Hen/6th May 2014

**Improved input data quality from administrative sources through the use of quality indicators**

**Mr. Coen Hendriks,** Senior Adviser, Division for Statistical Populations, Statistics Norway

***Acknowledgement****: Many thanks to Anders Haglund, Johan Åmberg, Grete Smerud and Jan Furseth from the Division for Statistical Populations at Statistics Norway for valuable collaboration while developing the quality indicators and reports, and for useful comments on this paper. They are not responsible for the content of the paper.*

***Abstract***

*Statistics Norway (SN) has processed the quality indicators for administrative sources from Blue-Ets Work Package 4. This resulted in lists with quality checks for three administrative base registers, the Central Population Register, the Cadaster and the Register for Legal Entities. The quality checks are run on a regular basis and summarized in quality reports. The quality reports give an overall view of the quality of each separate register and how it develops over time. The quality reports are in demand by the register owners and used as a tool for quality improvement by the owners of the administrative base registers. Recently SN has started to discuss quality across the administrative base registers, involving several administrative base registers and their owners. Again this initiative was well received by the register owners and will inevitably result in improved input data quality. The paper describes how SN cooperates with owners of administrative base registers on quality improvement.*

**1 The use of administrative sources in Statistics Norway**

Statistics Norway (SN) is using three administrative registers as base registers for the production of statistics. The base registers are:

* The Central Coordinating Register for Legal Entities (CCRLE, owned by the Brønnøysund Register Centre)
* The Cadastre (ground properties, addresses, buildings and dwellings, owned by the Mapping Authority)
* The Central Population Register (CPR, owned by the Tax Directorate)

To facilitate the statistical use of these registers, SN has established an integrated database solution for the statistical versions of the base registers. The statistical database solution includes:

* The Business Register (BR, businesses/Local Kind of Activity Units and enterprises, based on the CCRLE)
* The Statistical Cadastre (SC, based on the Cadastre)
* The Statistical Population Register (SPR, based on the CPR)

The databases are updated daily from the administrative sources. They are extended with information from other sources. Examples of extensions are aggregated information on employment which is added to businesses in the BR and information on jobs which is added to persons in the SPR. The solution allows for data inspection at micro level, browsing from one database to the other and extraction of combined data from different sources. The personal identification number (PIN), the business identification number (BIN) and the numerical address are used as keys for linkage. Managing the databases for statistical purposes is called statistical population management in SN [2].

In addition to the base registers, SN uses a variety of other administrative sources [6]. Some of them are well established in the statistical system. An example is employment statistics which is based on The Register on Employers and Employees, the Register of Personal Tax Payers and several other sources. Another example is income statistics which is based on data from tax returns, the Tax Register, the End of the Year Certificate Register and several other sources. It goes without saying that these and many other statistics are produced in combination with data from the base registers.

SN is constantly adding new administrative sources to the statistical system. The sources keep track of units (persons, multinational enterprise group, enterprises, businesses, dwellings etc.) by means of an identification number (personal or business identification number) and/or by address. They are being used to fill out missing values, analyse and correct over and under coverage, reduce inconsistency or simple to add new data to the base registers. Some examples of new sources are:

* Euro Group Register (EGR)
* The Taxation Property Register (SERG) from the Tax Authorities
* Homes for sale in Finn.no (a commercial advertising service on the internet)
* Service Based Personal Data System form the Norwegian Labor and Welfare Administration (NAV)
* The Register on Licensees from the Norwegian National Broadcasting (NRK)
* The Common Register on Mail Recipients from Norway Post (Posten Norge AS)
* The Register on Customers from the State Educational Loan Fund (Lånekassen)
* The National Digital Contact Register from the Agency for Public Management and eGovernment (Difi)

EGR is used to improve the external dimension in the Business Register. The quality of the Cadaster is improved by using information from SERG and Finn.no. Within social statistics, some of these sources are being used to improve contact information, in order to reduce non-contact in interview surveys. In future the new sources might also be used to estimate the number of unregistered emigrations in the CPR [4] or as sources of information for the actual distribution of the population in Norway.

