DISTRIBUTIONAL INFORMATION ON HOUSEHOLD INCOME, CONSUMPTION AND SAVING IN LINE WITH NATIONAL ACCOUNTS

GUIDELINES

Version - December 2020

Prepared by the Statistics and Data Directorate of the OECD

Table of Contents

INTRODUCTION

1. The present guidelines are intended to assist national compilers in the implementation of distributional estimates of household income, consumption and saving compatible with the System of National Accounts (SNA).

2. While the aggregate nature of the SNA provides a comprehensive, consistent and flexible set of macroeconomic accounts for policymaking, analysis and research purposes, it prevents capturing the disparities in income, consumption and saving across different household groups. However, measures of the distribution of income, patterns of consumption, redistribution through tax and transfer systems provide information that is critical to the design of economic and social policies. Although micro data (e.g. surveys, administrative records and censuses) could be used to provide measures of inequalities across individuals, results might diverge from SNA macro aggregates, which can cause problems for users.

3. In 2011, the OECD and Eurostat launched a joint Expert Group, which aimed to explore whether it is possible to devise an internationally comparable methodology to produce measures of disparities that are consistent with national accounts concepts and totals, using existing micro data sources. In 2014, this work was followed up by an OECD Expert Group on Disparities in National Accounts (EG DNA) and in 2018 by a joint OECD-Eurostat Expert Group[1]. The material presented in these guidelines is based on the work done by these Expert Groups.

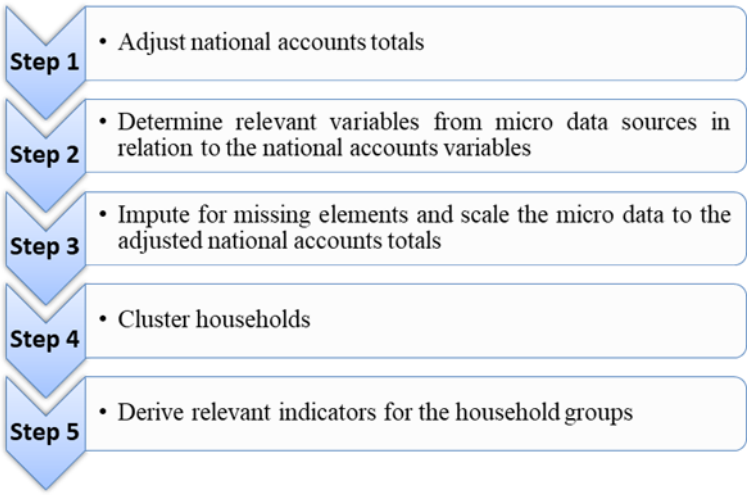
4. The guidelines are organised as follows: the first part gives an overview of the step-by-step procedure for the estimation of the distributional information consistent with the national accounts aggregates, while the second part analyses in depth some methodological aspects. The third section gives an overview of the questionnaire template for the collection of data and metadata, and some instructions for its completion.

[1] The Expert Groups have engaged in three exercises. The results of the first two exercises have been published in Fesseau and Mattonetti (2013) and Zwijnenburg, Bournot and Giovannelli (2017) respectively. Results of the third exercise will be published in Zwijnenburg, Bournot, Grahn and Guidetti (2021, forthcoming).

PART I: A step by step procedure for the estimation of the distributional information

5. This section proposes a step-by-step procedure for the estimation of distributional information of households' income, consumption and saving in line with national accounts aggregates. The procedure is set up in 5 steps, starting with adjustments to the national accounts' totals that are required to remove any amounts that do not relate to resident private households. This is followed by determining relevant items in micro data sources in relation to the variables from national accounts into which income and consumption need to be broken down. In the third step, imputations are made in the case of missing elements in the micro data and subsequently these micro data are scaled to the adjusted national accounts totals. On the basis of these benchmarked micro data, in step 4, households can be clustered into groups according to their equivalized quintile or according to other household characteristics. In the final step, the relevant indicators for the underlying household groups can be derived. Figure 1 provides an overview of these various steps.

Figure 1. A step-by-step approach for the estimation of distributional information



Step 1: Adjustment of the national accounts totals

6. Distributional information on income, consumption and saving concerns private households only. Therefore, national accounts totals may need to be adjusted to exclude non-profit institutions serving households (NPISHs) should the latter be reported together with the former. Further amendments might be required to exclude the expenditures of non-resident households on the national territory at the detailed expenditure category. Finally, national accounts totals should be adjusted to exclude the income and the consumption by the part of the population that does not relate to private households, such as institutional households. The following paragraphs provide a brief overview of these adjustments.

- *Exclusion of NPISHs*

7. Households and NPISHs are sometimes grouped together in the institutional sector accounts. The aggregation could be due to the lack of specific information on NPISH activities but can also be based on the notion that, as these institutions are largely financed by households and their purpose is to serve households, their accounts can be assimilated to those of households. For the purpose of this study, however, the Secretariat recommends excluding NPISHs as they do not concern the target population of the exercise. Such an adjustment should be made for all income and consumption components, even if this implies the use of very rough estimates. In the absence of information on the main variables for

NPISHs, experts are encouraged to apply assumptions. Countries should be aware that the share of NPISHs may vary significantly across components, so applying identical percentages to all components should be avoided.

- *Expenditures of non-resident households on the territory and of resident households abroad*

8. The national accounts aggregate of total final consumption expenditure should include the expenditures that resident households make abroad while excluding the consumption of goods and services made by non-resident households on the economic territory of the country. It depends on the setup of supply-and-use tables in countries whether this is already provided for at the detailed level of consumption components in the national accounts or whether these 'corrections' are only applied at the aggregated level (i.e. items P33 'Final consumption expenditure of resident households abroad' and P34 'Final consumption expenditure of non-resident households on the territory'). Furthermore, it will depend on the coverage of the underlying micro data whether adjustments of the national account totals are needed at the detailed level to arrive at better alignment between the micro and macro data.

Expenditures of non-resident households on the territory

9. If the detailed consumption items from the national accounts totals still contain expenditures of non-resident households on the territory, it is recommended to subtract these expenditures at this detailed level (instead of at the aggregated level) before applying the distributional information. This will lead to the best alignment with the micro data as these will only relate to expenditures of resident households. Alternative would be to correct at an aggregated level (P34), but as the amount of consumption expenditure by non-residents may vary significantly across consumption items, it is expected that this will lead to sub-optimal results in aligning micro and macro data in step 3.

10. Countries' experiences suggest that information from tourism satellite accounts, balance of payments, surveys on foreign tourism, and credit card information could be used to adjust the various consumption categories at the detailed level. Countries may also use the breakdown of expenditure of residents abroad (in case detailed information is available) as a proxy, or make estimates for the specific categories which this spending will most likely relate to (mainly recreation and culture, restaurants and hotels, and transport).

Expenditures of resident households abroad

11. Consumption expenditure of resident households abroad may be included at the detailed level in the national accounts or may only be available at the aggregated level (P33). In the latter case, it will depend on the type of information that is available from micro data sources whether there is a need to make an adjustment at the detailed level. If information on expenditures of residents abroad is included at the detailed level of consumption in the micro data source, it is recommended to adjust the national accounts totals accordingly in order to align them with the micro data source. However, if - as is often the case in countries[2] - information on consumption expenditure abroad is only available at an aggregated level in the micro data sources, such a correction is not necessary. These expenditures can in that case better be included at the aggregated level (P33) as this would provide the best linkage between the micro and the macro data.

