

# 9 Clustering households into household groups

---

In the fourth step, households can be clustered into household groups. This chapter describes how this can be done for the main household groupings targeted in the work.

---

## 9.1. Introduction

In the fourth step, households can be clustered into household groups. This may be done on the basis of equivalised disposable income, but the clustering can also be based on alternative characteristics such as main source of income, household type or age of the head of the household. It will depend on the available information and on the quality of the distributional results what household groups can be targeted.

This chapter describes how distributional results can be derived for the various household groups. For breakdowns based on socio-demographic characteristics, this will be relatively straightforward, by simply allocating households on the basis of their underlying characteristics. This is explained in Section 9.4. However, for classification according to standard of living (i.e. equivalised disposable income) or main source of income, more guidance may be needed. Section 9.2 describes the classification according to standard of living and Section 9.3 according to main source of income.

## 9.2. Clustering according to standard of living

In the classification according to standard of living, households are clustered on the basis of their equivalised disposable income. For this purpose, household disposable income in line with national accounts totals (i.e. after imputation for missing elements and alignment of the micro data to the macro results) is recalculated into per consumption unit results, taking into account the size and composition of the household, to arrive at comparable results across households. This is known as *equivalised disposable income*. The Handbook uses the OECD-modified equivalence scale as reference method, but compilers may also decide to apply a different scale if this is deemed more appropriate in relation to country specific circumstances. The OECD-modified scale assigns a value of 1 consumption unit to the first adult in the household, a value of 0.5 for each additional person aged 14 and over, and 0.3 for all children under 14 (see also Box 2.1).

After disposable income for each household has been divided by its number of consumption units, households can be ranked on the basis of this equivalised disposable income. On the basis of this ranking, households can then be clustered into income groups, for example into income deciles. In that case, the clustering should be done in such a way that each decile represents 10% of the households. Hence, each decile represents 10% of the total number of households (not consumption units). Depending on the reliability of the distributional results, more detailed breakdowns may be envisaged as well, such as into income percentiles, which may be particularly relevant for the top end of the distribution. However, this would require a careful assessment of the robustness of the results at these levels of detail (see also Chapter 12).

## 9.3. Clustering according to main source of income

Households can also be clustered into household groups on the basis of their main source of income. For that purpose, households should be clustered in the category which shows the highest contribution to their adjusted disposable income. It should be borne in mind that this should be based on the adjusted disposable income in line with national accounts, thus after imputation for missing elements and alignment of the micro data to the national accounts totals.

The DNA work distinguishes four main sources of income, namely a) *wages and salaries* (based on item D11R), b) *income from self-employment* (based on item B3R3), c) *net property income* (based on item D4N), and d) *current transfers received* (based on items D62R (social benefits in cash received), D63R (social benefits in kind received) and D7R (other current transfers received)).<sup>1</sup> As explained above,

households should be ranked in the category which shows the highest contribution to its national accounts aligned adjusted disposable income.

#### 9.4. Clustering according to socio-demographic information

Households can also be clustered on the basis of socio-demographic characteristics. One specific classification distinguished in the DNA work is according to household type. This takes into account the presence, number and age of the members of the household. In the DNA approach, eight categories of household types are distinguished, i.e. a) *single less than 65 years old*, b) *single 65 and older*, c) *single with children living at home*, d) *two adults less than 65 without children living at home*, e) *two adults at least one 65 or older without children living at home*, f) *two adults with less than 3 children living at home*, g) *two adults with at least 3 children living at home*, and h) *others*. In this classification, an adult is defined as anyone 18 years or older.<sup>2</sup> Furthermore, the delineation of “children living at home” is based on all individuals up until the age of 16 plus the individuals whose age is between 17 and 24 and are offspring of one of the household members and are still living at home. Depending on user needs and the quality and available detail from the underlying data, more granular breakdowns can be envisaged as well.

With regard to socio-demographic information, several alternative classifications can be envisaged. This will depend on user needs, availability of the necessary information to cluster the households accordingly, and the quality of the underlying distributional results. Examples of breakdowns that could be envisaged are according to the *age of the head of the household* (e.g. into a) 0-24, b) 25-34, c) 35-44, d) 45-54, e) 55-64, and f) 65 and above),<sup>3</sup> *housing status* (e.g. a) rental, b) owner-occupied with mortgage, and c) owner-occupied without mortgage), or *main activity of the head of the household* (e.g. a) unemployed, b) employee, c) employer, d) own-account worker, e) unpaid family worker, f) member of producer’s cooperative, g) student, h) retired, and i) not classifiable (see Section 2.6.5 for more information)). Countries apply different rules to determine the head of the household, but most of them define it as the person with the highest income (see also (United Nations Economic Commission for Europe, 2011<sup>[1]</sup>) and (OECD, 2013<sup>[2]</sup>)). In assessing these alternative breakdowns, compilers should check the availability of the relevant information needed to cluster the information on the basis of these specific characteristics. However, they should also carefully assess the robustness of the results at these levels of detail (see also Chapter 12).

The clustering according to socio-demographic information is relatively straightforward in the sense that households should simply be allocated on the basis of their underlying characteristics. In case of changes throughout the year, compilers should look at the duration of the different situations and allocate households according to the situation it was in for the longest period of time.

## References

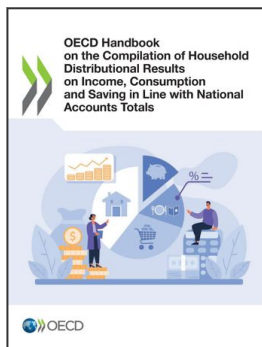
- OECD (2013), *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264194830-en>. [2]
- United Nations Economic Commission for Europe (2011), *Canberra Group Handbook on Household Income Statistics*, [https://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canberra\\_Handbook\\_2011\\_WEB.pdf](https://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canberra_Handbook_2011_WEB.pdf) (accessed on 27 September 2017). [1]

## Notes

<sup>1</sup> As mentioned in Section 2.5.2, the latter category could be further broken down into pension benefits received and other current transfers received, in case the relevant information is available at that level of detail.

<sup>2</sup> In line with general principles of the System of National Accounts, the age of a person for a given reference year should be derived on the basis of his/her age during the largest part of the year. This means that anyone born after the 1st of July should be assigned its age at the start of the year, whereas anyone born on or before the 1st of July should be assigned its age at the end of the year. If this is not feasible, it could be decided to take one cut-off point in the year (e.g. at the start or at the end of the reference period), bearing in mind that this may generate slightly different results.

<sup>3</sup> For national purposes, it may also be of interest to delineate the last two groups on the basis of the retirement age in the country. However, for international comparability, it is recommended to maintain the breakdowns as suggested here. Furthermore, in using the retirement age, it has to be borne in mind that time series analysis may be affected, when the retirement age is changing over time.



From:

## OECD Handbook on the Compilation of Household Distributional Results on Income, Consumption and Saving in Line with National Accounts Totals

Access the complete publication at:

<https://doi.org/10.1787/5a3b9119-en>

### Please cite this chapter as:

OECD (2024), "Clustering households into household groups", in *OECD Handbook on the Compilation of Household Distributional Results on Income, Consumption and Saving in Line with National Accounts Totals*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/0b8c5a09-en>

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.