**2 Measuring and documenting input quality**

*2.1 The quality indicators*

Based on the quality indicators for administrative sources from Blue-Ets WP4 [1], SN has developed quality checks for the base registers. This was done for the benefit of the users of the registers in SN and for the owners of the administrative registers. The quality checks are run on a regular basis. For each check, an indicator is added to the dataset (table 1). A positive value indicates that there might, but not necessarily, be something wrong with a specific unit in the register. The indicators are summarized in quality reports. The quality reports give an overall view of the quality of each separate register and how quality develops over time. The quality checks indicate quality issues at micro level.

*Table 1. Dataset with quality indicators.*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Ind1** | **Ind2** | **Ind3** | **Ind4** | **Ind5** | **Ind6** | **Ind7** | **Ind..** | **IndN** | **Sum** |
| **Unit1** | 1 | 1 | 0 | 1 | 0 | 0 | 1 | .. | .. | 4 |
| **Unit2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .. | .. | 0 |
| **Unit3** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | .. | .. | 1 |
| **Unit4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .. | .. | 0 |
| **Unit5** | 0 | 1 | 0 | 1 | 0 | 0 | 1 | .. | .. | 3 |
| **Unit6** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | .. | .. | 1 |
| **Unit..** | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| **UnitM** | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| **Sum** | 1 | 2 | 0 | 2 | 0 | 0 | 4 | .. | .. | P |

The number of positive indicators P can be calculated from the dataset. Let N be the number of checks and M the number of units. The number of positive indicators P in the dataset can be calculated as follows:

P =

A general quality indicator Q can be calculated by dividing the total number of positive checks by the number of cells in the dataset. To improve readability Q is calculated per thousand:

Q = (P/(N\*M))\*1000

Q can be calculated for the dataset as a whole and for subpopulations. Q from two different registers is not suited for comparing because the denominator varies between registers. However, the general quality indicator can be used to monitor quality in one register and its subpopulations over time and to compare quality between subpopulations. Units/rows with many positive indicators can be of interest to the data owners. Checks/columns with many positive indicators can also be of interest.

*2.2 Defining and operationalizing the checks*

A working group was established at the Division for Statistical Populations. The coordinators/group leaders for the statistical versions of the base registers participated. The coordinators have a good overview of the subject matter, long experience from working on registers and have excellent programming skills in SAS. The work started by translating the list of indicators from Blue-ets WP 4. Each coordinator proposed quality checks for the administrative register they are experts on. The proposed checks where discussed with the project leader and grouped according to the dimensions from Blue-ets WP 4 in technical checks, accuracy, completeness, integrability and time-related dimensions. The checks where compared across the other base registers to ensure some degree of comparability.

Then the programming started. This resulted in three sets of indicator files, one for each base register. A typical set of indicator files for the CPR is made up from an indicator file for registered persons[[1]](#footnote-1), one for families and one for dwelling addresses which are being used in the CPR. This means that in the CPR each registered person, each formal family and each residential address is checked and indicators are produced. The units which are being checked in the Cadastre are addresses, buildings, ground properties and functional units in buildings (e.g. dwellings). The units which are being checked in the CCRLE are legal units and local kind of activity units.

*2.3 The quality reports*

The positive indicators are counted for each register and grouped in the five quality dimensions from Blue‑ets WP 4. Finally they are entered into a quality report (refer to Appendix for an example). Thus SN came closer to the intended quantitative approach from Blue-ets WP4. The first versions of the quality reports in SN where based on a descriptive approach [3].

The quality reports are made available for users of the registers in SN on the intranet.

The quality reports are in demand by the register owners and used as a tool for quality improvement. SN has quarterly meetings with the owners of the CPR and the Cadaster, and the specific statistics divisions. E.g. the division in charge of Construction Statistics might join the quarterly meeting on quality in the Cadaster. Depending on the agenda, other public agencies which might benefit from the meeting are invited. The Co-operation Forum for the administrative Central Coordinating Register for Legal Entities (CCRLE) showed great interest in the quality report for the CCRLE.

