



Use of Administrative Records in GGS

Preliminary report of the Administrative Records Group (ARG) for the Third Informal Working Group Meeting of GGP, Spetses, 24-26 May 2004

By Helge Brunborg, Statistics Norway¹

1. Mandate and membership

The *Administrative Records Group (ARG)* was created at the Prague Informal Working Group meeting of GGP in February 2003. Its mandate is to investigate to what extent existing administrative records can be used to provide part of the information required by the Generations and Gender Survey questionnaire.

The members of the group represent the statistical institutes of the four Nordic countries Finland, Norway, Sweden and Denmark (as an observer). Belgium, Netherlands and Poland have also expressed an interest but have not yet attended any meeting or been active in any other way.

A first meeting of the group was held at Statistics Norway in Oslo in October 2003, with the following participants:

- Helge Brunborg and Svein Gåsemyr, Statistics Norway
- Anita Lange, Statistics Denmark
- Timo Nikander, Statistics Finland
- Gun Alm-Stenflo, Statistics Sweden
- Alphonse MacDonald, GGP manager at Population Activities Unit, ECE

The meeting explored the extent and modalities of using existing administrative records and surveys to provide part of the information required by the GGS. The group decided to conduct a detailed question-by-question analysis for each of the four countries in order to establish which GGS-type information may be drawn from administrative records and surveys. For each question, the following information would be provided: (1) degree of coverage (complete, partial, or none) in various administrative registers along with the degree of complexity of extracting the information (easy, difficult) and (2) degree of coverage (complete, partial, or none) in various surveys.

¹ E-mail address: helge.brunborg@ssb.no

2. The Nordic model of population registration

Since all active members of ARG represent Nordic countries with similar, although not identical, systems of population registration, it may be useful to present a brief summary of the main points of the characteristics of the Nordic model of population registration:

- a) Unique personal identification numbers (PINs) are allocated to all residents of the country, regardless of citizenship. Persons who have never been residents but who have economic interests in the country are allocated special numbers.
- b) There is a central population register (CPR), but with local civil registration offices.
- c) The CPR includes links which can be used to form couples, families and households.²
- d) The CPR is widely used for administrative purposes, such as passports, elections, taxes, education, health and social security.
- e) The PIN is also widely used by private institutions, such as banks and insurance companies. A bank account cannot be opened without a PIN.
- f) Government agencies and several private institutions have access to parts of the CPR, according to their needs.
- g) Vital statistics are closely integrated into the population registration system.
- h) All vital and migration statistics and most other demographic statistics are based on registers.
- i) Linking of individual records from different registers, and also with censuses, has become technically simple, and is widely done for administrative, statistical and research purposes.
- j) There is strong legal protection of administrative records with data on individuals. An independent data inspectorate oversees privacy protection and the use (and misuse) of such data. The legislation can sometimes cause problems for research.

Most of the CPRs in the Nordic countries were established in the 1960s based on the 1960 population census. The use of the PIN and the linking of different administrative registers have gradually been expanded and the quality is continuously improving.

One important lesson after four decades of experience is that the quality of the CPR improves with the use: The more frequent use and the more different users, the better quality. Linking of records from different administrative registers detects errors and improves the quality. For example, Statistics Norway has combined a large number of administrative registers with economic data to identify persons in Norway with no registered income, economic transfers or wealth in a given year, as this indicates that these persons do not live in Norway any longer. They have probably emigrated without registering the move.

The existence of all of the properties (a) - (i) above is not necessary for the use of administrative records. The possibility of linking sample surveys to registers may, for legal or technical reasons, be limited to the population census, for example. This is the case for Belgium and Poland.

² The household units are based on dwelling units. Denmark and Finland has been able to do this for several years, whereas Norway established this link in the 2001 Population and Housing Census.

The use of administrative registers to provide data for official statistics is routine in the Nordic countries. The creation of administrative registers may be long, complex and resource intensive, but once the system is in place and well maintained the production of statistics is relatively easy and cost-effective. However, the use of some information requires resource intensive procedures, while in other cases, access to the data is restricted and long procedures need to be followed to obtain permission.

It has become normal to supplement sample surveys with data from administrative registers in the Nordic countries. In Norway, for example, sample surveys almost never ask about age and marital status, and almost never about variables such as income and educational attainment. As the administrative registers are gradually expanded and improved more variables can be covered, at least to some extent, such as labour force participation and disease.

The Norwegian government policy on data collection is that the same item of information should only be transmitted to public authorities once, be it from private individuals or companies. Thus, supplementing sample surveys with administrative records is fully in line with this policy.

3. Supplementing interviews with administrative records

In all Nordic countries there exists a large set of administrative registers, and almost all official statistics are derived from these registers. Nevertheless, in each country a number of specific surveys are routinely carried out to supplement the administrative records. These surveys are usually drawn from the Central Population Register. It has become common practice to use administrative records to supplement the data collected in surveys, such as age, marital status, educational attainment and income.

