Experiment Report Work Package(s): WP1, WP5

October 2016

Name: **Linking the Statistical Register of Employment and the Labour Force Survey**

**1 Introduction**

The purpose of the experiment was to link and compare microdata on persons in employment in two sources: the Statistical Register of Employment (SRDAP) and the Labour Force Survey (LFS). Data in SRDAP are updated monthly by administrative data collected for registration, changes and deregistration for mandatory social insurance in Slovenia. The social insurance data are used by the Health Insurance Institute of Slovenia, the Pension and Disability Insurance Institute of Slovenia, the Employment Service of Slovenia, the Tax Administration of the Republic of Slovenia and the Statistical Office of the Republic of Slovenia (SURS). Employers are obliged by law to turn in the forms for registration of persons in employment and they are the ones to code the data in the forms. The sources for SRDAP are also the Slovenian Business Register, the Register of Spatial Units and some other SURS’s sources. The LFS is carried out in compliance with the International Labour Organisation (ILO) guidelines for labour force statistics and with Eurostat requirements which refer to the harmonised EU survey.

The aim was to get a more thorough insight into the quality of data (educational attainment, occupation, work contract, economic activity and work location) in both sources and also to determine if data validation and data editing rules in SRDAP can be improved. The quality of data in SRDAP is pretty important, since the data are further used in many surveys (e.g. register-based Population Census, EU-SILC, Structure of Earnings Survey, Annual Job Vacancy Statistics, Average Annual Gross Earnings, Statistical Business Register, National Accounts, etc.) and also published monthly.

A test has also been done for the work location of persons in employment by linking the data also to a telephone directory. In SRDAP it is determined by the identification number of a local unit which is entered in the administrative forms and then linked to the Business Register.

**2 General characteristics of SRDAP and LFS for persons in employment**

|  | ***Statistical Register of Employment (SRDAP)*** | ***Labour Force Survey (LFS)*** |
| --- | --- | --- |
| *Unit:* | *Person in employment* | *Person in employment* |
| *Coverage:* | *Full coverage* | *Based on a statistical sample* |
| *Main source:* | *Administrative data* | *Sample survey* |
| *Observation period:* | *Monthly* | *Quarterly* |
| *Reference period:* | *On the last day of the month* | *The activity of the respondent in the week before the interview (from Monday to Sunday)* |
| *Target population:* | *Persons in employment are persons in paid employment with employment contracts and self-employed persons* | *Persons in employment are unpaid family workers and persons working under contracts for work/service or for direct payment, i.e. persons who in the week (from Monday to Sunday) before the interview performed any work for payment (in money or in kind), profit or family gain.* |
| *Definition of the population:* | *All persons are taken into account, irrespective of the definition of the population, i.e. as soon as a person appears in the labour market in Slovenia* | *The sample frame included persons pursuant to the definition of the population; Residents of the republic of Slovenia are persons with registered residence in Slovenia who live or intend to live in Slovenia for a year or more and are not temporarily absent from Slovenia for a year or more. Nationality of a person is not important.* |
| *Relevance:* | *Information on employment at lower territorial levels (statistical regions, municipalities)* | *International comparisons* |
| *Publishing:* | *Monthly in the following fields: Labour market; Labour force, Slovenia* | *Quarterly in the following fields: Labour market; Labour force (in the LFS)* |

**3 Linking and matching data**

We linked the microdata in the two sources by the SID (statistical identifier). In SRDAP the SID is assigned to each person according to the personal identification number each time microdata are obtained from the Health Insurance Institute of Slovenia. The personal identification number is later on dropped and only the SID is used in the microdata according to statistical confidentiality policies of SURS.

The starting point was a sample of 11,950 persons who were in employment in SRDAP and the LFS in the same month. SRDAP monthly situations and LFS quarterly situations were by the month of the last LFS interview in 2015. However, since some changes in employment can occur during the month, we also did a matching on the name of the employer in both sources. In SRDAP the name is imported from the administrative Business Register and in the LFS it is as reported by the interviewed person, so in some cases the name in the data can be quite different although it is actually the same company. There is also variable non-response in the LFS. We only kept the records where the employer in both sources was undoubtedly the same and the variables had no missing values. The sample for the analysis consisted of 10,081 persons who were in employment according to SRDAP and the LFS in the same month, were employed by the same employer and their matching through the SID was undoubted.