[2] Paragraph 9.80 of the 2008 SNA explains that in calculating household final consumption expenditure "*it may be convenient to calculate the total expenditure made by all households, whether resident or not, within the economic territory and to adjust this figure by adding expenditures by residents abroad and subtracting expenditures by non-residents within the economy territory*". The rationale is that it may be easier to balance supply and use of goods and services at the detailed level if the total of domestic consumption expenditure (i.e. expenditure by residents and non-residents) is considered.

- *Population not covered by micro sources*

12. In most countries, a part of the resident population falls outside the scope of micro data surveys, while national accounts typically cover the entire population whose center of economic interest lies in the country. Usually, the population falling outside the sampling frame of micro sources includes people without a permanent address, those living in non-private dwellings (such as prisons, boarding schools, retirement homes, hospitals, nursing homes, religious institutions and hotels) as well as people living in overseas territories or in sparsely populated areas. As the composition and behaviour of people living in non-private dwellings may be completely different from private households, the Secretariat recommends separately identifying and estimating the transactions of these households in a specific household sub-group, i.e. to deduct from national accounts data (component by component) the flows related to non-private households and to show the relevant totals separately[3]. This is not the case for the other groups that may be missing from micro sources as these still concern private households. Countries are encouraged to make imputations for these groups to include them in the distributional results.

13. As is the case with excluding transactions of NPISHs, countries should avoid applying identical percentages to all components in adjusting for non-private households, as this group may have more substantial shares in some specific income and consumption components (e.g. social protection spending and Social Transfers in Kind related to health for people living in retirement homes). Furthermore, it must be stressed that this step deals with population scope adjustments only. It does not concern any deletion of specific variables that may be missing in micro data sources.

Step 2: Determine relevant variables from micro sources in relation to national accounts variables

14. The second step of the proposed procedure is to determine the relevant components from the micro data sources in relation to the national accounts variables for income and consumption. The relevant variables from national accounts are presented in Annex 1. For a couple of transactions more detail is presented, as sometimes variables from the micro data sources need to be linked to specific national accounts components at a more detailed level. In addition, some specific sub-items may not be directly available from micro data sources (such as investment income attributed to insurance policyholders or employers' imputed social contributions). The template includes these extra details as this may lead to better linking of micro data to the underlying national accounts totals and to specific imputations for only those variables that cannot be linked to micro data sources, and therefore to more accurate results.

15. All components from the micro data sources for the most recent year available (if information referring to the same year is also available for the annual national accounts data) that best correspond to the national accounts variables should be identified. A single micro source may be used for the various income and consumption components, but sometimes multiple data sources may be available. Whenever multiple micro data sources are used, it is recommended that the various data sources refer to the same year and that the data on the various components describing the same households (or household types) are linked as good as possible. The best way to do the latter would be via micro data linking (e.g. by linking the data on a micro economic level on the basis of unique identifying information such as social security number). Otherwise, assumptions will have to be made how data from the various data sources can be combined to best describe various household types, making sure that the most reliable data on full sets of accounts of various household types are derived. In that case, socio-demographic characteristics that best correspond to certain household types may be used to link the data. This will especially become an issue when information on consumption is derived from other micro data sources

[3] By presenting this information as separate household group in addition to the distributional results, all information would add up to national accounts totals for the household sector as a whole.

than information on income, and no good denominator can be found to link the information at a micro level. More information on data linking can be found in Annex 2.

16. Moreover, attention should be paid to the fact that micro data sources may follow different classifications than the SNA. It may indeed occur that the same transactions are classified differently. For instance, wages and salaries paid while on sick or maternity leave may be recorded as wages and salaries in micro sources while they are classified as social benefits in the SNA. Also, the income received by a sleeping or silent partner participating in an unincorporated enterprise is typically considered as property income by micro sources but as mixed income in national accounts. Under those circumstances, the micro components should be reallocated to match the SNA classification.

Step 3: Impute and scale the micro data to the adjusted national accounts totals

- *Imputation for missing items*

17. Some national accounts transactions have no counterpart in the micro data, due to either their conceptual nature or practical considerations. For instance, imputed transactions such as Employers' imputed social contributions (D122 and D612) and Financial Intermediation Services Indirectly Measured (FISIM) have no direct match in the micro data. The same occurs for Social transfers in kind (D63), Investment income attributable to insurance policyholders (D441) and Investment income payable on pension entitlements (D442). In addition, some economic activities are deliberately concealed to avoid tax payments (underground production) and therefore income needs to be estimated indirectly. Also, specific household groups, such as people living in overseas territories or in sparsely populated areas may be missing from the micro data. In those cases, imputations to micro data sources may be required.

18. Should imputations be required due to lack of reliable data, three methods can be employed, all making use of indirect information (corresponding to method B – D in the template). The first method (method B in the template) proxies the missing information by using the distribution of another component, assuming that the two are distributed in a similar way. The distribution for Employers' imputed social contributions (D122) may for example be derived on the basis of the distribution for Wages and salaries (D11), whereas the distribution of FISIM may be linked to Interest paid (D41P') and Interest received (D41R') (see also Part II where FISIM is described in more detail). The second method (method C in the template) imputes missing distributional information according to exogenous data (e.g. socio-demographic information used for the distribution of Social Transfers in Kind) available at the individual or at the household level. In both cases, the assumptions made to allocate the totals can impact the overall final distribution. It is therefore very relevant how these imputations are linked to various (groups of) individuals or households. It is preferred to employ the imputations as detailed as possible as it enables to classify households into different groupings in the remainder of the process. When imputations are made at group level, this may need to be done for the various classifications that are needed. If no information is available, a third method can be used (method D in the template) where the imputations are made in a manner that the inclusion or exclusion of the component does not have an impact on the commonly used distributional indicators. Therefore, this can only be done at the end of the exercise when the distributional information has been derived on the basis of the other variables. In applying this solution, it has to be decided to which balancing item to best link the specific item. For consumption items, it makes sense to link it to either Final domestic consumption expenditure (P31DC) or Final national consumption expenditure (P31NC). For income components, the distributions may be linked to the Balance of primary incomes (B5), Disposable income (B6) or Adjusted disposable income (B7). It will depend on the underlying item which aggregate will provide the best proxy.

- *Scale the micro data to the adjusted NA totals - Bridging micro-macro gaps*

19. After the necessary imputations have been made, the micro data (including the imputations) should be scaled to the adjusted national accounts totals. For that purpose, it is advised to explore the possible or most likely reasons for gaps between the micro and macro aggregates (step 1), especially in case of large gaps between the two, and to allocate the gaps to the relevant households or household groups on the basis of these underlying reasons (step 2). A framework has been developed that may be applied for that purpose. Table 1 presents an overview of this framework.

