



# THE *WAGEINDICATOR* WEB-SURVEY: THE WAGEINDICATOR QUESTIONNAIRE OUTLINE AND KEY CONCEPTS

Kea Tjeldens, AIAS/University of Amsterdam, Netherlands

15/01/2007

For downloadable documents related to this text, please visit:

[www.wageindicator.org](http://www.wageindicator.org) > Research lab > [Downloads](#)

## Measuring employment status, contract and informal labor

Employment status is asked in the first question. The related variable in the dataset is called **contst**. It distinguishes the individuals in self-employment and dependent employment. However, the boundaries between the two categories are not always so clear. For some countries, therefore additional categories are included. For the routing, however the distinction between self-employed and employees is clear.

In many Labor Force Surveys, the information applies to the job held in the reference week. In the *WageIndicator* survey, the job refers to the job held at the time of completing the questionnaire. For the unemployed, the questionnaire is completed for the last job held. An additional question asks for the calendar year when the last job ended.

For individuals in dependent employment, a large variety of employment contracts vary greatly across countries. Yet, a permanent contract is very common in most countries. Therefore, the first question in section E is '*Do you have a permanent contract?*', to be answered with Yes or No. The related variable is called **CONTR7**. If ticked 'No', a question pops up '*What kind of employment contract do you have?*'. The Table shows a list of non-permanent contracts, one of which can be ticked. **contract**.

*The variation in non-permanent employment contracts*

Value	Non-permanent employment
2	Fixed-term contract with the prospect of permanent employment
3	Fixed-term contract without the prospect of permanent employment
4	On-call contract
5	Zero-hours contract
6	Apprentice, trainee
7	Contract with temp agency
8	Labour pool / job pool
9	Seasonal work
10	Other type of contract
11	Combination of contracts
12	No written contract
13	Replacement contract (fixed-term contract to replace another
14	Community work scheme / subsidised employment
15	In community or military service
16	Contractor / labour-only subcontractor
17	Project work
18	Occasional work
19	Sub-contracting or 'dependent' self-employed
20	Interim contract
21	Fixed-term contract to cover another worker
22	New Deal
23	Job creation scheme
26	No written terms and conditions
27	Sub-contracting

Workers in the informal labor market are identified by questions, asking if the respondent has a signed work card, receives payment in cash, is paid in kind, and is paying taxes. In India, for example, a few additional questions ask who provides for the equipment the worker uses to perform the job

## Measuring education

Education is measured with the question: 'What is the highest level of education you have attained (with certificate)?'. For respondents having indicated that they are a school pupil or student in full-time education, the question is: 'At what stage of education are you at the moment?'. The two variables are recoded into one variable **EDUCAT**.

The answer consists of a radio-button list of country-specific educational categories (not more than 20). The national list does include the major national educational categories, but also previous major educational categories, i.e. a 50-year employee who has performed an education that today is replaced by another still must be able to tick the appropriate button in the question.

For reasons of cross-country comparison of the data, the national educational categories are recoded into the worldwide International Standard Classification of Education (ISCED) classification, which was designed by UNESCO, known as ISCED 1997. The variable **EDUISCED** ranges from 0 (No education) to 6 (Second stage of tertiary education). See for details about ISCED [www.unesco.org/education/information/nfsunesco/doc/isced\\_1997.htm](http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm). However, for a few countries, a recoding scheme is not yet available.

As for the type of education, in the *WageIndicator\_NL* survey number 1a and 1b this question was asked in addition to the education-question. In later survey numbers, the question was dropped, because too many respondents used the open-ended option to indicate their special diploma they had ever received. Not only did this require a huge effort for recoding, many answers could not be identified too.

## Measuring job level

A few questions allow for identifying the respondents' job level independent of the educational level attained. First, for the Netherlands the respondents' occupation is coded according to the Standard Occupational Classification of Statistics Netherlands, which includes a code for the required education in five levels. For other countries, occupations are coded according to the International Classification of Occupations (ISCO). ISCO also identifies a job level, but here the levels are not tested, as is the case for the Netherlands.

