the preservation of the distribution. These results were obtained by comparing each individual imputed value with the "truth". Then we compared the rate of differences with the "truth" due to the imputations by

[2] The difference of the variable 'situation in the household' and the commonly used variable 'relationship to the household representative' is discussed in the paragraph 16.

NIM and those due to the deterministic imputations. The results for the variables marital status and mother tongue are shown in tables 1 and 2. The results for the 'situation in the household' were similar to those of the marital status.

Table 1: Comparison of NIM with CORA by means of the restitution of the true values of the marital status.

Marital status	Fribourg		Valais		Zurich (city)	
	NIM	CORA	NIM	CORA	NIM	CORA
Resituated	91%	57%	90%	54%	84%	47%
Not resituated	9%	43%	10%	46%	16%	53%

Table 2: Comparison of NIM with CORA by means of the restitution of the true values of the mother tongue.

Mother tongue	Fribourg		Valais		Zurich (city)	
	NIM	CORA	NIM	CORA	NIM	CORA
Resituated	84%	90%	90%	96%	87%	92%
Not resituated	16%	10%	10%	4%	13%	8%

13. Based on the above results the census managers decided that NIM would be used for the demographic variables in the population census. And if encouraging results were found through the tests on the building census data 2000 then NIM would also be used for this part of the census. The remaining population variables would be processed with deterministic imputations.

IV. THE APPLICATION OF NIM IN THE SWISS POPULATION CENSUS 2000

A. Data characteristics

14. The people who couldn't be linked to a dwelling unit were linked to a special household in their respective municipality called the 'remaining-household' (of the municipality). NIM processed private households and the 'remaining-households' and people in 'collective households' (hospitals, old people's homes, prisons, etc.). We limit the present discussion on the private households and the 'remaining-households'.

15. The private households were stratified using the canton variable and the number of household members when they had less than 7 household members. The households with 7 to 10 members were stratified on the NUTS2 level. This resulted in 7 regions, two of them being equal to the cantons of Zurich and Ticino. The households with 11 to 15 members were processed on the national level without any geographical stratification. This stratification procedure was necessary because of the little number of big households in most of the cantons and because of the need to attain for the whole population plausible values for the processed variables. About 30 households with more than 15 members were processed manually, because they all failed at least one edit rule. Hence, they could not be imputed with NIM.

B. Population variable characteristics

16. NIM applied to the population census 2000 (NIM-pop) processed the age by classes, the sex, the marital status, the questions about parenthood and the 'situation in the household'. The questions about parenthood consisted of three parts. In the first part the respondent was asked if he had ever a child. The number of children was asked in the second part. Birthdays of the first four children and from the youngest child were asked in the third part. The variable 'situation in the household' is not like the variable 'relationship to the household representative' used for example in the Canadian census. In fact, a clear definition of the household representative as the reference person for all the other household members is missing. This is due to the fact that in most of the households there is more than one member who defines himself as one of the heads of household. There are five categories designing the head of

household (person living alone, husband/wife, common law partner, single parent and other head of household). Thus, the household representatives were not necessarily unique. Some people answered this question according to their marital status, for example widows/widowers who say that they live alone but in fact their children live in the same household. These people were put alone in a household during the definition of the household process if no information indicating that this person does not live alone was available, for example from the household questionnaire or the person questionnaires of the other household members. Others answered the question about the 'situation in the household' depending on their "social status", for example people of the same sex living in a community and having a girlfriend/boyfriend and answering 'common law partner' instead of 'member of a community' because most of the spare time they spent with their girlfriend/boyfriend. Again, if no further information was available, such communities did not change and were then seen as a community of homosexuals. The edit rules can normally not handle such problems because it is impossible to find a hard rule which defines when such a household is what it looks like and when it is erroneous. It was inevitable to process such households manually, in some way with subjective rules. Unfortunately, this was only done after NIM-pop had processed the data and amplified some of the mentioned household types.

17. We introduced edit rules based on the age differences to prevent that parents and children of the head of household were changed by NIM-pop to head of household. This edit rules were not enough efficient, especially in incomplete households where one or more head of household were missing.