*2.4 Cooperation with the data owners*

Co-operation on quality with register owners has been SN-policy for many years, but the approach was recently renewed. The management of SN initiated the process, starting in November 2011. It was decided to establish general agreements on co-operation with external register owners to improve input quality in the registers which SN uses. The approach has become a routine during the annual updating of next year’s operating plan. So far SN has received many positive reactions from the major register owners.

The purpose is to improve and standardise the contact with register owners, in order to improve the quality of administrative data sources before they are transferred to and checked by SN. SN has 25 data providers of administrative information who provide information for 89 sources. Currently SN has signed agreements with 20 data providers and has produced quality reports for 82 sources. The quality reports are attachments to the agreements. However only the quality reports for the three base registers are quantified as described in this paper. The quality reports for the other sources are more of a descriptive nature. This approach has shown to be beneficiary for SN, for the register owners and for other users of the registers. In all, SN expects this will contribute to a reduction in public spending [3].

SN has regular meetings with the data providers to follow up the quality reports. The meetings with the owners of the base registers are held on a quarterly or semi-annual basis and are organised by The Division for Statistical Populations. Other meetings with register owners are organised by the Divisions which are in charge of the specific statistics using the source.

SN can inform the data owners about quality aspects, either by means of a quality report or by transferring the units with positive indicators. Table 2 shows 12 municipalities with the highest value of the general quality indicator Q for registered persons. The municipalities with the worst scores concentrate mainly in the county of Troms (municipality code starts with 18). This indicates that population registration in Troms might need some extra attention. A further inspection of the quality report for registered persons (Appendix, table 1) shows that indicators for inconsistent values have the highest score. Invalid links between family members (indicators C25A-C) is a major problem. The quality report also indicates measurement errors, more specifically quality issues with the registered addresses in the CPR (indicators C24F-H). However this issue is improving (table 3).

***Table 2****. General quality indicators Q for registered persons in the CPR. Twelve municipalities with the highest general quality indicator Q, four large municipalities and Norway, 1st January 2014.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Municipality** | | **Records checked** | **Positive indicators** | **Records without positive indicators** | **General quality indicator Q** |
| 1849 | Hamarøy | 1 819 | 1 244 | 1 154 | 24 ‰ |
| 0817 | Drangedal | 4 132 | 1 561 | 3 302 | 13 ‰ |
| 1854 | Ballangen | 2 587 | 976 | 1 880 | 13 ‰ |
| 1857 | Værøy | 777 | 288 | 613 | 13 ‰ |
| 1874 | Moskenes | 1 108 | 394 | 882 | 12 ‰ |
| 2018 | Måsøy | 1 244 | 450 | 1 014 | 12 ‰ |
| 1514 | Sande | 2 632 | 872 | 2 258 | 11 ‰ |
| 1835 | Træna | 489 | 163 | 388 | 11 ‰ |
| 1840 | Saltdal | 4 691 | 1 458 | 1 940 | 11 ‰ |
| 1850 | Tysfjord | 2 004 | 623 | 1 617 | 11 ‰ |
| 1851 | Lødingen | 2 246 | 735 | 1 855 | 11 ‰ |
| 2014 | Loppa | 1 027 | 318 | 843 | 11 ‰ |
| 0301 | Oslo | 634 249 | 135 547 | 556 138 | 7 ‰ |
| 1201 | Bergen | 271 854 | 46 889 | 245 024 | 6 ‰ |
| 1103 | Stavanger | 130 755 | 17 357 | 121 071 | 5 ‰ |
| 1601 | Trondheim | 182 122 | 22 166 | 169 173 | 4 ‰ |
|  | Norway | 5 107 477 | 777 584 | 4 638 325 | 5 ‰ |

Sharing information from a dataset which combines information from different sources is in conflict with the Statistics Act and is not practised in general. In a few cases SN has received requests from register owners to assess data quality in one source by linking it to another source. Provided the register owner is authorised to use the other source and has a copy of the other source available, an agreement on data processing can be drawn up between SN and the register owner. An agreement on data processing regulates the linking of the sources for quality improvement purposes and the reporting of errors at micro level (the multiple source approach, [3]).

