Classifying people by social class in population based health surveys:
Two methods compared

Steinar Krokstad¹, Kristen Ringdal² and Steinar Westin¹

¹) Department of Community Medicine and General Practice, Norwegian University of Science and Technology, Trondheim, Norway
²) Department of Sociology, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence to: Steinar Krokstad, Department of Community Medicine and General Practice, Norwegian University of Science and Technology, HUNT Research Centre, Neptunveien 1, NO-7650 Verdal, Norway
Telephone: +47 74075180  Telefax: +47 74075181  e-mail: steinar.krokstad@medisin.ntnu.no

ABSTRACT

Aims: In this study we evaluate the accuracy of a reclassification from a 10-category questionnaire-based occupational classification used in health surveys into the Erikson Goldthorpe Portocarero (EGP) social class scheme, by comparing it to the standard procedure based on occupational codes. Comparisons are based on socioeconomic inequalities in self-rated health.

Methods: Individual data on occupation and health in a Norwegian cross sectional total county population, The Nord-Trøndelag Health Study (HUNT) in 1984-86, was linked to 1980 national census occupational code data from Statistics Norway. A cross tabulation comparison of two classification methods was done using Kappa statistics. Inequalities in health were measured by logistic regression models. The study population was economically active men aged 20-59 years.

Results: 57% of all respondents were assigned to the same social class in both social class schemes, 23% of the respondents were classified to the nearby classes, Kappa = 0.47 suggested moderate agreement. The value of Kappa was 0.66, suggesting good agreement, for the most occupationally stable groups using three broad social classes in the analysis. Differences in health inequalities measured by the two different elaborated social class schemes were small. The prevalence odds ratio between social class V+VI+VII versus I+II for perceived health less than good was 2.11 (1.86, 2.38) using the HUNT reclassification method, and 2.07 (1.88, 2.32) using the Nordic Occupational Classification (NYK) reclassification.

Conclusion: Reclassification into the EGP social class scheme from a 10-category occupational classification used in population based questionnaire showed moderate to good agreement compared to the more resource demanding standard method. Fairly similar health inequality estimates were found in the two methods.

NORSK SAMMENDRAG


Resultater: 57% av alle menn ble klassifisert til samme klasse, 23% til nærmeste klasse. Kappa = 0,47 indikerte moderat samsvar. Kappaverdien økte til 0,66 som anses som godt samsvar, når man analyserte på den mest yrkesstabile gruppen og benyttet tre brede yrkesklasser. Det ble funnet små forskjeller i helse målt med de to skjemene: Prevalens odds ratio mellom sosial klasse V+VI+VII og sosial klasse I+II for dårlig selvpåvirket helse var 2,11 (1,86, 2,38) for HUNT-inndelingen og 2,07 (1,88, 2,32) for NYK-inndelingen.

INTRODUCTION

Classification of people by social class in population based health surveys may be done in different ways (1). In questionnaires, participants may be asked to write their occupational title, which later has to be coded and reclassified. In surveys with large numbers of participants this manual coding of self reported occupational titles is very resource demanding. Therefore most health surveys in Norway in recent years apply a simpler procedure by asking participants to fill in a form based on a classification with a limited number of alternatives. The different methods have advantages and disadvantages with respect to validity, economy, use of other resources and comparison to other studies.

Social circumstances across the entire life span influence peoples health. Socioeconomic inequalities in health are found in all countries where social gradients have been studied. Despite rapid economic growth and expanding health care systems after the Second World War, there are persistent (2,3) and even perhaps widening health inequalities in Europe (4). Thus, to investigate, monitor and compare such inequalities have become an increasingly important task for public health research.

In Britain, the well known Registrar General’s classification has been used since early in the 20th century (5,6). Norway has a much weaker tradition in monitoring health inequalities (7). Statistics Norway has published a Norwegian standard of socioeconomic status based on a large number of occupational groups. This allows for interpretation of social inequalities in Norway, but has not been suitable for international comparisons (8). The standard has not been regularly used in measuring health inequalities and is perhaps better suited for research on risk and characteristics of specific occupations.

The socioeconomic status of a person is determined by occupation, education and income together (9). However, these factors are sufficiently distinct to require that they be studied separately in relation to health. Occupational status is relevant because it determines people’s place in the social hierarchy. Many different social class schemes are used in health inequality research (9). Different countries have different schemes to classify people into social classes on the basis of job titles, and even within countries different approaches have been used, like in Norway (7,10). This heterogeneity might seriously impede the comparison, exchange and accumulation of findings from different studies.