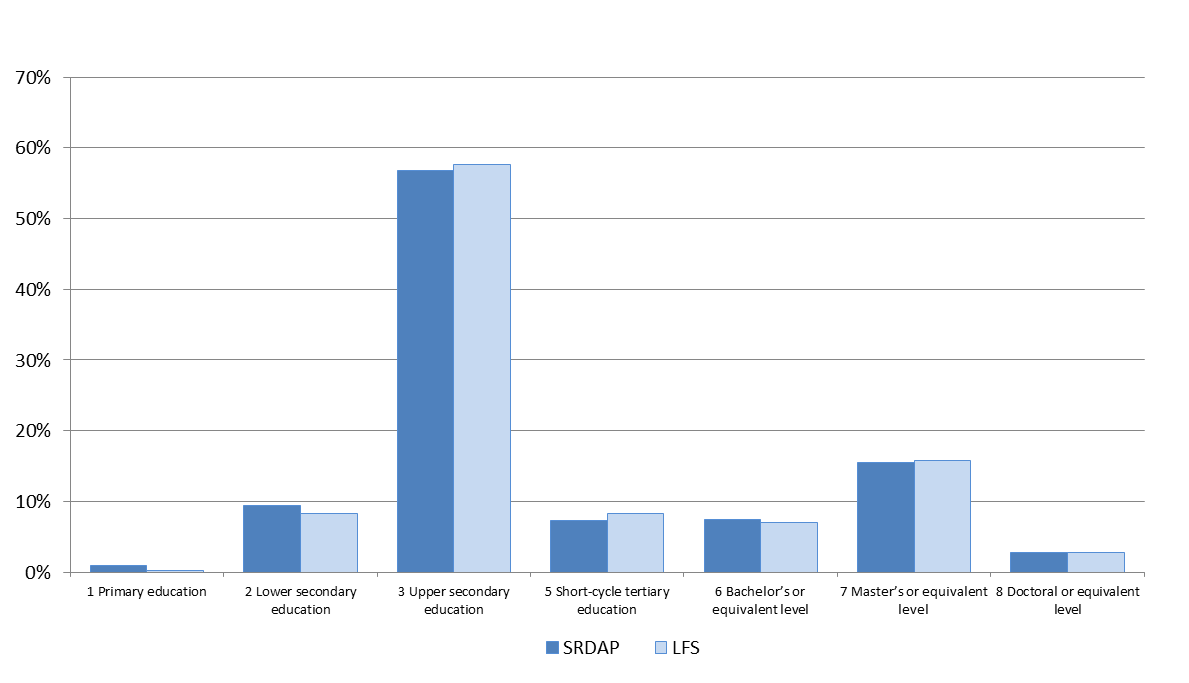
**4 Comparison of register and survey data**

The data on educational attainment, occupation, work contract, economic activities and work location in both sources were compared to identify the differences and to evaluate the quality of data. Most evaluations were done at the lowest level of classifications. Only basic information is provided in this report.

* 1. Educational attainment

Data on educational attainment in administrative sources and sample surveys in Slovenia are collected and classified according to the national Classification System of Education and Training (KLASIUS). KLASIUS is compatible with and for international purposes translated into the International Standard Classification of Education (ISCED). Since employers sometimes do not report the changes in educational attainment for their employees in the administrative source, SRDAP is also annually updated with the data on college and university graduates from education statistics. We compared the data according to the first level of ISCED 2011. Although at the micro level 16% of the sample units have different educational attainment in SRDAP than in the LFS, these discrepancies become hidden in the overall structure (Figure 1) which is quite close for both sources. This is often the case if some classification groups are close or if it is hard to distinguish among them.

Figure 1: Structure of persons in employment in the SRDAP-LFS sample by ISCED 2011, 2015