Second, one question addresses the learning time for the respondent's occupation: 'How much time is needed to become fully effective in your job for someone with your qualifications?'. The answers range from no training period required to more than 1 year, using eight steps. This is a very commonly used indicator when tested job levels are absent. **edujobtr**

Third, one question asks 'Do your qualifications match your job?'. Response categories are Yes/No I am overqualified for my job/ No, I am under-qualified for my job/I don't know. This question measures self-perceived discrepancies between the education level and the job level. **educlevl**

Fourth, training efforts at the workplace are measured with two questions. These are 'How much training have you received, paid for or provided by your EMPLOYER, over the past year in order to improve your skills?', and 'How much training have you paid for YOURSELF over the past year in order to improve your skills?', using the same eight response categories. **training trainin1**

Finally, the self-perceived importance of training for a respondent's labor market mobility is measured in the question 'If you were looking for another job, how important would the following aspects be? .... Suitable training opportunities'. The response ranges from 1 Not at all important to 5 Very important. **seektra1**

## Measuring blue and white collar workers

In some countries, blue-collar workers are clearly distinct from white-collar workers and from professional or managerial jobs or technical staff. In other countries, the distinction is not relevant. Respondents from the former countries get a question asking about blue or white collar in the private sector. For respondents in the public sector, a similar question is asked, distinguishing the civil servant. **CAOCATE**

## Measuring employees' workplace representation

Countries vary with regard to their workplace representation system. *WageIndicator* asks first for presence of workplace representation at the workplace. It lists a variety of systems, ranging from shop steward systems co-operative committees, whereby in each country only the applicable items are switched on. The answers are Yes/No/Not applicable.

Second, questions ask whether the respondent is acting as a workplace representative him or herself. Here a variety of answers can be ticked, ranging from being a shop steward to being a member of a works council.

Third, questions ask for membership of trade unions, staff associations or professional organizations. Here the answers vary also per country.

Once ticked 'Yes' for trade union membership, in some countries a follow-up question is posted: 'Of which trade union are you a member?'. **memtrad4** Here, a country-specific TRADE UNION search tree is used, allowing for a choice of the national or regional trade unions or trade union confederations. Currently, the search tree is used in seven countries (FI, DE, UK, NL, BE, KR, HU), for the remaining countries the question is switched off.

## Measuring collective bargaining coverage

Collective bargaining systems vary greatly across countries. Therefore, the questions about bargaining coverage slightly vary across countries. The answers are Yes/No/I don't know/Not applicable. In the dataset, the related variable is called **CAOFIRM**. Some countries also have one, two or even three follow-up questions. These questions ask whether the agreement is a company, industry-wide or national agreement.

For three countries, it is not necessarily the case that if the respondent's firm is covered by an agreement, the respondent is also covered by the agreement. In Finland, Denmark and Netherlands, respondents may not be covered by an agreement, but the company that they work for is covered.

The breach of the collective agreement matters in a few countries. In Poland and Hungary, it is important whether apart from wages working conditions are also negotiated in the agreement. In these countries a follow-up question asks about this issue.

Finally, one question asks about the employees' opinions whether it is important to be covered by a collective bargaining agreement, regardless whether they are covered or not. **CAOIMPO**

As for collective agreements, three countries deploy a question 'By which collective agreement are you covered', shown if the respondent has ticked 'yes' on the question of collective bargaining coverage. Brazil, Belgium and the Netherlands have provided a list of all collective agreements, bargained in their country. The remaining countries either could not provide a list of all agreements in their country, or do not have a system of collective agreements. Belgium has approximately 200 agreements, Brazil some 600 and the Netherlands almost 800. In Belgium and the Netherlands, the first tier of the search tree is according to industry, allowing the respondents to choose their collective agreements easily in two tiers. In Brazil, the search tree is structured according to region. At the end of 2006 the Netherlands collective agreement code-set was updated. The dataset holds the names of the agreements, but in the national languages only.