**3 Quality across registers**

Once the indicators where established for the separate registers, SN started measuring quality across registers. By matching information from different sources, inconsistency can be measured. Table 3 shows a preliminary result at an aggregated level. The table shows that some municipalities have “fewer mistakes” in the CPR, while addresses in the Cadaster show “better quality”[[2]](#footnote-2). SN checks the quality of the numerical addresses in the CPR by matching to the address register form the Cadaster. This is an example of the way information from two sources can be used for quality assessment and improvement.

***Table 3****. Municipalities showing improved quality in the CPR 1st January 2013- 1st January 2014, and in the Cadaster 1st January 2012 - 1st January 2014.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Municipality** | **Percentage of registered persons improved quality (a reduction in the number of positive indicators) from 1.1.2013 to 1.1.2014 (source: CPR)** | **Percentage of addresses with improved quality (source: CPR and Cadaster)** | | |
| **1.1.2012** | **1.1.2013** | **1.1.2014** |
| 0511 Dovre | 12 | 36 | 65 | 83 |
| 0520 Ringebu | 45 | 68 | 85 | 94 |
| 0941 Bykle | 12 | 4,0 | 8,8 | 71 |
| 1750 Vikna | 28 | 51 | 52 | 99 |
| 1903 Harstad | 93 | 75 | 81 | 81 |
| 1913 Skånland | 72 | 7,6 | 85 | 86 |
| 1919 Gratangen | 33 | 8,4 | 78 | 79 |
| Norway | 2 | 71 | 74 | 77 |

SN is currently developing a check for under coverage in the Business Register (BR) which combines information from different register based sources. The Register on Employers and Employees is matched to the Business Register (BR) and the CPR. The BR provides addresses of the employers. The CPR provides addresses of the employees. A geographical cluster of employees without a local kind of activity unit of the employer in reasonable distance might indicate under coverage in the BR (a missing local kind of activity unit). The check is yet to be developed, and should of course take into account specific properties of the NACE-code of the enterprise (e.g. long distance commuting in building activity). But the first tests look promising.

Register owners are interested in the results of quality measurements across registers. SN is at the very beginning of exploring quality issues across registers. Work ahead will be defining quality checks across sources, programming the checks, measuring and reporting quality, and presenting the result to the users and the owners of the registers. Finally SN will make recommendations on how register quality can be improved across the sources. This must be done within the provision on secrecy from the Norwegian Statistics Act and preferable be supported by an agreement on data processing.

**4 Final remarks**

There is a difference between good quality data from registers and good quality register based statistics. Zhang and Hendriks (2012) have outlined problems with input data from registers to the 2011 Norwegian Census. Statistical methods of micro integration were needed to create a complete (“one-number”) census data file of linked dwellings and households. They developed an approach that meets the need for detailed tabulation required by the census, while maintaining the existing ‘marginal’ statistics on the dwelling stock and the households from register based sources.

One of the problems with input data from registers is relevance or definitional error. The problem of definition errors was also addressed by Sagelvmo [4]. Administrative registers are developed for a specific purpose which not necessarily is complying with a statisticians need for data. The purpose of a statistician is to reflect the real world as close as possible. Therefor a statistician must be able to handle changes in the register which are the result of political decisions. One has to know the purpose of the register, a sound knowledge of the subject matter and not at least, have sufficient resources to make the transition from good quality register based data to good quality statistics from registers.