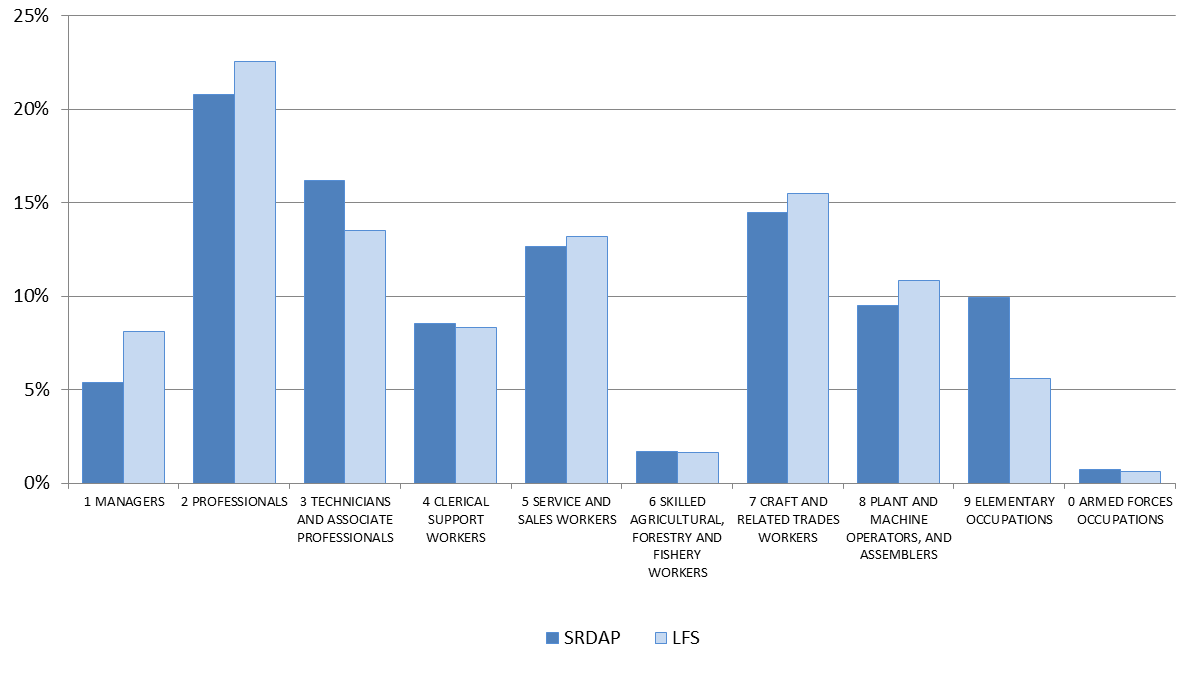


* 1. Occupation

Data on occupations are coded according to national Standard Classification of Occupations 2008 (SKP-08), the structure of which is based on and compatible with ISCO-08. Comparing ISCO-08 and the SKP-08, the first three levels are directly translatable, while at the 4th classification level 10 additional groups are included in the SKP-08. Also some groups are in the SKP-08 deleted at all levels.

In the case of ISCO-08 the differences in the two sources concerning the occupational data are significant (Figure 2). There are several reasons for these differences. First, there is a different methodology in collecting the data. In the administrative source respondent coding is applied - employers code the occupation for their employees, while in the survey expert coding based on open questions about respondent's occupation is applied. ISCO-08 is a complex classification with many groups at the lowest level and understanding of the coding rules is a demanding task. Different coding of occupations with broad scope of tasks is likely to happen. Some groups of occupations are similar and it is hard to define the borders among groups. The classification provides definitions and descriptions, but for some groups an important part of determining the code is left to subjective judgment. Coding at the lowest level of ISCO-08 only from the questions in the LFS proves to be quite difficult for certain occupations. Problematic groups and common mistakes in administrative data had been analysed over the years and the data in SRDAP are monthly edited to minimize coding errors.

Figure 2: Structure of persons in employment in the SRDAP-LFS sample according to ISCO-08, 2015



The most obvious distinctions concerning occupational data in the sample are:

**Major group 1 Managers**

The number of managers in the LFS is obviously higher than in SRDAP. This is mostly the consequence of different methodology in obtaining data, and unclear distinction between managers and other groups. Coding in the LFS results from job titles and short task descriptions (which were answered as managerial); while employers (SRDAP) decided in a lot of cases that other tasks dominated the managerial tasks. Also data in SRDAP are validated and in some cases edited. For example, self-employed persons with no employees reported as managers are in the editing process redistributed to other occupations taking into account other variables such as economic activity, number of employees and educational attainment.

**Major groups 2 Professionals and 3** [**Technicians and associate professionals**](http://www.ilo.org/public/english/bureau/stat/isco/isco88/3.htm)

The number of Professionals in the LFS is higher than in SRDAP, while the number of Technicians and associate professionals is lower. There are several reasons for these differences. Sometimes employers tend to believe that skills needed for the job are not adequate for Professionals, while the LFS coding takes into account obtained education; data in SRDAP are edited according to required skill and only partly according to education obtained; and also some of these differences arise from employers not reporting a change of occupation.