## Measuring working hours

The *WageIndicator* questionnaire has a smart routing as far as working hours is concerned. This routing was based on an overview study on questions about working hours in surveys (Tijdens and Dragstra, 2007). The questionnaire starts asking the

respondents in dependent employment whether they have agreed their working hours with their employer, either in writing or verbally **hrscont3**. If yes, they are asked about the sort of hours agreed, notably full-time, part-time, annualized, flexible, on call, opt-out or other **hrscont4**. The respondents with full-time, part-time or other hours are asked how many hours per week they work under the terms of their contract **hrscontr**. The respondents with annualized hours are asked how many hours per year they work **hrscont5**. All respondents with annualized, flexible or on call hours are asked whether they have agreed a minimum or a maximum number of hours **hrscont6**. If so, the minimum and/or the maximum number of hours per week are asked **hrscont7** and **hrscont8**. Next, the respondents with part-time, annualized, flexible, on call, opt-out or other hours are asked what the full-time working week is at their workplace **hrsfirm**. Then, the respondents with full-time, part-time, minimum or maximum hours are asked how overtime hours, defined as 'more hours than laid down in your contract', are compensated **wageotim**. The response categories are: paid as normal hours with premium, paid as normal hours, time-off in lieu for overtime hours, partly paid and partly compensated with time-off in lieu, not compensated, don't know or not applicable. Note that this question is asked to all respondents with contractually agreed hours, regardless of having overtime hours or not. The incidence of overtime hours is asked, using the question 'Do you usually work the number of hours laid down in your contract' **hrsreal0**.

After having determined the contractual and collectively agreed working hours, the questionnaire continues asking about the usual hours **hrsreal**. This question is asked to the respondents who have reported usually working more hours than contractually agreed, to respondents in dependent employment with no contractually agreed working hours, to respondents in dependent employment with annualized, flexible or on call hours without a maximum number of hours, and to all respondents not in dependent employment. Thus, here the self-employed are asked for their usual working hours. Finally, a question asks to all respondents who ever had a job how many days a week they work. This question is only asked to check the reliability of the reported hours **hrsday2**.

One question asks whether the respondent has worked the usual number of hours in the last seven days **ddwork1**. If having worked either less or not at all, a follow-up question asks for the reasons, offering a checkbox with some twelve items, ranging from days off to labor dispute, weather conditions, or seasonal work. This question is asked because it facilitates estimations of annual working hours when only weekly hours are reported. Additionally, a number of countries ask this question in their Labor Force Surveys.

In section E, when the respondents in dependent employment are asked for their wage, a follow-up question asks about the number of waged hours in case the contractual and usual hours differ more than one hour **hrswage**. For the calculations of the hourly wage, the hours-input is the variable **hrswag1**. In case of discrepancies in the dataset across the reported hours, this variable is initially based on contractual hours. In case paid overtime is reported and included in the reported wage, the usual hours are considered the weekly waged hours. This is particularly checked for part-time workers with paid overtime hours. In case of self-employment, the usual hours are considered the weekly waged hours. In case of missing data on contractual or usual hours, the waged hours are considered the weekly waged hours.

Summarizing, the *WageIndicator* questionnaire measures working hours in four ways:

- contractual working hours, in case the respondent has an employment contract in which weekly working hours are agreed; in case flexible, on call or annualized hours are agreed, the minimum and maximum weekly hours or the annual hours are asked;
- usual working hours, registered for those individuals not having agreed hours or not having an employment contract and for those individuals whose usual hours differ from their contractual hours; in India and South Africa, the usual working hours per day and the days worked monthly are asked, which are calculated into weekly working hours;
- standard working week at the workplace, asked to part-timers or flexible workers only;
- waged hours per week for the last wage.