There are both advantages and drawbacks of using administrative records to supplement data collected in interviews. We will here discuss some of the issues at a general level. ??

Advantages of using administrative records

a. Lower costs: The most immediate reason for wanting to supplement interview data with administrative records is to save interviewing time. An initial and tentative assessment indicates that between ? and ½ of the variables in the core questionnaire can be obtained from administrative records. Administrative records are almost free.

b. Better data quality: Some variables are more accurately recorded in administrative registers than in a sample survey, for example date of birth and income.

c. Wider coverage: Administrative registers include data for the total population, both on events (births, moves, etc) and on status (marital status, educational attainment, etc) at certain time points, e.g. the end of the year. They can, in fact, be considered as censuses of the population, and may in principle be taken every day.

d. Time series: Administrative registers include data for many years. Thus, administrative registers can be used to establish longitudinal data sets.

Drawbacks of using administrative records:

e. No administrative records exist: Some data cannot be obtained at all from registers, in particular qualitative or subjective variables such as attitudes, preferences, intentions, plans, and wishes. But in some cases it may be possible to get indirect data on the intentions by linking data on actual behaviour *ex post*, e.g. whether a couple had another child or not in a given period, and use this as a proxy for intentions.

f. Different definitions: Some variables, which may be drawn from registers, may appear to be the same as those in the core questionnaire, but the use and definitions may be somewhat different. One well-known example of this is *marriage*, which is a registered legal status but whose practice and social connotation have been "polluted" by the increasing number of consensual unions. Another example is income, where registers and interviews give somewhat different values.³ It is not always clear which source is the most reliable, however. A final example is date of union formation and dissolution: The dates that can be obtained from registers are reported dates of moving to or from a partner, which may differ from the actual dates – which is what we are interested in.

g. Limited coverage: Some administrative records do not cover the total population. In Norway, for example, in most cases there are no registered ID numbers of parents of people born before 1953. Residents born abroad are a special challenge, as educational attainment and other background information of immigrants are often missing.

h. Complicated and costly extraction: Some variables may be drawn from registers but at large costs, requiring complex programming and the use of many registers. This is the case, for example, with housework in Norway. In such cases, it may be faster and less costly to collect the data in the survey.

i. Time lag: Some administrative records only become available some time after the event or period they cover, such as income in the previous year (almost a year later in Norway). This should not be a serious problem for GGS, however.

j. Restrictions on linking: There may be legal restrictions on the linking of interview data with administrative records. In Norway the letter of introduction to the respondents need to mention that the collected data may be used for such purposes, otherwise this cannot be done.

k. Restrictions on use: There may also be restrictions on the type of use of linked data. In Norway, for example, linked data may only be used for statistical and research projects, implying that anonymised linked data cannot be made available to everybody (such as the European Social Survey data).

³ Epland and Kirkeberg

4. Availability of administrative records to supplement the GGS

Andres Vikat of MPIDR, the coordinator of the Questionnaire Development Group, kindly provided ARG with a spreadsheet version of the GSS questionnaire, with one row for each question. This spreadsheet was sent to all members of ARG, with a request to indicate the following information, for each question:

1. Degree of coverage (completely, partially, or not)
2. Source of information (name of administrative register, such as CPR - Central population Register)
3. Degree of complexity to create the information (easy, difficult)
4. Which survey could provide the information
5. Degree of coverage (Complete, partially, or not)

Unfortunately, two different versions of the spreadsheet based on the questionnaire were used, with Sweden using a more recent version, but this does not matter much for the main findings. The final spreadsheet will be used later.

In this preliminary round of reporting the four Nordic countries Denmark, Finland, Norway and Sweden followed somewhat different practices, although all of them answered the first two questions, which are the most important. Only Finland reported the ease of using administrative records and only Sweden mentioned other existing sample surveys that could provide the data for each GGS question.

Table 1 shows a summary of the responses from the four countries. It shows that of the approximately 600 questions in the GGS, somewhat more than half cannot be covered at all by registers (between 51 and 66 per cent). On the other hand, this implies that administrative records *might* cover somewhat less than half of the questions. However, some of the data are only partially available, i.e. for subgroups of the population or with only a few of the specified response categories. Moreover, some of the responses were marked with question marks, indicating that more time and probably also that some testing will be required to determine if the GSS questions can be replaced by administrative records.

Table 1. Availability of administrative records for GGS-1 questionnaire variables

	Denmark	Finland	Norway	Sweden
Number of questions				
Not available	353	371	306	394
Perhaps partially available	16	0	31	7
Partially available	24	92	39	45
Available	171	120	149	117
Perhaps available	16	0	56	3
Total	564	583	550	576
Approximate number of questions	596	596	596	601
Per cent				
Partially or perhaps partially available	7 %	15 %	12 %	9 %
Available or perhaps available	33 %	36 %	32 %	27 %
Total (some data probably available)	39 %	51 %	43 %	36 %

Based on two different versions of the questionnaire.