The Nord-Trøndelag Health Study (Helseundersøkelsen i Nord-Trøndelag, HUNT), a Norwegian total county population study, was performed as two separate cross-sectional surveys with approximately ten years interval in the mid 1980s and mid 1990s (11). The participants returned a questionnaire (Q1) that was mailed with an invitation to participate in a medical screening. A second questionnaire (Q2) was distributed at the screening stations, which the participants were asked to complete and return by mail. The original occupational classification in the HUNT Study was a version of the Norwegian standard provided by Statistics Norway (8), where the participants had to choose between 10 predefined occupational classes.

When we started to classify people based on their position in the labour market to estimate inequalities in health in the HUNT Study, we constructed an approximation to the international Erikson Goldthorpe Porto-carero (EGP) social class scheme (12,13). Our choice of the EGP scheme was influenced by a recommendation promoted in a WHO report to overcome differences in classification between studies on different populations, thus allowing our data to be compared with results from other studies (9).

Figure 1 shows the original occupational classification in the HUNT Study questionnaire and the reclassification into the EGP scheme used in a recent study on socioeconomic gradients in health (12). This reclassification of pre-formed broad groups may introduce various degrees of random misclassification bias. Without access to International Standard Classification of Occupation (ISCO) or Norwegian (Nordic) occupational classification (NYK) codes, this approach seemed to be the best available to provide a social gradient scale based on occupational status (12). However, the NYK classification has now been available in the HUNT Study, and in the 1980 national census data this classification is available for the total population, and not only for a sample as in later censuses.

The purpose of this study was to make a comparison of the reclassification of the occupational classification in the HUNT Study into the EGP social class scheme (HUNT-EGP) (12), to a standard method using official NYK codes provided by Statistics Norway (NYK-EGP) (14). Furthermore, this was an opportunity to see to what extent differences in method affected estimates of health inequalities.

METHODS

Material

The study population selected for this study consists of men who were economically active in 1980 and aged 20-59 years in the HUNT I Study (1984-86). The NYK codes were taken from the Norwegian national census in 1980. Only economically active people were asked to assign their job title in the census. In the HUNT Study, however, people were asked to assign their present or last held occupation.

The time lag between collecting the NYK data and HUNT data, may introduce random misclassification bias owing to some people having changed occupation in the period in-between. Thus, we compared groups where we expected high occupational stability to groups where we expected a high rate of occupational mobility. Occupational mobility is more frequent among younger people than among older.
Every citizen in Norway is given a unique “national identity number” of 11 digits at the time of birth, containing information on birth date and gender. This identity number enabled the individual linkage between collected information in the HUNT Study and the official NYK occupational codes from Statistics Norway. A previously published algorithm for reclassification of the NYK codes into the EGP scheme in Norway was applied (14).

Erikson Goldthorpe Portocarero (EGP) social class scheme

The EGP scheme is considered a reasonable and internationally applicable socioeconomic “gradient scale”, although it is not intended to produce a one-dimensio- nal measure. The scheme is developed on the basis of an explicit set of principles for grouping occupations and is a validated measure of employment conditions designed without reference to health data. Its use in the analysis of morbidity and mortality differences may therefore be regarded as a test of the hypothesis that the employment relations encountered in the different occupational groups are related to health experience in those groups (9,12,15).

The EGP scheme has been compared to the British Registrar General’s scheme; the same differences by social class using both schemes on identical materials have been demonstrated (15).

Classifying women by socioeconomic status always raises the question of whether their occupational status should be derived from their own occupation or from the partners’. Thus, results are mainly presented for men in this article, although the analyses were performed for both genders.

Health outcome variables

The health or morbidity indicators used were self perceived health and any long-standing health problem.

- **Perceived health** was measured by the question “How is your present state of health?” (translated from Norwegian) and the answer categories were “very good”, “good”, “fair” and “poor”. The variable was dichotomised (fair and poor versus good and very good) to yield the variable “perceived health less than good”.

- **Any long-standing health problem** was recorded by asking, “Do you suffer from any long-standing limiting somatic or psychiatric illness, disease or disabil- ity”? The answer categories were “yes” and “no”.