Table 1: Framework for allocating micro-macro gaps to households or household groups

Step 1: A*ttributing micro-macro gaps to underlying causes*

	Item xx.	**Original estimate**	**Correction**	**Ultimate Estimate**
I	**National account total (A)**	…	…	…
	- Adjustment for NPISH (B1)	…	…	…
	- Adjustment for non-private households (B2)	…	…	…
	- Adjustment for expenditures of non-resident households on the territory (B3)	…	…	…
	= Adjusted NA total (C=A-B1-B2-B3)	…	…	…
II	**Micro total (D)**	…		…
	= Macro-Micro gap (E=C-D)	…	…	…
III	**Conceptual or classification issues (F)**	…	…	…
	Underground and illegal activities (G)	…	…	…
	Other elements missing in micro data (H)	…	…	…
	Estimation errors (under-/overcoverage) (I)			…
	Measurement errors (under-/overreporting) (J)			…
	Reasons n.e.c. (K)			…
	= Remaining gap (L=E-F-G-H-I-J-K)			…

Step 2: Allocating the gaps to the quintiles

	Item xx.	**Estimate**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
IV	***Allocation on the basis of micro data***						
	Original micro aggregate (P)	…	…	…	…	…	…
	Revised micro aggregate (Q)	…	…	…	…	…	…
V	***Allocation on the basis of meso corrections***						
	Conceptual or classification issues (R)	…	…	…	…	…	…
	Underground and illegal activities (S)	…	…	…	…	…	…
	Other elements missing in micro data (T)	…	…	…	…	…	…
	Estimation errors (under-/overcoverage) (U)	…	…	…	…	…	…
	Measurement errors (under-/overreporting) (V)	…	…	…	…	…	…
	Reasons n.e.c. (W)	…	…	…	…	…	…
VI	**Alignment of remaining gap (X=C-Q-R-S-T-U-V-W)**	…	…	…	…	…	…
	Final estimate (Y=Q+R+S+T+U+V+W+X)	…	…	…	…	…	…

20. In the first step, gaps between micro and macro aggregates have to be allocated to the underlying reason. As the gap can be due to a quality issue with the adjusted national accounts total, Block I first focuses on the derivation of this amount, starting from the national accounts total and showing the adjustments for a) NPISH, b) non-private households and c) expenditures of non-resident households on the territory respectively. The first column in this block presents the original estimates used to derive the adjusted national accounts figure. The second column provides the possibility to

correct any of these original figures to close part of the gap between the micro and the macro results. The result is presented in the third column.

21. The second block of the framework (Block II) confronts the adjusted national accounts total with the micro aggregate, showing the gap between the two. The initial micro-macro gap is presented in the first column. The third column shows the gap that still remains after corrections have been made to the adjusted national accounts aggregate. This remaining gap still needs to be attributed to other reasons. This is done in the third block (Block III). Note that in this respect, no corrections should be reported in Block II for changes to the micro aggregate, as all changes should ideally be attributed to one of the underlying reasons listed in Block III. This block lists possible causes related to a) conceptual or classification differences, b) missing information related to underground and illegal activities, c) other elements that may be missing from the micro data, or errors with regard to the micro data, broken down into d) estimation errors (i.e. related to under- or overcoverage) and e) measurement errors (related to under- and overreporting)[4]. It also contains an item for f) reasons that are not covered by the other categories. The block ends with the gap that still remains after attributing parts of the gap to the underlying reasons. Ideally, the amount of this remaining gap is zero, so that the complete gap is explained by the various causes.

22. After the attribution of the micro-macro gaps to the underlying causes, the related amounts have to be allocated to the relevant household groups. This is dealt with in the second step. Block IV focuses on the allocation on the basis of revised micro data, which is the preferred option. In that case, corrections are processed at the micro level and new results are derived following the standard step-by-step approach. However, in some cases this may be deemed too time-consuming or too complex. In those cases, corrections may be allocated at an aggregated level. This can be done in Block V which provides the opportunity to allocate the remaining gaps on the level of the quintiles (or other household groups). Finally, Block VI deals with allocating the remaining gap that could not be linked to any of the possible causes. In most cases, this step may consist of adjusting the amount reported by each individual or household by the same proportion (this corresponds to method A in the template). The underlying assumption is that the distributional information from the micro source provides an adequate representation of the underlying distribution. The sum of the corrected micro data and the consecutive meso-corrections leads to the distributional results for the quintiles (or other household groups).

23. Countries are encouraged to apply the framework for at least those items that show large gaps between the micro and the macro aggregates. Simply applying a proportional allocation on the basis of the underlying micro data is assumed to lead to sub-optimal results for those items. On the other hand, a proportional allocation (i.e. applying an adjustment coefficient (macro total/micro total) to all households[5]) is deemed acceptable for items with smaller gaps (or for dealing with small remaining gaps as was explained above), especially when no information is available on the reasons for the gaps.

Step 4: Clustering households on the basis of the benchmarked micro data

24. When all steps that are described above have been processed at the micro level, in the next step, on the basis of the underlying micro data, the households can be clustered into household groups. As a minimum, the Secretariat would like to receive the information broken down by income quintile, on the basis of *equivalized disposable income* according to 2008 SNA definitions, as this is the main focus of the exercise. When countries are also capable of compiling distributional results by household

[4] More detailed information can be found in Annex 3.

[5] In case of negative values, however, the scaling procedure using a proportional adjustment may lead to distorting results. In that case it is advised to leave the negative values untouched and only scale up the positive values to the adjusted national accounts total. The adjustment coefficient should in that case be based only on the households reporting positive values.

type and/or main source of income, they are kindly requested to do so, as this may also provide very valuable information. However, the compilation of this additional distributional information is only optional.

25. For the breakdown into income quintiles, households should be clustered on the basis of *equivalized* household disposable income (as presented in Annex 1). This means that the disposable income estimates for the households should first be adjusted to take into account differences in size and composition of households, recalculating results according to the number of consumption units in each household, before allocating them to the relevant income quintile. For the purpose of this study, the Oxford-modified equivalence scale has been chosen as reference method. Accordingly, the first adult counts as 1 consumption unit, any additional person aged 14 and over counts as 0.5 per person, while all children under 14 count as 0.3 per child. Household income and consumption are then divided by the sum of consumption units to obtain a comparable measure across households. The equivalized disposable income for each household is estimated by dividing its disposable income (from step 3) by its number of consumption units. After this step, households can be ranked according to their equivalized disposable income and allocated to income quintiles (Q1, Q2, Q3, Q4 and Q5), in such a way that each quintile represents 20% of the households (with sampling weights applied)[6].

26. As was explained before, optionally countries can also provide the Secretariat with distributional data according to household type and/or main source of income. In the case of household types eight categories are identified, looking at the composition of the household and the age of the adult(s) of which the household comprises. In the case of main source of income, four categories are identified: wages and salaries, income from self-employment, net property income, and current transfers received.

Step 5: Derive relevant indicators for the household groups

27. When all the micro data have been aligned with the adjusted national accounts totals and households have been clustered into household groups, the totals for the various groups can be calculated and relevant indicators can be derived.

[6] The classification is done on the basis of disposable income per consumption unit, but the distribution is done on the basis of the number of households. This means that in the classification each quintile represents 20% of the total number of households (and not consumption units).