**Litterature**

[1] Daas, P. and S. Ossen 2011: Report on methods preferred for the quality indicators of administrative data sources. *Deliverable 4.2 BLUE-Enterprise and Trade Statistics. European Commission European Research Area, Seventh framework programme*

[2] Hendriks, C. and J. Åmberg 2011: Building and maintaining quality in register populations. *Paper to the 58th ISI Congress (Dublin, Ireland, 21-26 August 2011)*

[3] Hendriks, C. 2012: Input data quality in register based statistics – The Norwegian experience. *Paper to the Joint Statistical Meetings (San Diego, USA, 28 July – 2 August 2012)*

[4] Sagelvmo, I. 2013: Bruk av registerdata til statistikkformål – når er kvaliteten bra nok? *Paper to the Nordic Statistician Meeting (Bergen, 14-16 August 2013), Norwegian only*

[5] Tønnesen, M., S. V. Pedersen, K. Vassenden and J. Åmberg 2014: Uregistrert flytting fra Norge – kartlegging av omfanget. *Unpublished paper, Norwegian only*

[6] UNECE 2007: Register based statistics in the Nordic countries. Review on best practices with focus on population and social statistics

**Appendix: Example of a quality report for registered persons in the Central Population Register**

**Table 1:** Quality report for registered persons in the Central Population Register, 2012-2014.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **01.01.2012** | **01.01.2013** | **01.01.2014** | **Change 2013-14** |
| Quality indicator, per 1000 | 5,4 | 5,7 | 5,2 | 0,5 |
| Number of checks | 29 | 29 | 29 | - |
| Number of positive indicators | 784 396 | 841 040 | 777 584 | (63 456) |
| Number of records (registered persons) | 4 983 756 | 5 049 958 | 5 107 477 | 57 519 |
| Number of records (registered persons) with positive indicators | 496 247 | 523 357 | 469 152 | (54 205) |
| Number of records (registered persons) with negative indicators | 4 487 509 | 4 526 601 | 4 638 325 | 111 724 |
|  |  |  |  |  |
| **C2 Accuracy** |  |  |  |  |
| **C21 Identifiability** |  |  |  |  |
| C21A Use of dicontinued PIN on the transaction | - | 7 | - | (7) |
| **C22 Inconsistent units** |  |  |  |  |
| C22A Spouses/partners have different civil status | 49 015 | 54 768 | 58 814 | 4 046 |
| C22B Spouses/partners have different date for civil status | 86 502 | 93 039 | 97 359 | 4 320 |
| C22C Registered as a resident on Svalbard in the CPR, but not in the Svalbard reg. | 8 | 19 | 5 | (14) |
| C22D Persons in a family have a common family id, but not a common dwelling id | 1 614 | 1 409 | 1 369 | (40) |
| **C24 Measurement errors** |  |  |  |  |
| C24A Invalid municipality number according to the Cadaster | - | - | - | - |
| C24B Invalid street number according to the Cadaster | - | - | - | - |
| C24C Invalid house number according to the Cadaster | 3 701 | 7 028 | 2 293 | (4 735) |
| C24D Invalid entrance number according to the Cadaster | 1 488 | 1 643 | 945 | (698) |
| C24E Invalid sub-number according to the Cadaster | 490 | 689 | 355 | (334) |
| C24F Invalid dwelling number according to the Cadaster | 34 634 | 48 248 | 33 800 | (14 448) |
| C24G Use of property id as address while street address is available in the Cadaster | 4 181 | 7 500 | 2 538 | (4 962) |
| C24H Use of invalid address according to the Cadaster | 116 215 | 128 977 | 103 067 | (25 910) |
| **C25 Inconsistent values** |  |  |  |  |
| C25A Mother’s PIN is invalid | 106 283 | 102 018 | 97 808 | (4 210) |
| C25D Father’s PIN is invalid | 134 902 | 130 588 | 126 460 | (4 128) |
| C25C Spouse’s/partner’s PIN is invalid | 36 800 | 43 590 | 48 660 | 5 070 |
| C25D Dnr indicates birth before 1900 | - | - | - | - |
| C25E Illogical sequence for civil status values | 2 431 | 2 986 | 3 605 | 619 |
| **C26 Dubious objects** |  |  |  |  |
| C26A More than ten registered persons per dwelling | 71 204 | 72 340 | 68 636 | (3 704) |
|  |  |  |  |  |
| **C3 Completeness** |  |  |  |  |
| **C35 Missing values** |  |  |  |  |
| C35A Registered in a multi-dwelling building, but missing dwelling number | 55 445 | 67 203 | 60 070 | (7 133) |
| **C36 Imputed values** |  |  |  |  |
| C36A Incomplete dwelling number (H0000) | 748 | 764 | 816 | 52 |
|  |  |  |  |  |
| **C4 Time-related checks** |  |  |  |  |
| **C44 Object dynamics** |  |  |  |  |
| C44A Birth registered more than 31 days after official date of birth | 2 120 | 2 367 | 2 630 | 263 |
| C44B Death registered more than 31 days after official date of death | 2 001 | 1 094 | 97 | (997) |
| C44C Immigration registered more than 31 days after official date of immigration | 8 802 | 11 065 | 12 391 | 1 326 |
| C44D Emigration registered more than 31 days after official date of emigration | 11 977 | 8 467 | 1 109 | (7 358) |
| C44E Moving registered more than 31 days after official date of moving | 53 286 | 54 294 | 53 800 | (494) |
| C44F Svalbard: Immigration reg. more than 31 days after official date of immigration | 199 | 162 | 143 | (19) |
| C44G Svalbard: Emigration reg. more than 31 days after official date of emigration | - | 419 | 491 | 72 |
| C44H Svalbard: Moving reg. more than 31 days after official date of moving | 350 | 356 | 323 | (33) |