The main reason for the differences between the Nordic countries is that the administrative registers differ with regard to their contents. For example, Sweden does not yet have a dwelling register that is required for linking persons living in the same household, whereas Norway does not have registers with data on abortions and hospital admissions.

In the appendix we give an overview of the availability of administrative records for each block of questions in the GGS core questionnaire. A worksheet with the availability of administrative records for each core question, for each country, is available on request.

With regard to the *household roster* for the established households all information, except on temporary absence, can be obtained from existing administrative registers. There is, however, a need to establish international definitions for the couple, family and household, possibly based on different typologies. What is missing in particular is a definition of when a family nucleus unit begins and ends.

With regard to the questions on *disability* the use of a general health indicator, based on a life path approach, may be preferable. However, there is no international recommendation for such an indicator, but that there are ongoing discussions on the theoretical aspects of such an indicator. The proposed health indicator has not been introduced in Norway.

The administrative registers may provide all information required on *education* in the household roster, except for some categories of immigrants and perhaps also parts of the elderly population.

The section on *childcare* does not follow the standard practice of time use surveys. This information is not available from administrative registers, but may be obtained from time use surveys. On the basis of the available information a time-use model may be developed and be used to impute the necessary values for specific questions. There is also some limited information about this in administrative registers, which may indicate whether a person is active in household work full time.

Administrative registers may provide the objective factual information required from the *partnership* section of the questionnaire, but no subjective issues can be covered. Surveys are needed for this.

There are a large number of administrative registers in the Nordic countries but some of these are not controlled by the statistical offices. Complex administrative procedures are required to be given access to some of the data. These refer mostly to health and health related issues.

5. Data from other sample surveys

There is ongoing work in the Nordic countries to create synthetic samples by combining the results from different samples to create linkages between variables covered in different surveys. This would require some kind of imputation or synthetic linking. The analysis would

either have to be done at group level or values would have to be imputed for the GGS sample, based on data from other sample surveys.

The GGP Consortium Board has expressed resentment against this. The problem is, however, that it is virtually impossible for statistical offices in some countries to secure funding for collecting interview data that are already collected regularly in other surveys. An example of this is the living conditions survey (ULF) in Sweden.

6. Conclusions

In the Nordic countries it seems possible to cover between ? and ½ of the questions in the GGS core questionnaire with administrative records. This should imply a similar reduction of the interview time.

The GGS questionnaire needs to be redesigned for countries using administrative records. The skip patterns, for example, probably need to be changed. Since most of these countries have similar administrative registers, development of a common “administrative records questionnaire” could be considered.

Further investigations are necessary to find out more about the quality, validity, reliability and coverage etc of using administrative records, and also the costs of using such records to supplement the interviews in a GGS.

With regard to the use of data from other sample surveys, more methodological work needs to be done before a final conclusion can be drawn on the feasibility and desirability of this.

Appendix: Coverage of the core questionnaire by administrative records in the Nordic countries

D = Denmark, F = Finland, N = Norway, S = Sweden

Generally, the following variables cannot be covered by administrative records:

- *Satisfaction with relationship, etc., as e.g. in Q 202.*
- *Intentions* (but data on actual events may be obtained ex post from administrative registers, for example, as proxies for intentions to have another child (Q 624) or intentions to resume education (Q 124).
- *Time distance*, e.g. to father's residence (Q 529). As proxy for this we may use the *geographical* distance, drawn from GIS information in the dwelling register. This variable can be used to estimate the time distance, based on average travel times inside and outside cities, if the time factor itself is important.
- *Number of meetings*, e.g. with stepchildren (Q 234)

1. Household

Household Roster: D & F fully, F partly, S not. (Relationship of R to hh members only partly.)

Dwelling Unit: D, F & N fully except for satisfaction and intention to move

Education: Fully, except for resumption of study (but may be obtained ex post)

2. Children

Childcare: Nothing, except for parental leave

Non-Resident Children: Fully, except for meetings with child and satisfaction

Stepchildren: Fully, except for meetings with child and satisfaction

Grandchildren: Fully, except for meetings with child and satisfaction.

3. Partner

Current Co-Resident Partner or Spouse: All questions, except that cohabitation is not fully covered, esp. for older people. Sweden less well covered than the other countries.

Current Non-Resident Partner or Spouse: Q306-309 only partly covered or not at all (start of relationship, reasons for living apart). Particulars about the non-resident partner are fully known, such as date of birth, according to the country reports. But how do the registers know who the non-resident partner is, unless they are or have been married or living together or have a child together?

Intentions of Union Formation: No administrative records exist, except perhaps ex post.