## Measuring wages and calculating hourly wages

Before asking employees about their wages, a first question identifies whether the reported pay is primarily based on individual performance. **wageflex**. If so, it is asked what share of the pay is depending on individual performance. The answers range from 0-20% up to 80-100%. **wagefle1**

Next, a question is asked whether the individual is paid into a bank account or cash in hand. **wageregu**

In countries where applicable, a question asks about the currency of the wage. **WAGECUR**

Then, the employees are asked 'Do you know your GROSS and your NET wage?'. Depending on the answer, questions follow for the last gross and/or net wages. If paid cash, only the net wage is asked.

The next question asks for the payment period, ranging from 1 hour to 1 year. The questionnaire has an open-ended question in case the respondent wants to add information to the ticked payment period. **WAGEPER3** This is an obligatory question, so the respondent can't proceed without having ticked an answer. If reported different contractual and usual hours, the hours-base for the wage is asked.

These questions are followed by three questions about monetary and non-monetary benefits. These vary greatly across countries.

The self-employed have a question about their gross annual income, **wagegr3**, followed by a question whether this income was earned in 12 months or less. **wagemon4**. If the answer is less than 12 months, it is asked how many months.

In order to compare wages, the reported wages have been converted into hourly rates based on the number of hours per week and corrected for the period covered by the payment. This is usually one month but could be four weeks, one week or otherwise. Where the reported contractual hours per week were zero or close to zero, the actual hours worked have been used for the calculations. Thus the calculated hourly wage rates exclude allowances, variable income elements, holiday allowances, expense allowances and overtime bonuses.

## Measuring employment history and future employment

The questionnaire addresses in great detail the respondents' work and family history. For each respondent minimal four and maximal nine calendar years are collected. As regards family history, all respondents are asked for their year of birth **yybirth**. In case the respondent has children, year of birth of eldest child **yyolchld** and year of birth of youngest child are asked **yyyochld**. In some countries, respondents are also asked whether they or their partner expect to have a baby within three years time **hhbaby**.

As regards work history, all respondents tick year of labor market entrance **yyfstjob**. In case of job mobility, respondents in dependent employment are asked in what year they joined their current employer **yycuempl**. In case respondents have occupied more than one many position with your current employer, they are asked when they attained their current job **yycuposi**. If self-employed, respondents are asked in what year they started their own or their family businesses **yycuemp1**. If the respondent has had a career break for at least one year (some countries three months), the year of the break and the year of re-entering the labor market are asked **yybreak yyreenter**.

Note that all above-mentioned questions are asked in calendar years instead of years. The reason to do so is that it is assumed that the visitor has a better memory for calendar years than for years, i.e. what is your year of birth instead of what is your age? Second, in comparison to calendar years, years may more likely cause mistakes in data analyses, because all years have to be controlled for survey year.

As regards future work status, a question addresses whether the respondents in dependent employment expect to be with their employer in a year's time **jobfutu1**. If yes, will this be in the same position or in another position **jobfutu4**? Unemployed are asked whether they expect to have a job in a year's time **jobfutu2**. Self-employed are asked whether they expect to have their your businesses in a year's time **jobfutu3**. If no, all groups are asked why this would not be the case **jobfutu5**. Do they expect to be unemployed, without their business, in education, retired, performing housework or so?

## Measuring supervisory position

The question on supervisory position is an important one, but causes measurement difficulties. These must be solved by explaining how the concept of supervisory position must be understood. In Germany, with its large vocational training system, many workers supervise trainees. In Spain with its large share of small enterprises, supervision is primarily understood as foreman.

In September 2004, we used the question 'How many people work directly under your supervision?', offering the respondents a text box where they could type in a number. For two reasons, the *WageIndicator* team was not fully satisfied with the response. First, it turned out that respondents instead of typing a 0 when they did not supervise anyone rather skipped the question. Second, initially the textbox had no check on whether only numbers were typed and no check on an upper limit.