PART II: Methodological aspects

28. Some methodological issues may be encountered during the compilation process. For instance, some national accounts aggregates have no direct counterpart in the micro data. This could be due to conceptual differences (e.g. imputation for owner-occupied housing or FISIM) or practical reasons (e.g. the non-observed economy). Furthermore, there are some relations between transactions within the system that must be borne in mind when deriving the various distributions. This section of the guidelines provides some explanations and guidance on how to deal with these methodological issues.

Operating surplus (B2)

29. Imputed rent enters in both household disposable income and household consumption expenditure. Persons who own the dwelling in which they live are treated as producers of housing services for their own consumption. The production generates income that is recorded as operating surplus (B2). The operating surplus corresponds to the surplus accruing from the process of production after deducting intermediate consumption (maintenance and repairs, borrower financial services (i.e. FISIM[7])) and taxes less subsidies on production. The housing services produced are deemed to be equal in value to the rent that would be paid on the market for accommodation of the same size, quality and type. This imputed amount called "imputed rent" should be used to value both the related output and the related consumption expenditure (CP042). In the exercise, it has been decided to focus on gross instead of net operating surplus (the latter refers to operating surplus net of depreciation), mainly due to reasons of data availability in most of the countries.

Consequence for the EGDNA work

30. When allocating operating surplus to household groups, each of the underlying items should be distinguished separately (imputed rent, maintenance and repairs, FISIM, taxes on production) and linked to the most appropriate micro information. For imputed rent, the same distribution by household group should be applied to the relevant income (B2R1) and consumption component (CP042).

Mixed income (including adjustment for non-observed economy) (B3)

31. Mixed income is the item in the national accounts that most closely corresponds to income from self-employment. Mixed income measures the surplus or deficit accruing from production by unincorporated enterprises owned by households after deducting compensation of employees, taxes on production and intermediate consumption (including FISIM[8]), but before deducting any interest charges, rent or other property incomes payable on financial assets, land or other natural resources required to carry out the production. This also concerns mixed income on the basis of underground production (see below) and income from production for own consumption. As was the case with operating surplus, it has been decided to use gross figures instead of net figures, mainly because of data availability in most countries.

Delineation issue

[7] Please note the link with the FISIM correction that is made in the property income block. The part that relates to mortgage loans should be equal, i.e. the part included in the FISIM correction in D41P should equal the deduction as intermediate consumption in Operating surplus. For further explanation, see the section on FISIM.

[8] For Mixed income, the parts that relate to business deposits and loans should be equal to the parts recorded in the property income block, i.e. the parts included in the FISIM corrections in D41R and D41P respectively should equal the deduction as intermediate consumption in Mixed income. For further explanation, see the section on FISIM.

32. The SNA recommends that production activities undertaken by unincorporated enterprises owned by households should be identified separately and treated as "quasi-corporations", if they operate like corporations and keep completely separate accounts. If this is the case, they are included in the corporations' and not in the household sector. This national accounts delineation may differ from households' self-perception and may lead to divergences between what is recorded in national accounts and what people declare as income generated from self-employed business activities in surveys.

Adjustment for the non-observed economy

33. In the national accounts compilation process, adjustments are made to correct for the non-observed economy, such as underground production, i.e. the deliberate concealment of legal activities to avoid tax payments by registered and unregistered units. The nature of the national accounts adjustment and the way it is compiled are country specific. Depending on country practices, the adjustment for the non-observed economy may affect several components of the household accounts, in particular compensation of employees, mixed income and property income received.

Consequences for EGDNA work

34. The template separately distinguishes own account production (B3R1) and the adjustment for underground production that affects mixed income (B3R2) as specific components of mixed income. Since adjustments for underground production are not made in surveys, imputations have to be made by modelling the likelihood of a household to benefit from concealed mixed income. When other corrections need to be made for correcting for the non-observed economy, countries are requested to process that in the exercise and by adding a comment what kind of correction they have made. Furthermore, micro data on mixed income needs to be corrected for the part of income from self-employment that relates to quasi-corporations (see above) before being linked to the corresponding item from the national accounts (B3R3). As the delineation in the micro source may be different from that in the national accounts, this is not always straightforward.

Employer's social contributions (D121 and D122)

35. Compensation of employees includes employers' social contributions which consist of the actual contributions paid by employers to social security funds or to other employment-related social insurance schemes, and imputations for benefits provided by some employers directly to their employees, former employees or dependents without involving an insurance enterprise or autonomous pension fund and without creating a special fund or segregated reserve for the purpose. In the latter situation, the SNA considers existing employees as being protected against various specified needs or circumstances, even though no reserves are built up to provide future entitlements. Remuneration is therefore imputed for such employees equal in value to the amount of social contributions that would be needed to secure the de facto entitlements to the social benefits they accumulate (cf. 2008 SNA, chapter 7.68). These imputed social contributions are specific to the national accounts framework. They are missing in micro data and are therefore not taken into account in the household income definition of the Canberra Group Handbook.

36. Employers' social contributions, both actual and imputed, have a counterpart in social contributions paid by households (D61P), recorded in the re-distribution account. As is described in 2008 SNA paragraphs 8.83 and 8.84, these items are exactly the same as the items recorded under D12, and as a consequence cancel out at the level of disposable income.

Consequences for the EGDNA work

37. Because of the conceptual link, countries should make sure that the quintile results for Employers' actual social contributions as received by households as part of compensation of employees

(D121R) and as paid by households as part of their social contributions (D611P) are identical. The same goes for the Employers' imputed social contributions that are also both recorded as part of compensation of employees (D122R) and as part of net social contributions (D612P).

38. Especially for Employers' imputed social contributions, it may prove difficult to derive the allocation across households, as information will usually be lacking in micro data. If no information is available, it is recommended to derive the distribution on the basis of the distribution of Wages and salaries (D11), or of that of Employers' actual social contributions (D121). Although the inclusion of these items will not affect disposable income (as explained above), countries are still encouraged to provide breakdowns for these items, as they will affect the distribution of primary income and provide insight how re-distributional transactions changes the incomes of households.

Property income, including FISIM in relation to interest paid and received, reinvested earnings on foreign direct investment and investment income disbursements (D4)

FISIM (D41R_FISIM and D41P_FISIM)

39. As well as charging customers directly for banking services, banks also generate service income by lending at a higher rate of interest than they borrow. The service provided by bringing together borrowers and lenders is charged for indirectly, by means of charging higher interest rates to lenders than they pay to depositors.

40. In the SNA, the margin earned by the banking sector from both the difference between the interest rate received on loans and a reference rate, and the difference between a reference rate and the interest rate paid on deposits are treated as service payments, called Financial Intermediation Services Indirectly Measured (FISIM). Within the household sector accounts, FISIM adjustments are made to several components. These include intermediate and final consumption, and interest payments/receipts.