C. Adaptation of NIM

18. The version of the NIM prototype we received from STATCAN in 2000 used the nearest imputation action[3] as donor. We decided to select the imputation action randomly between the 10 nearest imputation actions. The probability of selection of the imputation action was proportional to its closeness given by the distance function defined in Bankier (2000). Further we had to adapt the definitions of the variables, edit rules and distance measures. In addition, we had to make some little changes to the source code of NIM because we processed the data on a UNIX server.

D. Technical aspects of the processing of NIM-pop

19. We processed the whole population on two UNIX servers with 12 processors and 4 processors respectively (900 MHz) in mid-August 2002. The processing of the private households, the 'remaining-households' and the collective households took 18 hours in total. NIM-pop used ASCII files for the data input, which were exported from the ORACLE database of the census. The imputed data sets were also written in ASCII files, which were then imported in intermediate tables of the database to be analysed before updating the database values. First analyses of the influence of NIM-pop on the data were carried out with SAS. Further analyses were executed on the database.

E. Evaluation of NIM-pop

20. We used the following evaluation criteria: plausibility of the person and household data, plausibility of the changes NIM-pop made and plausibility of the distributions compared to external data, like the Swiss census 1990 and the annual population statistic 2000.

21. The average rate of failed households was 31%, which is due to about 21% of the people which failed the edit rules. This high rate is mainly due to the fact that about 15% of the people in private households had an invalid or missing value in the 'situation in the household' variable. The 'situation in the household' variable was not edited for the 'remaining-households'. But these people were processed without considering any household structure.

22. The processed data did not fail any of the edit rules NIM-pop used. But the remarks in B of this section casted a little doubt on the results, although NIM-pop was "innocent": Evaluations on the type of

[3] Cf. Bankier (1999) for the definition.

household were done after the evaluations mentioned in this section. These evaluations revealed an important amount of homosexual couples. Further, we observed that children were changed to husband/wife in some households. The reason for this was that most of these households were not complete and mostly there was only one husband/wife in the household before NIM-pop processed the data.

23. The comparisons of the variable distributions after NIM-pop with the Swiss census 1990 and the annual Swiss population statistic 2000 (ESPOP) revealed very little unexpected values. For example, we had parents which were too young/old by the birth of their first/last child compared to the data of the other data sources if we had an older head of household in the household. This is due to the problem of defining parenthood in the household. These comparisons were made on the national level and on the canton level. In figure 1 we can see that the amount of singles and married people is slightly higher in the Swiss census compared to the ESPOP. This result could already be observed before NIM-pop processed the data and amplified this phenomenon. Further we crossed the variables age, marital status, 'ch-nationality' (Swiss citizen = yes/no), residential status permit and the 'situation in the household' on the national level and compared the results with the 1990 data. Again, these comparisons did not reveal important deviations from the expected values.

D. Processes after NIM-pop

24. Manual correction processes following NIM-pop were performed because of the reasons mentioned in B. An interactive tool including the edit rules was developed to process this data and the households described in 22.