In 2005, we changed the routing of the question. The first question was 'Do you have a supervisory position?' followed by an instruction 'This applies to people working directly under your supervision.' **supv0**. The answer was just Yes or No. If ticked Yes, the original question followed. **supv2**. Now, we offered a tick list with numbers ranging from 1 to 10. If ticked the answer '10 or more', a follow-up

question asks the respondent to enter a number. **supv3**. In the dataset, the variable remained the same all the time **supv1**. In all cases, the instruction was that if the numbers tend to vary, the respondent is asked to enter the average.

For investigation segregation by gender at the workplace, a next question asks how many women are supervised to those supervising at least 1 person. **supvwom0**. The answer consists again of a tick list from 1 to 10, and if ticked '10 or more', a follow-up question asks the respondent to enter a number **supvwom2**. The values for number of women supervised are checked against the total number of persons supervised.

## Measuring industry, using NACE industry classification

The questionnaire has a question 'Please select the main business activity of the organisation where you work'. Here, the respondents are offered the two-tier INDUSTRY search tree.

The INDUSTRY search tree is based on the NACE classification of industries, which is firmly standardized by Eurostat and used by almost all national statistical agencies in Europe. We used the Statistical Classification of Economic Activities in the European Community, Rev. 1.1 (2002) (NACE Rev.1.1). In North America, the North American Industry Classification System (NAICS), 2002, is used for the US, Canada and Mexico. A table of concordance between NAICS 2002 and NACE Rev 1.1 is available at the website of the United Nations' Statistics Division. NACE has four hierarchical levels with 33 divisions at the first level and 512 items identified by four-digit numerical codes at the fourth level.

The INDUSTRY search tree uses all NACE codes at the fourth level, except for manufacturing industry, where the third level is used only. This is done for two reasons. First, NACE details the manufacturing industry to a much larger extent than other branches of industry. Second, due to mergers and new technologies, these very detailed fourth level industries do not exist anymore as separate enterprises, making the third level a better approach of reality.

The INDUSTRY search tree is a two-tier search tree with 21 first-tier items and 237 second-tier items. For reasons of user-friendliness, a few NACE codes were split into two items in the INDUSTRY search tree. In addition, the search tree has a number of items that detail the goods or service producing activities of private households to a larger extent than is the case in NACE. The variables related to the INDUSTRY search tree are for the Netherlands **SEC5DIGT**, for Germany **SECTOR\_D**, and for all other countries **SECTOR\_I**. Netherlands has a search tree with more items than other countries, while Germany has one with fewer items.

All data from the search tree is recoded into a variable **NACE4NUM**, specifying industries at the four-digit NACE level, except for manufacturing industry, where the 3-digit level is used. A syntax file is available for recoding these industries into the **NACE4NUM** variable. The variables **NACE3NUM**, **NACE2NUM**, **NACE1NUM** and **NACEFNUM** present aggregated information about the industry on a 3-digit, 2-digit, respectively 1-digit code and on four major industrial categories, notably (1) agriculture, manufacturing, construction, (2) trade, transport, hospitality, (3) commercial services, (4) public sector, health, education.

## Measuring occupations, using ISCO occupation classification

In telephone or face-to-face interviews the respondent listens to the questions and answers. The presence of an interviewer is advantageous to these two modes when it comes to the question 'What is your occupation'. The interviewer can ask for additional information in case the provided occupational title cannot be recoded into a National Occupational Classification (NOC).

In paper or web-based surveys questionnaires, due to the absence of an interviewer the question about occupation is most commonly asked as either an open-ended question or a short aggregated tick list of occupational groups. Both methods have disadvantages. First, coding occupations is a time-consuming activity and in some surveys up to 10 per cent of the cases remain unidentifiable. For the UK, the Netherlands and some other countries, programs for automatically recoding exist, such as CASCOT2000 for the UK. For other countries these recoding programs do not exist or they are not available beyond statistical agencies. These schemes recode job titles and synonym occupational titles into an occupational title used in a National Occupational Classification (NOC). Statistics Netherlands has developed a recoding scheme that holds over 30,000 titles. Second, the answer to the question may present a short aggregated list of occupational categories, asking the respondent to tick the category that comes closest. In these surveys, the data may not be reliable, as respondents may not know how their occupation fits into any of the aggregated categories, introducing *aggregation bias*.