41. The introduction of FISIM into the national accounts resulted in a redefinition of interest (SNA code D41). The relevant national accounts transaction no longer reflects the actual interest flows on deposits and loans. Interest payments and receipts are recorded at a notional reference rate taking into account a FISIM adjustment, reflecting the payment of households for the indirect intermediary service provided by banks (recorded as consumption). This means that if a household receives interest on a deposit, the actual amount will be lower than the notional amount recorded in the national accounts as the bank deducts an amount related to the service it provides (i.e. a higher notional interest receipt for the household sector with an accompanying FISIM payment), whereas if a household pays interest on a loan, the actual payment is higher than the notional amount recorded in the national accounts as the bank adds a service fee for their intermediary service (i.e. a lower notional interest payment by the household sector with an additional FISIM payment). Offsetting adjustments are applied to consumption, increasing intermediate consumption in the case of interest receipts related to business deposits and to interest payments related to mortgage and business loans, and increasing final consumption in case of other deposits and loans (e.g. loans for purchasing household goods and cars).

Consequences for the EGDNA work

42. In the template, actual interest payments and receipts (i.e. D41P' and D41R') are requested, as these correspond best to the amounts that are reported in micro data surveys. Furthermore, a correction item for FISIM is presented to arrive at the notional amounts (D41P and D41R) (also requested in the template) that are needed to be in line with the 2008 SNA, having an upward effect on interest receipts and a downward effect on interest payments (both increasing net property income). More specifically, in the template, the FISIM adjustment is presented as two separate items, on the one hand on the receipt side, as a positive item increasing *interest received (not adjusted for FISIM)* (i.e. D41R') and, on the

other hand, on the payments side, as a negative item reducing *interest paid (not adjusted for FISIM)* (i.e. D41P').

43. For allocating FISIM to households or household groups, it is recommended to make a link to the actual interest payments and receipts by households. If detailed information is available on the different reference rates and actual interest rates for various types of deposits and loans and these can be linked to the various households groups, this can be used to allocate FISIM at a very detailed level. However, if such information is lacking, one could also assume equal margins for all types of deposits and loans, allocating FISIM proportional to the aggregated absolute amounts of interest receipts on the one hand by households or household groups and interest payments on the other hand.

44. Furthermore, in allocating FISIM to households, one should bear in mind that a corresponding correction should be made for the consumption of FISIM. Depending on the type of deposit or loan, it should be recorded as intermediate consumption or as final consumption. If FISIM relates to mortgage loans, it should be recorded as intermediate consumption in the production of housing services related to owner-occupied dwellings. In the template, this will be reflected in a lower value of Operating surplus (B2). If FISIM relates to business loans owed by households or deposits held by household businesses, it should also be recorded as intermediate consumption, but in this case reflected in a lower value of Mixed income (B3). For all other deposits and loans, the related FISIM should be recorded as part of final consumption (CP1261). As there is a direct link between the consumption of FISIM and the correction item recorded in the property income block, one should make sure that in allocating FISIM to household groups, this link is maintained. This means that the sum of FISIM recorded as final consumption (CP1261) and as intermediate consumption (reflected in lower values of B2 and B3) should equal the sum of the two amounts of FISIM recorded as correction items in property income (i.e. D41R_FISIM and D41P_FISIM) at the level of the household (or household group). The breakdown into type of consumption should ideally be made on the basis of information on the types of deposits and loans or type of interest receipt and payments. If that type of information is not available, assumptions should be made to break it down into these three types of consumption.

Reinvested earnings on foreign direct investment (D43R)

45. This item specifically relates to the retained earnings of corporations or quasi-corporations that are part of a foreign direct investment relation. The 2008 SNA defines foreign direct investment in paragraph 21.34 as "a category of cross-border investment associated with a resident in one country (the direct investor) having control or a significant degree of influence on the management of an enterprise (the direct investment enterprise) that is resident in another economy". Any earnings that are not actually distributed to the direct investor are treated as being distributed and reinvested, as the decision to retain some of the earnings is seen as representing a deliberate investment decision on the part of the foreign direct investor (see SNA paragraphs 7.137 and 7.138). Although foreign direct investment usually takes place between corporations or quasi-corporations, a couple of countries record reinvested earnings on foreign direct investment as being received by households. This would imply that they are regarded as being the direct investor in a foreign direct investment relation.

Consequences for the EGDNA work

46. This item was initially lacking from the template, as this item was deemed irrelevant for the household sector. However, as a couple of countries report data on this specific item as being received by households, it has been included to the template to arrive at a complete reconciliation of net property income. With regard to the allocation to the underlying households, this should ideally be based on micro information or information available from foreign direct investment statistics.

Investment income disbursements (D44R)

47. Investment income disbursements consist of three types of disbursements: (1) investment income attributed to insurance policy holders, (2) investment income payable on pension entitlements and (3) investment income attributed to investment fund shareholders. These three components are explained below.

48. For non-life policies, the insurance corporation holds technical reserves that are seen as a liability towards the insurance policy holders. The investment income on these reserves is treated as income attributable to the policyholders, that is distributed to policyholders in the allocation of primary income account (part of D441R) and paid back to the insurance corporation as a premium supplement in the secondary distribution of income account (part of D71P: net non-life insurance premiums). Net non-life insurance premiums comprise both the actual premiums payable by policyholders to obtain insurance cover during the accounting period (premiums earned) and the premium supplements payable out of the investment income attributed to insurance policyholders, less the service charges payable to the insurance corporation.

49. For life insurance policies and annuities, the insurance corporations have liabilities towards the policyholders and annuitants equal to the present value of expected claims. Bonuses declared in connection with life policies are treated as being distributed to policyholders and as premium supplements recorded in the financial account as payable by households and receivable by insurance corporations as changes in life insurance and annuities entitlements. As the recording of this item in the remainder of the accounts differs from the supplements on non-life policies, which are also treated as being distributed to policyholders but have a counterpart in the re-distribution account as premium supplements payable by households and receivable by insurance corporations (D71P), the template uses two codes to make a clear distinction between the two (D441A versus D441B). When countries are able to, they are encouraged to provide this detail in the template.

50. The second category concerns investment income payable on pension entitlements. The exact calculation of this item depends on the underlying type of pension scheme. For *defined contribution* pension schemes, contributions are invested on behalf of former and current employees as future pensioners and the investment income receivable by the pension funds is therefore recorded as property income for the households. The investment income payable on defined contribution entitlements is equal to the investment income on the funds plus any net operating surplus earned by renting land or buildings owned by the fund. For *defined benefit* schemes the increase in the present value of the entitlements due to the unwinding of the discount rate represents the investment income distributed to the former and current employees and should be recorded under this specific item. For both types of schemes, the investment income is attributed to the policyholders and then households are treated as paying an equal amount back into the funds as premium or contribution supplements in the secondary distribution of income accounts (as part of D61). It also forms part of the adjustment item for the change in pension entitlements (D8), which will be explained further on in this section.

51. A third category of investment disbursements is the investment income attributed to investment fund shareholders. Also for these funds, it is reasoned that the shareholders are actually the owners of the investments and therefore should receive all the earning on the investments. As only part of the earnings of investment funds is actually distributed to the shareholders in the form of dividends, the remainder of the earnings is also recorded in the SNA as being distributed to the shareholders (leaving the investment fund with no saving) and being reinvested into the fund via a transaction recorded in the financial accounts.