**Major groups 7** [**Craft and related trades workers**](http://www.ilo.org/public/english/bureau/stat/isco/isco88/7.htm) **and 8** [**Plant and machine operators and assemblers**](http://www.ilo.org/public/english/bureau/stat/isco/isco88/8.htm)

The borderline between these two groups is sometimes hard to determine when collecting data on occupations. Employers look at a job title not paying enough attention to major group description, resulting in more occupations in the group Craft and related trades workers. Also coding in the LFS is not always reliable concerning the distinction between craft-oriented and machine-oriented occupations since a job title and the description of tasks do not always give the necessary data (e.g. Sewer and Sewing-machine operator).

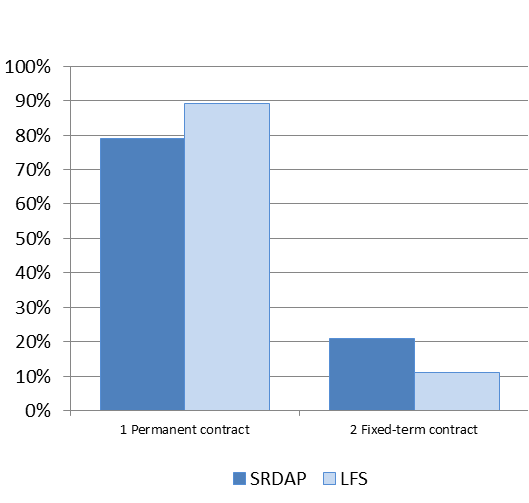
**Major group 9 Elementary occupations**

Employers tend to decide more often that it is an elementary occupation, while in the LFS it is more often coded according to work description in the groups 7 [Craft and related trades workers](http://www.ilo.org/public/english/bureau/stat/isco/isco88/7.htm) and 8 [Plant and machine operators and assemblers.](http://www.ilo.org/public/english/bureau/stat/isco/isco88/8.htm)

* 1. Employment contract

Data on employment contract in the two sources are quite different. In SRDAP there is a larger share of population with fixed-term contracts than in the LFS. This has been a long-known problem, the reason for it originating in employers not reporting the change after the employee is switched from a fixed-term to a permanent contract in the social insurance data. The values for this variable in SRDAP are not corrected, because there is a doubt which value is actually incorrect. Therefore, only the data from the LFS are published and the data from SRDAP concerning employment contracts are not.

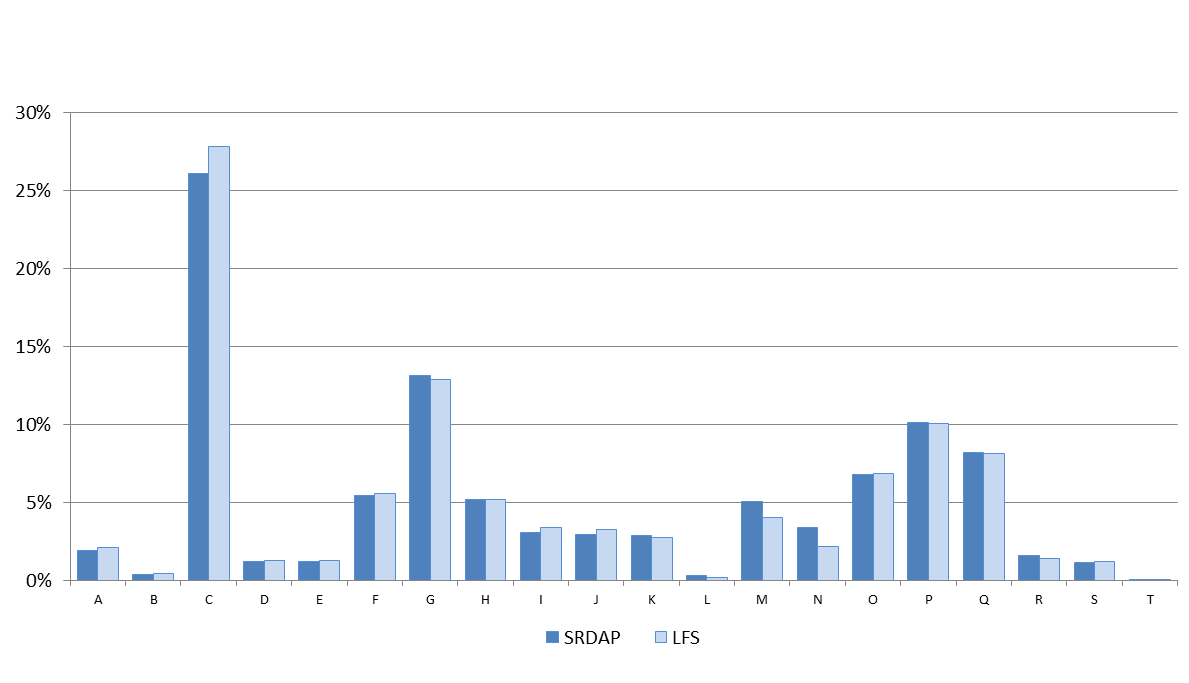
Figure 3: Structure of persons in employment in the SRDAP-LFS sample by employment contract, 2015