For the question 'Please select your occupation' the *WageIndicator* questionnaire has developed the three-tier OCCUPATION search tree. Since its start in 2001, the search tree has been improved several times. Initially, the search tree had two tiers, facilitating a choice of 25 occupations and an open-ended item. In 2002, the list was extended to 199 occupations, grouped into 12 first-tier items. Since, respondents who cannot find the appropriate title are asked to tick an occupational title that comes closest and to use an open-ended item for specification of the occupational title. By then, the National Occupational Classification of Statistics Netherlands was used. **OCCUPA\_N**

From the open-ended item response, it became clear that respondents to a large extent prefer to specify their occupation, and report that they are a senior or a junior in their occupation, that they supervise a team of job incumbents in the particular occupation, and so on. Therefore, a question was designed to measure the hierarchy within an occupation. After the question about occupation, this question asks whether the respondent wants to qualify the occupation in more detail, such as assistant, senior, junior, trainee, team leader, etcetera. **occtype**

Meanwhile, technical improvements have been made. The search tree technique in the *WageIndicator* questionnaire has two features that increase its user-friendliness. First, it allows the web-visitor to go easily back-and-forth in the search tree. Second, in each tier and each language the list of items is sorted alphabetically, allowing for an easy search.

In 2004, when the web survey was launched in countries outside the Netherlands, it turned out that the Netherlands NOC was not usable in a search tree for an international web survey, because countries wanted to refer to their own national classifications. The four-digit occupations of the International Classification of Occupations (ISCO) of the ILO could have been used for the search tree, but holds too few occupations (390) to meet the preferences of the respondents. Longer lists of occupations were required, though of course coded within the framework of ISCO.

Due to the absence of an adequate list of occupations and due to lack of funds for translating such a list, the countries could add their own occupations to the occupation search tree, as long as they had a code placing them in the international ISCO classification of occupations. This led to a list of more than 5,000 occupational titles, but many of them are available in only one or two languages. These occupations were either not at all or not to a very detailed level found in the countries, or translations were not available. By the end of 2005, most countries participating in *WageIndicator* had translated between 700 and 1500 occupations of the list of more than 5,000. The variables is called **OCCUPA\_I**. Germany and Denmark have a search tree different from the other countries. These variables are called **OCCUPA\_D**, **OCCUP\_DK**.

For 2007, the occupations database, used in the survey, was updated. This could be done due to the so-called EUROCCUPATIONS project (05/2006-04/2009), that aims at developing a detailed 8-country occupations database for comparative socio-economic research in the European Union, funded from the 6<sup>th</sup> Framework Program of the European Commission (FP6-028987). This EUROCCUPATIONS project helps to improve the international comparability of occupations classifications. Altogether, approximately 1500 occupations are listed and subsequently translated into the languages of the survey. The OCCUPATION search tree has therefore, compared to ISCO, far more detailed occupations, making it a better approach of reality for the survey respondents.

For 2007, the Netherlands occupation search tree was improved. Once the three-tier industry search tree is completed, a fourth screen presents a list of occupations that is most appropriate for the ticked industry, including an item 'Other occupation'. Once ticked this item, the respondents enters into the three-tier occupation search tree. It is assumed that the majority of the respondents will find their occupation in this fourth screen. The major advantage of this industry-based occupation list is the reduction of clicks from six to four screens and the reduction of the number of words to be read before being able to click the appropriate occupation for the majority of the respondents.