Consequences for the EGDNA work

52. As investment income disbursements are normally not recorded in micro data surveys, imputations will often be needed. This can be done for the aggregate of the investment disbursements, but as it is composed of components that may have very different distributions and may be linked to different types of auxiliary data, it may best be done at a more detailed level. For the estimation of the distribution of the components, balance sheet information may be used (e.g. insurance technical reserves, life insurance and pension entitlements, and investments in investment funds), premiums paid or other information that can serve as a proxy for distributing the national accounts totals to households.

53. It should be noted that part of the investment income disbursements is treated as premium supplements that need to be recorded in the distribution of income account as part of D61 (the investment income payable on pension entitlements) and D71 (the part of investment income attributed to insurance policy holders that relates to non-life insurance). In that regard, countries should make sure that the breakdown into household groups is identical for the corresponding parts. The other elements of investment income disbursements (i.e. the part of investment income attributed to insurance policyholders that relates to life insurance and the investment income attributed to investment fund shareholders) are also redistributed to the insurance corporation and the investment funds, but these are recorded entirely in the financial accounts. As was mentioned before, the different recording of these disbursements in the remainder of the accounts is another reason why detail on these transactions is provided in the template. It may help countries in attributing the correct amounts to the secondary distribution of income accounts.

Net social contributions paid (D61P)

54. In the SNA, all contributions to social insurance schemes are shown as made by households. They consist of employers' and households' social contributions. The former comprise actual premium payments by employers (D611P), reflecting actual contributions made to social insurance schemes, and imputed social contributions (D612P), related to social benefits provided by employers directly to their employees, former employees or dependents (see before). As is explained in paragraphs 8.83 and 8.84 of the 2008 SNA, both items are exactly the same as those recorded in the primary income account as part of compensation of employees, respectively D121R and D122R.

55. Households' contributions consist of households' actual social contributions (D613P) and social contribution supplements (D614P). The actual contributions reflect the contributions payable by employees on their own behalf, by self-employed and by non-employed persons to social insurance schemes. The contribution supplements consist of the property income earned during the accounting period on the stock of pension and non-pension entitlements, as recorded in the primary income account under item D442R (see before).

56. Set against these contributions is the service fee charged by the unit administering the social security scheme (D61xP). This may be an explicit or an implicit charge (e.g. equal to the sum of costs incurred by the employer administering the scheme) and this amount should be deducted to arrive at the net social contributions.

Consequences for the EGDNA work

57. In the template, the two components relating to employers' social contributions are requested separately. As both items have their counterpart in the primary income account, one should make sure that the reported amounts match, i.e. the results for D611P should be equal to D121R and the results for D612P equal to D122R. For households' social contributions the breakdown into actual premiums and premium supplements is only presented as optional. However, as the latter component has a direct

counterpart in D442R, countries are encouraged to provide the breakdown at the level of these two components. The service charge related to social insurance schemes is presented as a separate item in the template as well (D61xP), to arrive at consistency with the total of net social contributions and its underlying components. However, not all countries may have separate information available on this specific item. In that case, it is assumed that the service charge is already reflected in a reduced amount of imputed social contributions or premium supplements in which case no additional correction is needed.

58. Usually no separate information will be available on the service charge from micro data sources. In that case, countries are advised to determine its distribution on the basis of total social contributions.

Net social contributions received (D61R)

59. As the household sector may include unincorporated enterprises with paid employees, net social contributions may also be recorded as a receipt for the household sector (see 2008 SNA paragraph 8.16). This is the case when an unincorporated enterprise operates an employer-related social insurance scheme itself or provides social insurance benefits directly to its employees. In the former case, an actual social contribution is recorded (under category D611R), whereas in the latter case a social contribution is imputed (recorded under category D612R), equal to the amount of social contributions that would be needed to secure for the same social benefits.

Consequences for the EG DNA work

60. In the template, two categories are included to account for net social contributions received, the first one related to unincorporated enterprises that operate a social insurance scheme themselves, i.e. employers' actual social contributions (D611R). The second category relates to the direct provision of social benefits to employees by unincorporated enterprises. The related social contributions should be recorded under employers' imputed social contributions (D612R).

Social benefits other than STiK paid and received (D62)

61. This category covers "current transfers received by households intended to provide for the needs that arise from certain events or circumstances" (2008 SNA paragraph 8.87). However, whereas they mainly constitute a resource for households, it may also appear on the uses side of the household sector, relating to unincorporated enterprises that operate their own social insurance scheme or that directly provide social benefits to their employees.

Consequences for the EG DNA work

62. In the template, this category is both presented as a use (D62P), related to benefits provided by unincorporated enterprises, and as a resource (D62R), related to the receipt of social insurance benefits by households.

Other current transfers (D7)

63. In the exercise, distributional information is needed on net non-life insurance premiums and claims, and on net miscellaneous current transfers received and paid. However, as the payments and receipts may vary across groups of households, it may be better to derive the distribution on the basis of gross figures than on the basis of net figures. This may lead to more accurate results. Therefore, in the template, also the components of these net flows are presented. Although it is optional for countries to use them, it is strongly recommended.

64. Non-life insurance premiums are part of other current transfers. They consist of the actual premiums payable by policyholders and the premium supplements payable out of the investment income attributed to insurance policyholders less the service charges payable to the insurance corporation. In compiling the distributional results, one has to be aware that the part that relates to premium supplements is the counterpart of item D441AR in the primary income account, and that this link should be reflected in the results for non-life insurance premiums.

65. In the national accounts, transfers between resident households cancel out at the level of total households. As a consequence, usually no explicit estimates of these flows are available. However, at the level of individual households or at the level of groups of households, this consolidation does not apply, as the receipts and payments will concern different households. That is why this specific item is also shown in the template. In some cases, micro data may be available for their estimation. Otherwise, experts are encouraged to investigate how to best estimate and allocate these transfers across the distribution.

Consequences for the EGDNA work

66. As it is understood to lead to more accurate results, countries are encouraged to make distributions for the other current transfers at the more detailed level presented in the template. Furthermore, compilers have to be aware that part of non-life insurance premiums (D71P) relates to premium supplements as recorded under item D441AR. The distribution of these premium supplements should be reflected in the distribution of the non-life insurance premiums, i.e. for each household group the total non-life insurance premiums should at least be equal to or exceed the premium supplements. Finally, countries are encouraged to provide data on transfers between resident households.

Social transfers in kind (STiK) by type (D63)

67. The size of Social Transfers in Kind (STiK) is often significant compared to household disposable income and can show large variations across countries. Omitting STiK can have a major impact on inequalities; therefore, especially in international comparisons, the national accounts concept of adjusted disposable income, a concept that includes STiK, is recommended.

Health

68. Two approaches to allocate STiK have been distinguished within the EG DNA: the *actual value* approach, according to which the actual values of health benefits are allocated to the various household groups, and the *insurance value* approach, according to which an insurance premium equivalence is allocated to the households. In the absence of further information, the latter approach basically comes down to allocating the average per capita STiK for health to each individual.

69. It has become clear that few countries have the necessary information available to implement the actual value approach. Furthermore, it has also become clear that the insurance value approach can be refined by segmenting the population based on socio-demographic information and allocating STiK in line with the various needs/provision costs related to each population segment. For example, it is clear that health related spending is highly age dependent. Allocating STiK on Health to each individual proportional to the STiK spending by age, would also ensure that the results between the two approaches come much closer. Indeed, as part of the work of the EG DNA, experts are expected to implement a basic scenario following the *insurance value* approach by relying on as much socio-demographic information as possible to refine individual allocations.