* 1. Economic Activities

Economic activities of persons in employment are defined with the Standard Classification of Activities 2008 (SKD 2008), which is harmonized with the international classification NACE Rev. 2. In SRDAP economic activities are determined on the basis of local business units of a business entity from the Business Register at which an employee works and for the farmers the source are social insurance data. In the LFS the activities are coded according to answers about activity in the survey at the two-digit NACE Rev. 2 level. The methodology concerning coding and collection differs. The economic activity differs for 13% of the sample at the first classification level. The difference in the structures by economic activities is the highest in industry, which is larger in the LFS than in SRDAP, while for other activities discrepancies are minor.

Figure 4: Structure of persons in employment in the SRDAP-LFS sample by economic activities, 2015



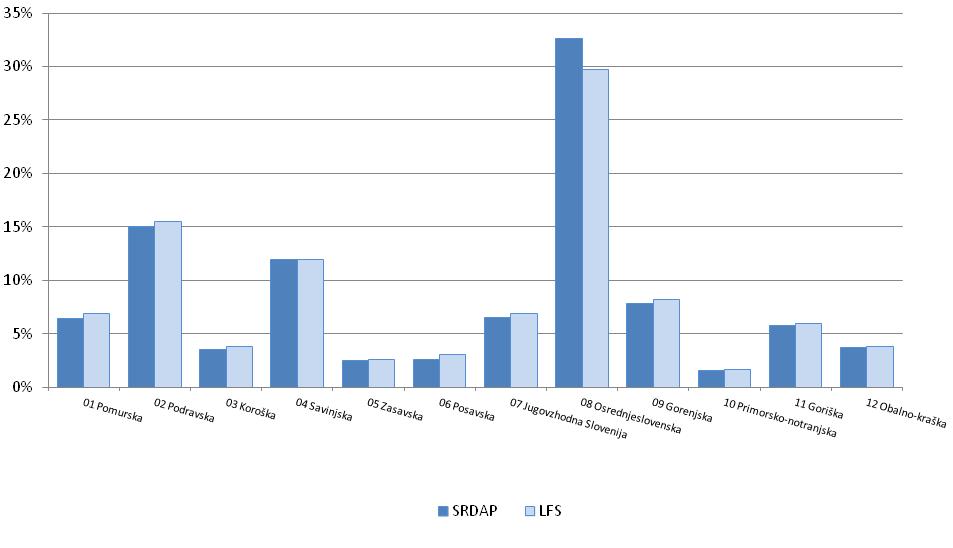
* 1. Work location

We compared the data on work location for 9,901 persons employed at business units (farmers were excluded). The work location of persons in employment in SRDAP is determined by the identification number of a local business unit of a business entity that is filled in the social insurance forms and then linked to the Business Register. At SURS there is also the Statistical Business Register, but these data are not used because SRDAP situations are monthly and the data from the Statistical Business Register are not available at the time, yet. In the past years a few business entities removed local business units from the Business Register to minimize their administration work; consequently, this can have an impact on the territorial distribution of persons in employment. The work location in the LFS is defined by the question on municipality of work; however, municipality in the LFS is not disseminated but only used to further determine the cohesion region. The distribution of the sample persons in employment by statistical regions is in Figure 5.