In the dataset, the variables **OCCUP\_DK** (Denmark), **OCCUPA\_D** (Germany), and **OCCUPA\_I** (International) are summarized in one variable, **ISCO9NUM**, using a 9-digit code, whereby the first 4 digits specify the ISCO-code and the last five digits is a follow-up number. The variable ISCO4NUM holds the first four digits, and therefore corresponds with the ISCO88 classification. The data of **OCCUPA\_N** (Netherlands) are only recoded into **ISCO4NUM**. The variables **ISCO3NUM**, **ISCO 2NUM**, and **ISCO1NUM** present aggregated information about the occupation on a 3-digit, 2-digit, respectively 1-digit code.

## Measuring migration and ethnic background

When studying labor market outcomes, both migration and ethnic background are important variables. The two variables are in some countries intertwined, but in others they are not. Worldwide, the definitions of migration and ethnic groups vary, and therefore are measured differently in surveys. Migration may apply to migration from another country as well as migration within the country, particularly in large countries such as India or China. Migration may be only important to the first generation, but sometimes also to the second generation. Related to migration is citizenship. In other countries migration is not of primary interest but ethnic background or language spoken at home are better variables to be used in studies

on labor market outcomes. The *WageIndicator* survey has quite some questions asking for migration or ethnic background.<sup>1</sup>

The questions about country of origin follow strictly the principle of the minimum average number of words and clicks. The question asks if the respondent is born in the country of survey Y/N **cobself**. If not, does he/she come from the most common immigrant countries, using a tick list, including the answer 'Other'? If 'Other' is ticked, the COUNTRY search tree pops up. The next question asks if 'your mother is born in the same country as you' Y/N **cobmothe**. If not, does she come from the most common immigrant countries, providing a tick-list, including the answer 'Other'? If 'Other', the COUNTRY search tree pops up. A similar series of questions is asked for the father Y/N **cobfathe**. Thus, the large majority of the respondents answers only three Y/N questions, and a minor part is asked a fourth question about the most common immigration country. A very small fraction is confronted with the three questions and all follow-up questions, notably in the very exceptional case that the person and his/her parents are all born in different countries, none of which belong to the most common immigrant countries. If the respondent is not born in the country of survey, in some countries questions are asked about the year of entering the country **yyarriva** as well as the reason for coming to the country **cobreaso**. Four reasons are optional, notably family reasons, reasons of work, as a refugee or other reasons.

The COUNTRY search tree is a 2-tier tree with 236 countries, subdivided into 5 continents. For the dataset, the codes of the worldwide used classification of the International Organisation for Standardisation were used (ISO, [www.iso.org](http://www.iso.org)).

In large countries, migration within the country may be important for labor market outcomes. Therefore, a sequence of questions asks for the home region, the region of birth and sometimes the region of the workplace. Here too the principle of minimized words is followed. The sequence starts to ask 'In which region do you live' **REGIHOME**. Here the REGION search tree is presented, as discussed in the previous section. In two countries, United States and Netherlands, the postal code is asked instead. For the dataset, the postal codes are recoded into regions. If not indicated that the respondent is born in another country, a next question asks if one is born in the region of living. If not, a follow-up question asks for the region of birth **REGIBIRT**, in India followed by a question about the reasons for moving, the means of moving and the time it took to find a job in the region of living. Finally, a question asks for the region of work **REGIWORK**, if different from the region of living. This question also aims to identify whether the respondent is a mobile worker, or is working from home.

In our attempt to measure ethnic background and migration as precise as possible, information is asked about the language the respondent mostly speaks at home **coblangu**. The answer consists of a country-specific limited list of most commonly spoken languages in the country of survey, including the native language(s), the item 'other' and if applicable an item 'local dialect'.