**Figure 1.** Transforming the original occupational group data in the Nord-Trøndelag Health Study into the Erikson Goldthorpe Portocarero (EGP) social class scheme (12,13).
Statistics

Agreement between the two methods for elaborating the EGP social class scheme was estimated by calculating the proportions of exact agreements and kappa statistics (16). The association between social class and health using the two different EGP schemes was measured by calculating the age adjusted odds ratio (OR) of having a health problem between the social classes and, as a summary measure, between two broad classes; blue collar workers (class V+VI+VII) and white collar workers (class I+II) by logistic regression analyses (17). A second summary measure applied in this study was the regression based Relative Index of Inequality (RII). This index is recommended when making comparisons of health inequalities over time or across populations (9), its advantage being that it takes into account the different prevalences of morbidity in all the different groups (not only the highest and lowest social class) and also the relative size and position of each group. When using this method, the socioeconomic status of each occupational group is quantified as the relative position of that group in the occupational hierarchy. This continuous measure of socioeconomic status is then related to morbidity prevalences in the groups by means of a logistic regression model, since the morbidity indicators were defined in a dichotomous way. The resulting OR can be interpreted as the relative risk for having a health problem at the bottom compared with the risk at the top of the occupational hierarchy. A more comprehensive explanation of this method is beyond the scope of this paper, but can be found elsewhere (18).

Results

Figure 1 shows a facsimile of the original occupational classification in the HUNT Study questionnaire (in Norwegian) and the reclassification into the EGP scheme as done in a previous study (12).

Table 1 provides a comparison of the EGP social class scheme elaborated by two different methods for men; from a reclassification of the HUNT occupational classes (HUNT-EGP) and from the standard NYK national census occupational code (NYK-EGP). 57% of all respondents were assigned to the same social class in both social class schemes, an additional 23% of the respondents were classified to a nearby class, Kappa = 0.47 suggested moderate agreement. The highest non-agreement was between the classifications into EGP social class I and II. 32.2% of the respondents could not be classified in the HUNT-EGP compared to 9.0% in the NYK-EGP. The value of Kappa for men aged 20-29, 30-39, 40-49 and 50-59 years was 0.30, 0.48, 0.51 and 0.52 correspondingly. Using three broad social classes (I+II, III+IV, V+VI+VII) the corresponding values were 0.33, 0.59, 0.65 and 0.66.

Table 2 shows age adjusted prevalence OR with 95% confidence interval (CI) for perceived health less than good and any long standing health problem by social class using the HUNT-EGP and the NYK-EGP. The OR between social class VII and I was 2.24 using the HUNT-EGP, 2.93 using the NYK-EGP for perceived health less than good. For any long standing health problem the VII versus I ORs was 2.30 for the HUNT-EGP and 2.21 for the NYK-EGP. For the other classes there was a tendency to higher ORs using the NYK-based reclassification.

In Table 3 we show the health inequalities found using two different summary measures to investigate whether the two different methods of elaborating the socioeconomic classification gave different results. Using broad classes (blue collar versus white collar workers) the agreement between the two methods was good. The OR for perceived health less than good between blue collar workers and white collar workers (V+VI+VII/I+II) was 2.11 for the HUNT-EGP and 2.07 for the NYK-EGP. For any long standing health problem the V+VI+VII versus I+II OR was 1.99 and 1.87 correspondingly. Lastly, we applied the Relative Index of Inequality, which takes into account all groups, the different prevalence in the groups and the relative size of the groups. The agreement between the two methods was an OR equal 2.90 for the HUNT-EGP compared to 3.02 for the NYK-EGP for perceived health less than good and 2.62 compared to 2.51 for any long standing health problem.

Discussion

The comparison made in this study suggested a moderate agreement between the two methods, with a Kappa estimated at 0.47. However, when applying the two different schemes in health inequality analyses, the agreement between the results was good when using broad groups as well as when using the regression-based summary measure Relative Index of Inequality, which takes all groups into account.

The time lag between the National census in 1980 and the HUNT Study in 1984-86 may have introduced some random misclassification. It is difficult to interpret exactly how much this could have biased the results. However, the comparison of the value of Kappa for the different age groups suggests that the Kappa could have been somewhat higher at 0.52 for men. Using three broad social classes (I+II, III+IV,
Table 1. Classification of economically active men aged 20-59 years in the Nord-Trøndelag Health Study according to Erikson Goldthorpe Social Class Scheme elaborated by two different methods. Kappa = 0.47.