**Table 2:** Analysis of positive indicators by municipality, registered persons in the Central Population Register, 1st January 2014.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Registered persons,**  **1st January 2013 and 1st January 2014** | | | **Newly registered persons, 1st January 2014** | |  |  | **Registered persons,**  **1st January 2013 and 1st January 2014** | | | **Newly registered persons,**  **1st January 2014** | |  |
| **Municipality nr** | **Nr of positive indicators unchanged** | **Nr of positive indicators increased** | **Nr of positive indicators reduced** | **No positive indicators** | **At least one positive indicator** | **Total** |  | **Nr of positive indicators unchanged** | **Nr of positive indicators increased** | **Nr of positive indicators reduced** | **No positive indicators** | **At least one positive indicator** | **Total** |
| Total | 4 832 714 | 47 259 | 96 091 | 107 441 | 23 972 | 5 107 477 |  | 95 % | 1 % | 2 % | 2 % | 0 % | 100 % |
| 0101 Halden | 29 098 | 226 | 258 | 473 | 64 | 30 119 |  | 97 % | 1 % | 1 % | 2 % | 0 % | 100 % |
| 0104 Moss | 29 982 | 293 | 352 | 542 | 106 | 31 275 |  | 96 % | 1 % | 1 % | 2 % | 0 % | 100 % |
| 0105 Sarpsborg | 51 978 | 476 | 584 | 864 | 141 | 54 043 |  | 96 % | 1 % | 1 % | 2 % | 0 % | 100 % |
| 0106 Fredrikstad | 74 684 | 643 | 869 | 1 185 | 195 | 77 576 |  | 96 % | 1 % | 1 % | 2 % | 0 % | 100 % |
| 0111 Hvaler | 4 087 | 45 | 194 | 50 | 4 | 4 380 |  | 93 % | 1 % | 4 % | 1 % | 0 % | 100 % |
| 0118 Aremark | 1 354 | 18 | 19 | 18 | 1 | 1 410 |  | 96 % | 1 % | 1 % | 1 % | 0 % | 100 % |
| 0119 Marker | 3 480 | 16 | 52 | 39 | 8 | 3 595 |  | 97 % | 0 % | 1 % | 1 % | 0 % | 100 % |
| 0121 Rømskog | 653 | 5 | 3 | 9 | 1 | 671 |  | 97 % | 1 % | 0 % | 1 % | 0 % | 100 % |
| 0122 Trøgstad | 5 177 | 41 | 45 | 90 | 18 | 5 371 |  | 96 % | 1 % | 1 % | 2 % | 0 % | 100 % |
| .. | .. | .. | .. | .. | .. | .. |  | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. |  | .. | .. | .. | .. | .. | .. |
| 0301 Oslo | 587 824 | 8 949 | 11 838 | 21 034 | 4 604 | 634 249 |  | 93 % | 1 % | 2 % | 3 % | 1 % | 100 % |
| .. | .. | .. | .. | .. | .. | .. |  | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. |  | .. | .. | .. | .. | .. | .. |
| 2027 Unjárga Nesseby | 853 | 11 | 22 | 18 | 8 | 912 |  | 94 % | 1 % | 2 % | 2 % | 1 % | 100 % |
| 2028 Båtsfjord | 2 026 | 30 | 36 | 86 | 29 | 2 207 |  | 92 % | 1 % | 2 % | 4 % | 1 % | 100 % |
| 2030 Sør-Varanger | 9 496 | 113 | 183 | 202 | 91 | 10 085 |  | 94 % | 1 % | 2 % | 2 % | 1 % | 100 % |