In some countries, the country of birth is not important in identifying ethnic groups, because ethnic groups are in the country for more than two generations. This applies for example to the UK and US. In these countries, a question is asked 'To which ethnic group do you belong?', followed by a radio-button list of the most common ethnic groups **cobethgr**. In other countries, it is more polite to let respondents self-

---

<sup>1</sup> Information about ethnic groups and languages spoken at home can be found in the World Fact Book of the CIA, see [www.cia.gov/cia/publications/factbook/index.html](http://www.cia.gov/cia/publications/factbook/index.html)

define whether they belong to a minority group, for example in Hungary **cobetmin**. Once ticked yes, a follow up question aims to identify the ethnic minority group **cobetmi1**.

In other countries, such as India for example, it is also important whether the respondent is a citizen **cobindia**. Therefore, a question is asked 'Are you an Indian citizen?' (Y/N).

## Measuring region, the NUTS regional classification

In the survey several questions address domiciliary region, region of birth and region of work. For European countries, the regions according to the NUTS regional classification are used. See <http://europa.eu.int/comm/eurostat/ramon/>. A worldwide classification is not available. For countries outside Europe, various solutions were applied. For South Africa, provinces are used, for India States and Territories, and for Brazil States and large cities within States.

The REGION search tree has a 1-tier, 2-tier or 3-tier choice, depending of the detail in the NUTS classification or the choices made for countries outside Europe. For tracing boundary work and boundary residence, the names of the neighboring countries have been added to the REGION search tree, wherever applicable.

In the dataset, the variables **REGIHOME**, **REGIBIRT** and **REGIWORK** are composed of the 3-digit numerical ISO country code, a follow-up number and the numerical part of the NUTS code. For example, the UK sub-region Brighton and Hove with the NUTS code UKJ21 has the *WageIndicator* code 8261021. The value label includes the two-letter ISO country code and the NUTS region code, in this case 'UKJ21 Brighton and Hove'. In Mexico, the code 48118000 is used to identify the region Nayarit, whereby 418 is the ISO code for Mexico and 18000 a follow-up code.

In the USA and in the Netherlands, the domiciliary region is asked by postal code. In the USA, the REGION search tree is only used in case the respondent doesn't know the postal code or doesn't want to say. In the dataset, the postal codes are assigned regions. For privacy reasons, the postal code in the Netherlands is truncated into a 2-digit code, called **POSTCOD2**. The postal code in the USA is a 5-digit code, called **POSTCUSA**.

**Download Region search tree.**

**For the Netherlands, a scheme is available to recode postal codes into regions.**

## Measuring household composition

Household composition is measured in several ways. First, a question asks with whom they live in your household, providing a checkbox with items such as partner **hhpartn1**, one or more children **hhchild**, grandchildren **hhgchild**, parents (in law) **hhstate7**, brothers or sisters **hhstat21**, non-family members **hhstate8**, or house helps **hhstat12**.

Second, marital status is asked, indicating whether the respondent is married, never married, widowed, or divorced **hhstat7**. Another question asks whether the respondent has children, including step and foster children **child**. If ticked yes, it is asked how many children are living at home **childhome** and how many children are not living at home **childout**. If the respondent has children up to the age of 12 year, a set of items asks where the children are when the respondent is at work.

A few questions ask how household tasks divided up in the respondent's home. These questions are partly similar to the ones asked in the Dublin working conditions survey.

### The time frame of events and perceptions

In work and employment surveys, it is tempting to ask for changes over time or for expectations of future events. Yet, when for example asking for changes at the workplace over a time period of five years, it is likely that almost 10 percent of the respondents will not have a work experience longer than five years, and therefore this group will not be able to provide a reliable answer. For all questions asking about the past or the future, *WageIndicator* has therefore for two reasons a strict policy to ask these questions in a time frame of the past year or the next year. First, the past and the future time frame are of an equal length. Second, data within a year's time frame are probably reliable data. Longer periods increase the likelihood that either the respondents' memory in case of the past or their imagination in case of the future becomes blurred. Moreover, it increases the likelihood that the memory or expectations are intertwined with other events, such as entering the labor market or retiring.