Partnership History: Most of the questions can be covered by registers.

Child Alimony/ Maintenance: Poor coverage but some of the questions may be partly covered by tax records and data from the Social Security Administration in Denmark and Norway.

Partner Alimony: The same very limited coverage as for child alimony.

4. Household Organisation and Partnership Quality

Household Organisation: No administrative records exist.

Decision-Making: No administrative records exist.

Partnership Quality: No administrative records exist.

5. Parents and Parental Home

Co-Residence with Biological Parents: Fully covered by registers.

Questions to Those Who Live with Biological Father, Not Mother: Fully covered by registers, except for living arrangement, illness and time distance.

Questions to Those Who Live with Biological Mother, Not Father: Fully covered by registers, except for living arrangement, illness and time distance.

Questions to Those Who Do Not Live with Either Biological Parent: Fully covered by registers.

Basic data on both parents: Fully covered by registers, but limited data for older cohorts.

Co-residence of parents: Fully covered by registers.

Father, if parents do not live together or if mother not alive: Living arrangement and illness partly covered, the other questions not at all.

Mother, if parents do not live together or if father not alive: Living arrangement and illness partly covered, the other questions not at all.

Both parents, if live together: Living arrangement and illness partly covered, the other questions not at all.

Questions to Those Who Live with Both Biological Parents: Fully covered by registers.

Brothers, Sisters, Grandparents: Fully covered by registers.

Parental Home During Childhood: Most questions fully covered by registers, except for older cohorts (varies from country to country).

Intentions to Start Living Separately from Parents: No administrative records exist.

6. Fertility

Current Pregnancy: Only partly, i.e. if the respondent is giving birth later.

Fecundity: No administrative records exist, except that Finland has administrative data on sterilization.

Intentions to Have Children: No administrative records exist.

7. Health and Well-Being

Health in General: No administrative records exist, except for disability in D, F and N.

Personal Care: No administrative records exist.

Emotional Support: No administrative records exist.

Locus of Control: No administrative records exist.

Well-Being: No administrative records exist.

8. Respondent's Activity and Income

Questions to Those on Maternity, Parental or Childcare Leave: Partly or fully covered by registers.

Questions to the Unemployed: Partly or fully covered by registers.

Questions to Students: Partly or fully covered by registers.

Questions to the Retired: Partly or fully covered by registers.

Questions to the Ill or Permanently Disabled: No administrative records exist.

Questions to Homemakers: No administrative records exist.

Questions to Those Who Are in Military Service or Social Service: No administrative records exist.

Previous Employment: Partly covered by registers.

Questions to Those Who Did Not Declare Working: Almost nothing in registers.

Questions to Those Who Are Working: More or less fully covered in D and N, almost nothing in F and S.

Questions to employees: More or less fully covered in D and N, almost nothing in F and S.

Questions to the self-employed: Only limited data in registers.

Questions to all those who are working: No administrative records exist.

Additional Job or Business: More or less fully covered by registers in D and N, partly in F, nothing in S.

Income: Mostly fully covered by registers.

9. Partner's Activity and Income

Questions to Those Whose Partner Is on Maternity, Parental or Childcare Leave: More or less fully covered in D and N, almost nothing in F and S.

Questions to Those Whose Partner Is Unemployed: More or less fully covered by registers in D and N, partly in F, nothing in S.

Questions to Those Whose Partner Is a Student: More or less fully covered.

Questions to Those Whose Partner Is Retired: More or less fully covered.

Questions to Those Whose Partner Is Ill or Permanently Disabled: More or less fully covered.

Questions to Those Whose Partner Is a Homemaker: More or less fully covered by registers in D and N, partly in F, nothing in S.

Questions to Those Whose Partner Is in Military Service or Social Service: More or less fully covered in D and N, almost nothing in F and S.

Partner's Previous Employment: More or less fully covered in D and N, almost nothing in F and S.

Questions to Those Who Did Not Declare Partner Working: Fully covered in D and N, almost nothing in F and S.

Questions to Those Whose Partner Is Working: More or less fully covered in D and N, partly in F and nothing in S.

Questions to those whose partner is an employee: Almost nothing in registers.

Question to those whose partner is self-employed: Covered by registers in D, F and N.

Partner's Additional Job or Business: More or less fully covered in D and N, partly in F and nothing in S.

Partner's Income: Mostly fully covered by registers, except for "number of received payments", which is not an essential question.

10. Household Possessions, Income and Transfers

Household Possessions and Economic Deprivation: Nothing in registers, except for ownership of telephone, cars and second home.

Income from Other Sources than Employment: Fully covered by income/tax registers.

Total Household Income: Fully covered by income/tax registers.

Monetary Transfers and Inheritance: Fully covered by income/tax registers.

11. Value Orientations and Attitudes

Nothing in registers except perhaps for data on religious denomination.