70. To implement the basic scenario following the insurance value approach, it is suggested to apply the following (minimum) procedure:

- Step1: Adjust the national accounts total for the part received by people living in non-private dwellings.

- Step 2: Try to find a source providing an estimate of average public health spending by age, and perhaps other categories (e.g. gender). If no adequate information is available at national level, experts may want to use the estimates provided in Annex 4. This table shows 2010 health care cost by age groups (on a per capita basis, as a percentage of GDP per capita). To obtain the amount of per capita expenditures by age these figures should be multiplied by the GDP per capita value of the breakdown year. In using this information, it is assumed that the age dependency profile of health expenditures is relatively stable when moving from the breakdown year to 2010.

- Step 3: Impute to each individual the average health care cost of a person with the corresponding age. Each individual is thus assumed to receive a public benefit determined by the average public spending of his/her group, irrespective of whether or not actual use of health care services has been made.

- Step 4: Scale up or down the imputations so that they match the adjusted national accounts totals (as determined in step 1).

Education

71. Similarly to health expenditures, an actual value approach or a modelled approach using socio-demographic information can be used to allocate spending on education to individuals and households. For example, if socio-demographic information is available on age, or schooling status/level of education (and whether or not the relevant students are in public education), and STiK spending per capita for all these sub-groups is available, then education related STiK allocations can be made fairly close to the actual value.

Other STiK

72. Social transfers in kind for health and education typically represent the largest share of overall STiK. If no further information is available, then other STiK could be allocated flatly to all households or all individuals. At the same time, countries are encouraged to work on a finer decomposition of other STiK, such as for housing, early childhood education and child care services, or long-term elderly services, and make imputations separately for each STiK category using all the available socio-demographic information.

Adjustment for the change in pension entitlements (D8)

73. An adjustment for the change in pension entitlements is necessary because of the way employment-related pension contributions paid to pension funds and pension benefits received from these funds are treated in the SNA. They are recorded as current expenditure/income, while on the other hand they are also considered as a (dis)saving, adding to the value of pension entitlements. For this reason, a correction has to be made when saving is calculated, reflecting that these flows both affect income and households' savings, increasing their pension entitlements as a consequence of pension premium payments and decreasing their entitlements via the receipt of pension benefits.

74. The adjustment item (D8) covers those parts of social insurance schemes for which the liabilities are recognised in the SNA (i.e. employment-related social insurance schemes). Pensions due under social assistance and social security schemes are excluded (see 2008 SNA, paragraphs 9.20-9.25).

75. The adjustment for the change in pension entitlements is equal to:

- the total value of the actual and imputed social contribution payable into pension schemes (parts of D611P, D612P and D613P that relate to pensions),
- plus the total value of contribution supplements payable out of the property income attributed to pension fund beneficiaries (part of D614P that relates to pensions),
- minus the value of the associated service charges (part of CP125 that relates to pension schemes),
- minus the total value of the pensions paid out as social insurance benefits by pension schemes (part of D62R that relates to pensions).

Consequences for the EGDNA work

76. As the adjustment for the change in pension entitlements is the result of various components with their own specific distributions across households, countries should make sure that they derive the distribution for this item accordingly. Ideally this is done at the level of the household or detailed household group, but as a minimum it is recommended to distinguish between the active population (for which actual pension contributions (part of D61P), related service charges (part of CP125) and their part of the premium supplements are relevant), and the retired population (for which pension benefits (part of D62), related service charges (part of CP125) and their part of the premium supplements are relevant). This provides for a better link with the main underlying flows.

Taxes less subsidies on production and imports (D2-D3)

77. As described in the 2008 SNA, "taxes are compulsory, unrequited payments, in cash or in kind, made by institutional units to government units" (see paragraph 7.71 of the 2008 SNA). These are defined as unrequited as the government provides nothing in return to the individual unit making the payment. The government can use the funds raised in taxes to provide goods or services to other units. The 2008 SNA also states (see paragraph 7.73) that "taxes on production and imports (D2) consist of taxes on products (D21) and other taxes on production (D29)". On the other hand, "subsidies are current unrequited payments that government units, including non-resident government units, make to enterprises on the basis of the levels of their production activities or the quantities or values of the goods or services that they produce, sell or import" (see 2008 SNA, paragraph 7.98). Similar to taxes, subsidies (D3) consist of subsidies on products (D31) and other subsidies on production (D39).

Consequences for the EGDNA work

78. As taxes and subsidies represent an important part of government redistribution policies (Mitchell, 2019), the inclusion of distributional information on household payments of taxes and receipts of subsidies on products and production is very relevant in allowing for a comprehensive analysis of the impact of these policies and their impact on various groups within the household sector[9].

79. Only taxes (and subsidies) on products and production that are explicitly paid (and received) by the households sector are considered in the EG DNA work. While it is possible that part of the taxes on products and production paid by other sectors may be passed on and in the end get paid by the households sector (as final consumers), to assume this in its entirety not only ignores the components of final demand used by government sector and non-residents, but it also removes the ability for the

[9] For this reason, the relevant items have been added to the collection template. However, as this has only been done in 2019, the results were only collected as optional in the third exercise and have not yet been included in the online databases of Eurostat and the OECD.

statistical outputs to properly reflect the behaviour of organisations in response to changes in taxation policy.

80. Due to the lack of micro information on household payment (and receipt) of taxes (and subsides) on products and production, a proxy indicator must be used. The different nature of taxes and subsidies on products and those levied on production necessitates separate indicators for each item. While these difficulties make estimation challenging, it is still possible to compile reliable estimates by using some standard assumptions and modeling, which is not uncommon in the production of distributional statistics.

81. The vast majority of taxes on products becomes payable when the good or service is consumed. Therefore, it is logical to use consumption as an indicator for the distributional split of taxes less subsidies on products. This is done by generating an applicable tax rate to apply to the amount of household consumption. Whereas this can be done at the level of total consumption expenditure, it is recommended to calculate appropriate tax rates for each consumption category, to arrive at more accurate distributions.

82. Taxes (less subsidies) on production are not related to the purchase of products but rather to the production of these products. While the amount of these taxes (less subsidies) as paid by the household sector will usually not be very large, the household sector may still pay this tax as part of their production activities. Due to the different nature of taxes and subsidies on production, it would be more suitable to distribute these amounts on the basis of the distribution of *gross mixed income*[10].

[10] More detailed information on conceptual issues, as well as the method that statistical agencies can apply to generate distributional outputs related to taxes less subsidies on products and production for the household sector, is available in "The distribution of taxes less subsidies on products and production into household groups" (prepared by John Mitchell for the meeting of the Expert Group on Disparities in the National Accounts, June 27-28 2019)..