Slovenia is divided into 212 heterogeneous municipalities. The work location by both sources at the level of municipalities differed for 14% of persons in employment, but there is a question which of the sources is more correct. In some cases work location can be regarded differently in the administrative source. In administrative sources it is linked to an existing local business unit, while in the survey it is reported according to the activity of the respondent in the week before the interview. There would be differences for example in the construction industry, since in the survey the location of the construction site could be reported and also differences in some occupations that are not bound to one location, for example maintainers of railway tracks.

Next, the aim was to analyse if the telephone directory can be applied as a useful source for work location. We linked the data from the telephone directory by business entities' IDs and under the condition that there was at least one local business unit registered in the telephone directory in the municipality that was recorded in the LFS. 1,064 records did not have matching entities ID's in the telephone directory. 7,655 persons (87%) had a work location in the same municipality by both sources, 651 persons had a work location in the LFS that was not equal to SRDAP's location while a local unit was registered in that municipality and the remaining 531 had a different municipality in the LFS than in SRDAP, but also no complying local unit was entered in the telephone directory. Next, we repeated this action with the Business Register and linked the data from the Business Register by business entities' IDs and under the condition that there was at least one local business unit registered in the telephone directory in the municipality that was recorded in the LFS. Out of the persons that had a different location, there were 545 cases with a local unit also registered in that municipality and 637 cases where no complying local business unit was registered in the Business Register. The usefulness of the telephone directory would require some further analyses. In our case it had some more local business units registered where persons reported to be employed by the LFS compared to the Business Register for the business entities that we were able to link.

Figure 5: Structure of persons in employment in the SRDAP-LFS sample by statistical regions, 2015



**5 Conclusions**

Data on educational attainment, occupation, employment contract, and work location of 10,081 persons who were in employment were compared between the two sources: the LFS and SRDAP. The differences between the sample survey and the register were the lowest for educational attainment and economic activities. There are two administrative sources for SRDAP: the primary source is data collected for social insurance and annually these are also updated with the data on college and university graduates from education statistics. There is more variety for occupational data which to some extent show that the way the data are collected impacts the data. The same classifications may result in different data, especially when classifications are complex or the rules of a classification are difficult to apply. There is a respondent coding in the administrative data, while the sample survey is based on open questions and coding is done by the coding personnel. Another issue of administrative sources can be that the data are up to date. There are differences between the LFS and SRDAP concerning the employment contracts, a large part of them originating in employers not reporting the change after the employee is switched from a fixed-term to a permanent contract in the social insurance data. We experimentally applied the telephone directory for the quality assessment of the work location. The weakness of the telephone directory is that not all local units have correct ID's, so some data could not be linked. In our case it had some more local units registered in municipalities in which persons reported to be employed by the LFS compared to the Business Register for the business entities which is the location source for SRDAP. To make a final conclusion on the usefulness of the telephone directory, further analyses would be required.

Some of the advantages and disadvantages of the assessed two sources are summarized below.

|  |  |
| --- | --- |
| **Statistical Register of Employment** | **Labour Force Survey** |
| ***Advantages*** | ***Disadvantages*** |
| *Full coverage of persons in employment* | *Weighted sample* |
| *No extra cost for collecting data* | *Costs related to data collection* |
| *No extra respondent burden* | *Possibility of non-response* |
| *Provides information on employment at lower territorial levels (statistical regions, municipalities)* | *Sample is not representative at the level of municipalities* |
| *Employers have all the data on one's job* | *Another person may answer for the surveyed person* |
| *Data for coding may be scarce* |
| *Coding is done by many reporting units - employers* | *Subjectivity in coding in case of complex classifications (e.g. ISCO-08) may have a significant impact on the results* |
| ***Disadvantages*** | ***Advantages*** |
| *Definition of persons in employment is different from the international LFS definition* | *Provides data according to international concepts and definitions* |
| *Change of data is not always reported by the reporting units to the administrative source (certain data in SRDAP are updated from other sources)* | *Data are collected in time with the reference period* |
| *Complexity of some classifications and because of that different understanding among the employers* | *Coding according to the classifications (occupation, economic activity) is performed by personnel especially qualified for this job* |
| *No uniformity of coding according to classifications* | *The same understanding of coding rules, and uniformity* |
| *Coding errors (shallow knowledge of classification, superficiality, fast coding, etc.)* | *Good knowledge of classifications* |