**Table 3:** Entries from 1st January 2013 with an increase in the number of positive indicators, registered persons in the Central Population Register, 1st January 2014.

|  |  |  |
| --- | --- | --- |
|  | **Registered persons with an increased number of positive indicators** | **%** |
| **C2 Accuracy** |  |  |
| **C21 Identifiability** |  |  |
| C21A Use of discontinued PIN on the transaction | - | 0 % |
| **C22 Inconsistent units** |  |  |
| C22A Spouses/partners have different civil status | 5 992 | 13 % |
| C22B Spouses/partners have different date for civil status | 6 456 | 14 % |
| C22C Registered as a registered on Svalbard in the CPR, but not in the Svalbard reg. | 3 | 0 % |
| C22D Persons in a family have a common family id, but not a common dwelling id | 140 | 0 % |
| **C24 Measurement errors** |  |  |
| C24A Invalid municipality number according to the Cadaster | - | 0 % |
| C24B Invalid street number according to the Cadaster | - | 0 % |
| C24C Invalid house number according to the Cadaster | 551 | 1 % |
| C24D Invalid entrance number according to the Cadaster | 153 | 0 % |
| C24E Invalid sub-number according to the Cadaster | 94 | 0 % |
| C24F Invalid dwelling number according to the Cadaster | 7 275 | 15 % |
| C24G Use of property id as address while street address is available in the Cadaster | 992 | 2 % |
| C24H Use of invalid address according to the Cadaster | 24 599 | 52 % |
| **C25 Inconsistent values** |  |  |
| C25A Mother’s PIN is invalid | 1 212 | 3 % |
| C25D Father’s PIN is invalid | 1 808 | 4 % |
| C25C Spouse’s/partner’s PIN is invalid | 4 583 | 10 % |
| C25D Dnr indicates birth before 1900 | - | 0 % |
| C25E Illogical sequence for civil status values | 137 | 0 % |
| **C26 Dubious objects** |  |  |
| C26A More than ten registered persons per dwelling | 22 216 | 47 % |
|  |  |  |
| **C3 Completeness** |  |  |
| **C35 Missing values** |  |  |
| C35A Registered in a multi-dwelling building, but missing dwelling number | 13 024 | 28 % |
| **C36 Imputed values** |  |  |
| C36A Incomplete dwelling number (H0000) | 248 | 1 % |
|  |  |  |
| **C4 Time-related checks** |  |  |
| **C44 Object dynamics** |  |  |
| C44A Birth registered more than 31 days after official date of birth | 21 | 0 % |
| C44B Death registered more than 31 days after official date of death | 4 | 0 % |
| C44C Immigration registered more than 31 days after official date of immigration | 298 | 1 % |
| C44D Emigration registered more than 31 days after official date of emigration | 41 | 0 % |
| C44E Moving registered more than 31 days after official date of moving | 3 098 | 7 % |
| C44F Svalbard: Immigration reg. more than 31 days after official date of immigration | 22 | 0 % |
| C44G Svalbard: Emigration reg. more than 31 days after official date of emigration | 68 | 0 % |
| C44H Svalbard: Moving reg. more than 31 days after official date of moving | 39 | 0 % |
|  | 47 259 |  |