<table>
<thead>
<tr>
<th>Social class, NYK (Nordic occupational classification) reclassification</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V+VI</th>
<th>VII</th>
<th>Unknown</th>
<th>Total (%)</th>
<th>Unknown excluded (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Higher-grade admin./prof.</td>
<td>591</td>
<td>900</td>
<td>322</td>
<td>146</td>
<td>162</td>
<td>121</td>
<td>2,311</td>
<td>8.3</td>
<td>12.6</td>
</tr>
<tr>
<td>II Lower-grade admin./prof.</td>
<td>148</td>
<td>1,131</td>
<td>196</td>
<td>54</td>
<td>175</td>
<td>163</td>
<td>1,952</td>
<td>7.0</td>
<td>10.7</td>
</tr>
<tr>
<td>III Routine nonmanual employees</td>
<td>48</td>
<td>307</td>
<td>585</td>
<td>68</td>
<td>134</td>
<td>352</td>
<td>1,565</td>
<td>5.6</td>
<td>8.6</td>
</tr>
<tr>
<td>IV Self-employed, farmers, fishermen</td>
<td>54</td>
<td>327</td>
<td>3,625</td>
<td>443</td>
<td>743</td>
<td>365</td>
<td>5,734</td>
<td>20.5</td>
<td>30.8</td>
</tr>
<tr>
<td>V+VI Foremen, skilled manual workers</td>
<td>11</td>
<td>95</td>
<td>127</td>
<td>297</td>
<td>2,187</td>
<td>924</td>
<td>4,018</td>
<td>14.4</td>
<td>20.9</td>
</tr>
<tr>
<td>VII Unskilled manual workers</td>
<td>9</td>
<td>38</td>
<td>115</td>
<td>280</td>
<td>605</td>
<td>1813</td>
<td>3,310</td>
<td>11.9</td>
<td>16.4</td>
</tr>
<tr>
<td>Unknown</td>
<td>292</td>
<td>1,154</td>
<td>870</td>
<td>1,515</td>
<td>1,761</td>
<td>1,043</td>
<td>9,019</td>
<td>32.3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,153</td>
<td>3,952</td>
<td>2,392</td>
<td>5,908</td>
<td>5,451</td>
<td>6,541</td>
<td>27,909</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td>4.1</td>
<td>14.2</td>
<td>8.6</td>
<td>21.2</td>
<td>19.5</td>
<td>9.0</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown excluded (%)</td>
<td>4.9</td>
<td>16.1</td>
<td>8.7</td>
<td>25.2</td>
<td>21.2</td>
<td>23.9</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Age adjusted prevalence odds ratio with 95% confidence interval (CI) for self perceived health less than good and any long-standing health problem by socioeconomic status using Erikson Goldthorpe Portocarero social class scheme elaborated by two different methods: Reclassification of occupational classes in the Nord-Trøndelag Health Study (HUNT) 1986-86 versus reclassification of Nordic occupational codes (NYK) from 1980. The study population was economically active men aged 20-59 years.

<table>
<thead>
<tr>
<th>Erikson Goldthorpe Portocarero social class scheme</th>
<th>Perceived health less than good, OR (95% CI)</th>
<th>Any long standing health problem, OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HUNT reclass. 95% CI</td>
<td>NYK reclass. 95% CI</td>
</tr>
<tr>
<td>I Higher-grade admin./prof.</td>
<td>1.00 Reference</td>
<td>1.00 Reference</td>
</tr>
<tr>
<td>II Lower-grade admin./prof.</td>
<td>1.05 (0.85, 1.29)</td>
<td>1.60 (1.24, 2.07)</td>
</tr>
<tr>
<td>III Routine nonmanual employees</td>
<td>1.40 (1.13, 1.73)</td>
<td>2.16 (1.65, 2.83)</td>
</tr>
<tr>
<td>IV Self-employed, farmers, fishermen</td>
<td>2.08 (1.78, 2.43)</td>
<td>2.88 (2.26, 3.67)</td>
</tr>
<tr>
<td>V+VI Foremen, skilled manual workers</td>
<td>2.07 (1.75, 2.43)</td>
<td>3.04 (2.38, 3.90)</td>
</tr>
<tr>
<td>VII Unskilled manual workers</td>
<td>2.24 (1.90, 2.66)</td>
<td>2.93 (2.29, 3.74)</td>
</tr>
</tbody>
</table>

Table 3. Summary measures, age adjusted odds ratio with 95% confidence interval (CI) for self perceived health less than good and any long-standing health problem by socioeconomic status using Erikson Goldthorpe Portocarero social class scheme elaborated by two different methods: Reclassification of occupational classes in the Nord-Trøndelag Health Study (HUNT) 1986-86 versus reclassification of Nordic occupational codes (NYK) from 1980. The study population was economically active men aged 20 - 59 years.