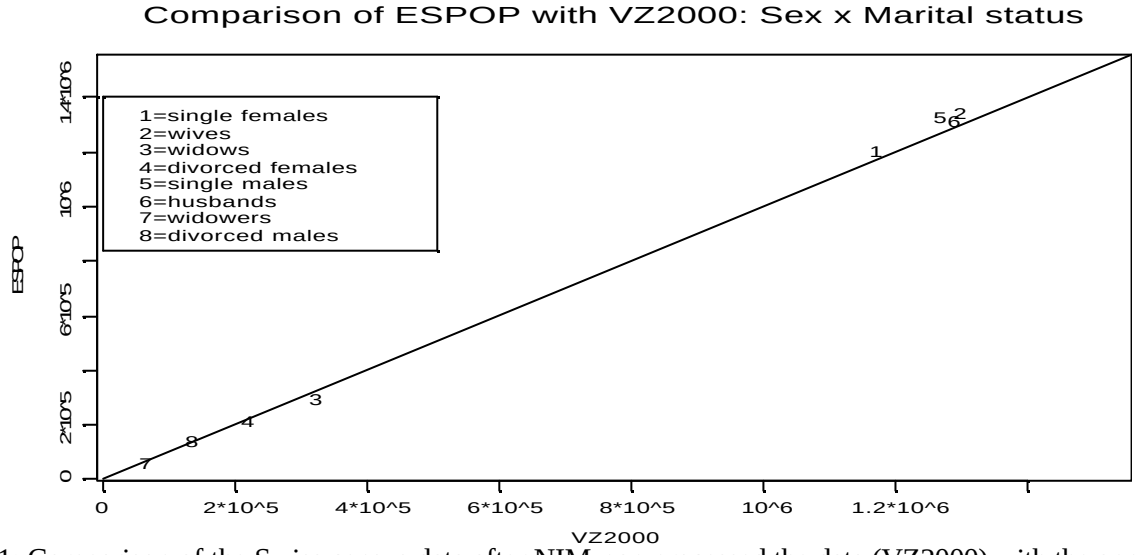

Fig. 1: Comparison of the Swiss census data after NIM-pop processed the data (VZ2000) with the annual Swiss population statistic (ESPOP), by sex and marital status on national level.

V. THE APPLICATION OF NIM IN THE SWISS BUILDING CENSUS

25. The Swiss building census is based on administrative data of the building addresses. The data of the Swiss building census was used to build up the Swiss register of buildings and dwellings. The imputed data were not included in this register as we explained above, cf. 4.

26. Every building owner had to fill in a questionnaire with questions about the building characteristics and characteristics of each dwelling in the building. SFSO developed a tool to fill in the questionnaires electronically for the administrators of numerous buildings.

27. The purpose of the edit and imputation process in the Swiss building census was to get a data set without missing items and inconsistent data by changing as few as possible of the existing values under the constraints which will be discussed in C of this section.

28. NIM can generally be seen as a multivariate nearest neighbour imputation tool. Actually, the algorithm for searching a donor depends only on the defined characteristics of the variables and is independent of whether the characteristics are person, household or building characteristics for example. We adapted NIM-pop for the building data by changing a few source code lines. We call this version NIM-b.

A. Data characteristics

29. The building characteristics are uncorrelated with their geographical location. For this reason we did not stratify by a geographical variable, but only by the number of floors of the buildings and the number of dwelling units in the buildings. We had first to impute the number of floors where it was missing or invalid before using this stratification. This imputation was carried out randomly in a stepwise backward process with existing building characteristics in the imputation model. The main variables in this model were the type of building, the number of dwelling units and the highest and the lowest floor of the building. All the strata with less than 200 buildings were put together in one stratum to ensure that a high rate of the failed buildings could be imputed. We will call this special stratum the 'remaining-stratum'. About 1.08% buildings of the 1.5 million buildings were in this 'remaining-stratum'.

B. Questionnaire and variable characteristics

30. NIM-b processed the following variables: type of building, construction period, variables about renovation, principal heating type, a variable concerning warm water and variables of the source of energy for both heating and warm water. We used the following variables as auxiliary variables: the municipality code, the variable saying if the building is a detached house or not, the number of floors and the number of dwellings in the building. Two further characteristics of the building whether the majority of the dwellings have an owner or not and the type of owner of the building. These two variables were not included in NIM-b because they interact with one of the variables of the dwelling part of the census. Hence, we preferred to process these two variables within the dwelling census part.

31. The question about renovation was split in two parts. In the first part the building owner was asked if an essential renovation took place in the last 30 years. In the second part he had to specify in which period the last renovation was finished. The questions about warm water and the source of energy for warm water in summer and winter were similar to this type of questions. The questions about principal heating type and the principal source of heating were also linked together but split in two parts. Such split variables deserve a deeper study when defining their distance functions to ensure that the probability of selection of the possible combinations is equal. For the questions about renovation that means, if an owner answers 'No' in the first part then the second part has to be blank. Hence, by treating the two parts independently, the imputation of 'No renovation took place' would be less expensive in NIM-b than imputing 'Yes' because this imputation needs also the imputation of a renovation period.

C. Deterministic error localisation aspects

32. The error localisation had a deterministic aspect under some particular conditions, as we will see in the next paragraphs. We observed that it is more efficient to include these deterministic error localisation rules in the distance definitions of the variables. We defined an upper limit of acceptable distance between the failed record and the potential donor. We assigned a higher distance than this upper limit to the imputation action based on a potential donor when at least one of the deterministic error localisation rules failed. In this way we eliminated most of the "not acceptable" donors within the search of the nearest neighbours of the failed record before trying to impute for any variable.