**Table 4:** New entries since 1st January 2013 with at least one positive indicator, registered persons in the Central Population Register, 1st January 2014.

|  |  |  |
| --- | --- | --- |
|  | **Newly registered persons with positive indicators** | **%** |
| **C2 Accuracy** |  |  |
| **C21 Identifiability** |  |  |
| C21A Use of discontinued PIN on the transaction | - | 0 % |
| **C22 Inconsistent units** |  |  |
| C22A Spouses/partners have different civil status | 9 700 | 40 % |
| C22B Spouses/partners have different date for civil status | 10 945 | 46 % |
| C22C Registered as a registered on Svalbard in the CPR, but not in the Svalbard reg. | 2 | 0 % |
| C22D Persons in a family have a common family id, but not a common dwelling id | 15 | 0 % |
| **C24 Measurement errors** |  |  |
| C24A Invalid municipality number according to the Cadaster | - | 0 % |
| C24B Invalid street number according to the Cadaster | - | 0 % |
| C24C Invalid house number according to the Cadaster | 63 | 0 % |
| C24D Invalid entrance number according to the Cadaster | 23 | 0 % |
| C24E Invalid sub-number according to the Cadaster | 4 | 0 % |
| C24F Invalid dwelling number according to the Cadaster | 1 377 | 6 % |
| C24G Use of property id as address while street address is available in the Cadaster | 50 | 0 % |
| C24H Use of invalid address according to the Cadaster | 4 905 | 20 % |
| **C25 Inconsistent values** |  |  |
| C25A Mother’s PIN is invalid | 735 | 3 % |
| C25D Father’s PIN is invalid | 1 643 | 7 % |
| C25C Spouse’s/partner’s PIN is invalid | 9 431 | 39 % |
| C25D Dnr indicates birth before 1900 | - | 0 % |
| C25E Illogical sequence for civil status values | 792 | 3 % |
| **C26 Dubious objects** |  |  |
| C26A More than ten registered persons per dwelling | 6 078 | 25 % |
|  |  |  |
| **C3 Completeness** |  |  |
| **C35 Missing values** |  |  |
| C35A Registered in a multi-dwelling building, but missing dwelling number | 2 338 | 10 % |
| **C36 Imputed values** |  |  |
| C36A Incomplete dwelling number (H0000) | - | 0 % |
|  |  |  |
| **C4 Time-related checks** |  |  |
| **C44 Object dynamics** |  |  |
| C44A Birth registered more than 31 days after official date of birth | 276 | 1 % |
| C44B Death registered more than 31 days after official date of death | 2 | 0 % |
| C44C Immigration registered more than 31 days after official date of immigration | 1 955 | 8 % |
| C44D Emigration registered more than 31 days after official date of emigration | 133 | 1 % |
| C44E Moving registered more than 31 days after official date of moving | 2 497 | 10 % |
| C44F Svalbard: Immigration reg. more than 31 days after official date of immigration | 13 | 0 % |
| C44G Svalbard: Emigration reg. more than 31 days after official date of emigration | 11 | 0 % |
| C44H Svalbard: Moving reg. more than 31 days after official date of moving | 2 | 0 % |
|  | 23 972 |  |

1. The expression Registered Persons is used in this paper to distinguish between persons who are registered in the CPR as being alive and with an address in Norway, and Resident Persons as statistical units in population statistics. Due to statistical inference, there might be a difference between the total number of registered persons in this paper and the number of resident persons in the official statistics. [↑](#footnote-ref-1)
2. Better quality in this case means the introduction of common street addresses. Many rural areas in Norway do not have street addresses. In these areas, the property identification code from the Cadaster is registered in the CPR as a substitute for the street address. [↑](#footnote-ref-2)