<table>
<thead>
<tr>
<th>Summary measure</th>
<th>Perceived health less than good, OR (95% CI)</th>
<th>Any long standing health problem, OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HUNT reclass. 95% CI</td>
<td>NYK reclass. 95% CI</td>
</tr>
<tr>
<td>EGP social class V+VI+VII / I+II</td>
<td>2.11 (1.86, 2.38)</td>
<td>2.07 (1.88, 2.32)</td>
</tr>
<tr>
<td>Relative Index of Inequality based on the EGP scheme</td>
<td>2.90 (2.46, 3.41)</td>
<td>3.02 (2.55, 3.57)</td>
</tr>
</tbody>
</table>

V+VI+VII) for the highest age groups in men, the Kappa reached 0.66. These values suggest moderate to good agreement, and are probably close to the agreement that would have been found in the case that the NYK census records and the HUNT data were collected at the same time.

When comparing the EGP social class schemes generated by the two different methods, a greater proportion of the population was assigned to social class I by the HUNT-reclassification compared to the NYK-based reclassification. A variable assessing the number of employees was not available for self-employed men in the national census data. Some of the non-agreement in classification may be explained by this problem, because this variable would have moved some people from social class IV to social class I in the NYK-reclassification program. But this misclassification does probably not explain the total difference.
The HUNT-reclassification assigned a higher proportion of people to social class IV and a lower proportion to social class VII compared to the NYK-reclassification. This may partly result from agricultural workers being assigned to group IV and not to group VII in the HUNT reclassification, because the HUNT classification did not differentiate between employed and self-employed farmers and fishermen.

Another difference between the NYK-EGP and the HUNT-EGP is that the respondents in the census gave information only about their occupation title. In the HUNT Study the respondents picked one of 10 occupational classes. The occupational classes in the HUNT questionnaire were not presented as social classes ranged from high to low or vice versa (Fig 1), but still there is an element of subjective assignment of occupational class. This might introduce some social desirability bias. This possible source of bias might explain some of the greater proportion of the population assigned to social class I in the HUNT-EGP compared to the NYK-EGP.

The proportion of unclassified people was higher in the HUNT-reclassification than in the NYK-reclassification. The relatively high non-response for these HUNT data was due to procedure: The question for occupational class was not included in the initial questionnaire (Q1), but to a second questionnaire (Q2) that should be returned after the screening day, and therefore subject to higher non-response rate. We do not know the effect of this selection, but sub-analyses we performed showed that the mean health levels among the responders and non-responders according to this question was equal, indicating that the selection problem was small.

Internationally, two different conversion schemes have been used for reclassification of occupational codes to the EGP scheme, the original scheme developed by Erikson, Goldthorpe and Portocarero (13) and a modification of their scheme by Gazzeboom, Luijkx and Treiman (19). These methods were compared using Swedish Level of Living Survey data of 1991. In that study 60% of all respondents were assigned to the same social class (20). Despite using the same occupational codes as source data, the agreement in this comparison was not much higher than in our comparison of methods.

Due to the time lag between collecting the NYK-data and the health data in HUNT, some changes in health and health related mobility might have occurred in-between. In spite of the relative short time period, some random misclassification might have occurred. The most likely result from such processes would be somewhat higher health inequalities measured with the NYK-EGP scheme compared to the HUNT-EGP scheme, owing to faster health deterioration in lower social classes compared to higher social classes. The results presented in Table 2 is consistent with this hypothesis, with slightly higher ORs in most social classes measured with the NYK-EGP scheme. However, the overall conclusion was that there was no systematic bias in results from the health inequality analyses using the two differently elaborated EGP social class schemes. As shown in Tables 2 and 3, what scheme creating the greatest inequalities depended just as well on choice of health variable and inequality measure. Very similar results on health inequalities emerged comparing broad classes like blue-collar versus white-collar workers and when the RII was applied.

These results suggest that the method originally used in the HUNT Study for social class stratification (12), a method that require relatively limited resources, might be used if more standard methods are unavailable. The national and international accumulation of knowledge on patterns and causes of health inequalities would greatly benefit from application of a common social class scheme. In this study we argue in favour of using the Erikson Goldthorpe Portocarero scheme.

The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between The HUNT Research Centre, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), Verdal, The Norwegian Institute of Public Health, and Nord-Trøndelag County Council. We want to thank Jon Ivar Elstad for valuable comments to this study.

Funding: The Norwegian Research Council financed the study.

Conflicts of interest: none.

REFERENCES