33. All building variables were included in the deterministic error localisation rules. In the following we will just give a few examples: We preferred to change the period of renovation rather than the construction period because the construction period was available from administrative data for most buildings. Equally, we preferred the answer to the question about the principal heating type to the answer concerning the principal source of heating. One exception was made for the last rule: if the answer in the principal heating type was 'no heating' and there was an answer for the principal source of heating then we preferred to change the variable principal heating type.

34. Preliminary comparisons of the census 2000 data with the census 1990 and the statistics of housing construction showed that the amount of buildings without warm water and without heating was too high in our data. In addition, we observed a high amount of buildings with heating but without warm water. Therefore we decided to prevent NIM-b from imputing these characteristics always if a building failed one of the rules concerning warm water or heating.

D. Restricted nearest neighbour search

35. We only accepted detached houses as donors for detached houses which failed the edit rules. This is due to the special character and the importance given to this type of buildings. We did not use the variable indicating a detached house as a stratification variable because detached houses were also allowed to be the donor for non-detached houses.

36. In the 'remaining-stratum' we only accepted buildings as donors which had at most 2 dwellings more or less than the failed building. This rule might have been omitted for buildings with a high amount of dwellings, because the building characteristics are only weakly correlated with the number of dwellings for big buildings. But because the 'remaining-stratum' contains also buildings with only a few dwellings and for simplification, we kept this rule for the nearest neighbour search for all buildings of this stratum.

E. 'Biased' selection of the imputation action

37. In general, the donors for any failed building could be from any part of Switzerland. This is admissible because the processed variables seem to be uncorrelated with the location of the building. Nevertheless, we decided to include the geographical characteristics by favouring buildings from the same canton as donors if they had the same distance as the randomly selected donor. The analyses of the imputation actions showed us that in general, the same variables were imputed for a given failed building by different imputation actions if their respective distances were equal. That means that the error localisation was not further biased by the bias in the selection of the imputation action. In addition, we sorted the buildings by a geographical code first. Thus about 50% of the donors were of the same municipality as the failed buildings and over 90% of the donors were from the same canton on average.

F. Characteristics of NIM-b

38. Especially the distance definitions we used in NIM-b were quite complex for some variables because we included the deterministic error localisation rules in the distance definitions.

39. Each potential donor could only generate one imputation action in NIM-b, because the building characteristics are in some way one-dimensional compared with the population census, where a household can have more than one member and hence the dimension of potential imputation actions increases by the number of household members. The imputation action was randomly selected among the 10 nearest imputation actions as we did in NIM-pop.

G. Preliminary tests

40. The first testing phase started in fall 2002. These tests lead to the decision that NIM would be used for the building census. We used the census 2000 data for the tests although these data were only

partially available at this time. The constraints, mentioned in C and D of this section, were only introduced in February and March 2003. The tests including these adaptations with all the census 2000 data took place in April 2003. We used the same evaluation criteria for these tests apart from some little changes.

H. Technical aspects of the processing of NIM-b

41. We processed the data of about 1.5 million buildings on a UNIX server with 4 processors (900 MHz) at the beginning of May 2003. The processing took 8 hours. The data flow of input data and output data was the same as for NIM-pop, cf. 19. The first analyses of the influence of NIM-b were performed with SAS.

I. Evaluation of NIM-b

42. The thematic specialists used the census 1990 and the statistics of housing construction to evaluate the distribution of the building variables after NIM-b. The evaluation of the behaviour of NIM-b included checking if no inconsistent data or missings persisted. Further we analysed the plausibility of the data changes made by NIM-b. A very simple univariate robust outlier detection method was used on each category on the municipality level to detect the values which have changed in a special way since the census 1990. All these analyses were very satisfying.

43. The rate of failing buildings was about 25%. The most important variables in terms of failing buildings were the variables of the source of energy for warm water in winter and in summer with 12% and 11% respectively. These high rates were also influenced by the constraint that NIM-b had to "over-impute" for these variables because of the reasons mentioned in 34.

44. NIM-b could not impute for 286 buildings. In fact, it was impossible for NIM-b to find a donor for these buildings under the geographical constraints we added only for the production version of NIM-b, cf. 37. We used the imputations of an older version of NIM-b for these buildings where the geographical constraints were not yet included.

VI. NIM IN THE SWISS DWELLING CENSUS 2000

45. The Swiss dwelling census is included in the Swiss building census. The questions about each dwelling of a building are asked on the same questionnaire. We first considered using NIM also for the dwelling census. But the reasons described later in this section and the deadlines forced us to abandon this idea.

A. Dwelling data characteristics

46. We had 6 dwelling variables and two building characteristics to process in the dwelling part of the Swiss census. The following variables were in the dwelling part of the census: the kitchen indicator variable, number of rooms, number of independent rooms, dwelling-surface, type of occupation of the dwelling, rent for the dwelling and the two building characteristic variables, floor property and type of owner of the building. These two building variables interact with the type of occupation of the dwelling, which is the reason for processing them in the dwelling part of the census. There were only two dwelling census variables for which missings were not allowed in the results of the census, although it is not desirable to have missing for the other variables. The two are the kitchen indicator variable and the type of occupation variable. In addition, we were asked to apply some deterministic imputations on these variables.

47. The building with the largest number of dwellings had more than 160 dwellings in April 2003. The dwellings of a building can be compared to the members of a household in the population census. But the NIM prototype cannot handle such huge amounts of "household" members. This number is

indeed limited to 19 members. We deemed the adaptation of the programming structures of NIM being too complicated. We decided to reprogram a considerable part of NIM with more dynamic structures. But we wanted to keep the same algorithms and its underlying methodology. We later abandoned this idea of reprogramming NIM because of our limited resources compared to the few imputations to do on the above-mentioned variables.

VII. CONCLUSIONS

48. We observed for the Swiss population census that preliminary tests on the real data with deep analyses are necessary for every E&I process. The time and resources used for these tests is a necessary investment for a satisfactory E&I production process.

49. NIM cannot compensate the problems with the design of the questionnaire and the production process; we mentioned the question about the 'situation in the household' of the Swiss census and the creation of the households. Actually, NIM can even increase such problems.

50. We observed in the preliminary test of NIM that it is not always possible to guarantee complex distributions on a geographically small region. However NIM is preferable to deterministic imputations whenever correction rules do not exist or if they are very complex.

51. The NIM software can be used for other problems than for imputing households in a population census. Actually, the application of NIM to the Swiss building census 2000 was successful and the adaptation did not cause excessive problems. We think that in other fields, like the Swiss dwelling census, NIM cannot be adapted without programming a part of the software in a different way. But we are convinced that the underlying methodology of NIM can be used for such kind of data.

References

Bankier M. (1999), Experience with the New Imputation Methodology used in the 1996 Canadian Census with Extensions for Future Censuses, Proceedings of the UN/ECE Work Session on Statistical Data Editing, Italy (Rome). (http://www.unece.org/stats/documents/1999.06.sde.htm)

Bankier M. (2000), Updated Version of the 1997 NIM Prototype, Social Survey Methods Division Report, Statistics Canada.

Bankier M., Lachance, M. and Poirier, P. (2000a), 2001 Canadian Census Minimum Change donor Imputation Methodology, Proceedings of the UN/ECE Work Session on Statistical Data Editing, United Kingdom (Cardiff). (http://www.unece.org/stats/documents/2000.10.sde.htm)

Bankier M., Lachance, M. and Poirier, P. (2000b), 2001 Canadian Census Minimum Change donor Imputation Methodology – Extended Version of Report, Social Survey Methods Division Report, Statistics Canada, Dated August, 2000.

Bankier M. (2002), Imputation of Demographic Variables from the 2001 Canadian Census of Population, Proceedings of the UN/ECE Work Session on Statistical Data Editing, Finland (Helsinki). (http://www.unece.org/stats/documents/2002.05.sde.htm)

Chambers R. (2000), Evaluation Criteria for Statistical Editing and Imputation, T0001.06, Euredit report. (http://www.cs.york.ac.uk/euredit/)