PART III: Overview of the questionnaire template for the data and metadata collection

83. This section of the guidelines provides a general overview of the collection template, which has been designed to report the data and the metadata for distributional information on household income, consumption and saving, as well as the additional socio-demographic information. The template consists of eight sheets: a 'read me' sheet with instructions on how to fill out the questionnaire; a sheet for general information; three tables for the collection of the data ('income', 'consumption & saving', and 'socio-demographic information'); an optional sheet for the collection of the distribution of underlying micro data; a sheet for the metadata; and a final sheet to run consistency checks.

Sheet 1: Read me sheet

84. The first sheet of the template provides instructions on how to fill out the questionnaire. It contains explanations of the separate sheets as well as general information on some of the conventions and technical features used in the template. Some of them are explained in more detail below.

Missing data and zero values

85. A first convention is how to deal with missing data and zero values in the data sheets. For missing data, the values should be reported either as "NAP" (i.e. non-applicable), i.e. if it is not applicable for a country, or as "NAV" (i.e. non-available), i.e. if it is applicable but not compiled in the exercise. Zero values ("0") should only be reported when data are applicable and computed as equal to zero. Furthermore, no cells in the data field should be left blank. Please see the following examples, showing a breakdown of an aggregate into its underlying components, which explain how to correctly fill out the template (more examples are provided in Annex 6):

Wrong: a cell is empty where the value of "10" should have been reported.

D1R	Compensation of employees	30
D11R	Wages and salaries	10
D121R	Employers" actual social contributions (*counterpart in D611)*	10
D122R	Employers' imputed social contributions (*counterpart in D612*)	

Wrong: a cell is empty where either a numeric value greater than "10" or "NAV" (according to country-specific circumstances) should have been reported.

D1R	Compensation of employees	
D11R	Wages and salaries	10
D121R	Employers" actual social contributions (*counterpart in D611)*	NAV
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Wrong: a cell is empty where "NAV" should have been reported.

D1R	Compensation of employees	30
D11R	Wages and salaries	10
D121R	Employers" actual social contributions (*counterpart in D611)*	
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Wrong: a cell is empty where the value of "10" should have been reported.

D1R	Compensation of employees	
D11R	Wages and salaries	10
D121R	Employers" actual social contributions (*counterpart in D611)*	NAP
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAP

Wrong: a cell reports a wrong value where either "NAP" or the value of "0" (according to country-specific circumstances) should have been reported.

D1R	Compensation of employees	30
D11R	Wages and salaries	10
D121R	Employers' actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R holds.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAP

86. To understand the importance of explicitly reporting "NAP"/"NAV"/"0", look at the following example. The item D8 (adjustment for the change in pension entitlements) is only relevant for countries with employment-related pension schemes in which pension entitlements accrue over time. In countries such as Australia and France, where employment-related pension schemes do not exist, a "NAP" should be reported for item D8. A "NAV" would imply that D8 is applicable, but not compiled, whereas a "zero" would be interpreted as if D8 is applicable and is computed as zero. This would lead to an incorrect interpretation of the results and may possibly affect the calculation of results for "savings" (B8). In this regard, reporting "NAV" would imply that it is not possible to derive the savings results, as one of its components (i.e. D8) would not be available.

87. Exceptionally and with particular regard to the Micro source column in the data sheets (discussed in detail hereinafter in this section), data compilers are encouraged to report a "NAP" for those items that are not relevant in micro data sources (e.g. not usually available from the surveys or other micro data sources), rather than a "NAV". This practice will avoid possible error messages in the "Checks" sheet (also discussed hereinafter in this section) which would show up on the basis of checking consistency based on national accounts (macro) concepts. As micro data sources are usually based on different concepts, some of the national accounts items are in that respect 'not applicable' from the point of view of the micro data source.

Signs

88. A second convention used in the template is that all values should be reported with a positive sign, even for those transactions which negatively contribute (components on the 'use' side, coded with 'P' for 'paid') to the computation of the aggregates and/or the balancing items (such as B5, B6, B8 etc.).

89. The only exceptions are (i) data for FISIM on the 'use' side (D41P_FISIM) on the "Income" sheet, which should always be reported with a negative sign, and (ii) data for those transactions which might assume a negative value by definition. The latter can also be reported with a negative sign if these turn negative as a consequence of the compilation process. On the "Income" sheet this concerns items B2R+B3R, B2R, B2R1, B2R2, B3R, B3R1, B3R2, B3R3, D4N, D43R, B5, D7N, D72R-D71P, D75N, D75x, B6 and B7; On the "Consumption & Saving" sheet it concerns items CP071, P33-P34, D8 and B8, D2-D3, D21-D31 and D29-D39. On the "Micro source" sheet, the same rule applies, whereas on the "Socio-demographic Information" sheet all items should be reported with a positive sign.

Embedded formulas

90. Tables in the data sheets already include formulas for the calculation of certain aggregates and sub-aggregates. This is done for convenience, but can be overwritten if countries do not have the appropriate underlying detail available. In that case, users can delete the formula and report the value directly.

Sheet 2: General information sheet

91. This sheet allows for reporting of general information on the year, the currency and the unit of measurement used in the questionnaire. It also includes a section on the contact details. Information on

the contact details will be of particular importance once the data collection has matured beyond the current experimental stage. Countries can indicate two contact persons: the first is typically a senior official overseeing the whole estimation process, whereas the second is a more hands-on official able to directly address any technical question that might arise during the validation process.

92. The Secretariat recommends using information for the most recent year available for micro data sources (if information referring to the same year is also available for annual national accounts data). However, breakdowns on income and consumption components should preferably be compiled for the same year within a country; otherwise estimating savings rates may not be possible.

93. Furthermore, for the purpose of analyses, it is important that results from the current data collection and the previous exercise(s) are comparable. When countries have made significant changes in their methodologies or use different data sources compared with the previous one(s) (or if there are other changes that lead to incomparability of the results), which lead to incomparability between the results of the exercises, they are kindly requested to provide updated results for the years covered in the previous exercise(s).

Sheets 3 and 4: Data sheets for 'income' and 'consumption & saving'

94. Data are collected for income in one sheet and consumption and saving in another. The two data sheets share the same structure (see Table 2) with the aggregates in the columns and the transactions in the rows.

Table 2: Data sheets

Transactions		SNA Households aggregates		Micro Source	Adjusted household aggregates broken down by quintile				
		Original	Adjusted		Q1	Q2	Q3	Q4	Q5
Transaction Codes	**1st level**								
	2nd level								
	3rd level								
	1st level								
	…								
	1st level								
	2nd level								
	3rd level								

95. The first two columns include the SNA aggregates, with the original value of the aggregate (income, consumption or saving depending on the table) followed by the scope-adjusted aggregate (correcting for NPISH, expenditure of non-residents, and non-private households). The third column includes the total values of the corresponding item in the micro source. Finally, the last columns are dedicated to the distributional information. As was mentioned before, the Secretariat would like to receive distributional information by income quintile, but when countries are also capable of providing distributional data by household type and main source of income, they are kindly requested to do so, as this provides very valuable information. For this purpose, there is also an optional part in these tables where data can be recorded regarding these two classifications.

96. Transactions are reported in the rows and are broken down into three levels:

- The first (top) level includes a small number of highly aggregated key income, consumption and savings items. These most likely will be analysed and used by end users (especially in a macroeconomic context), and correspond to a minimum to be disseminated to the users.

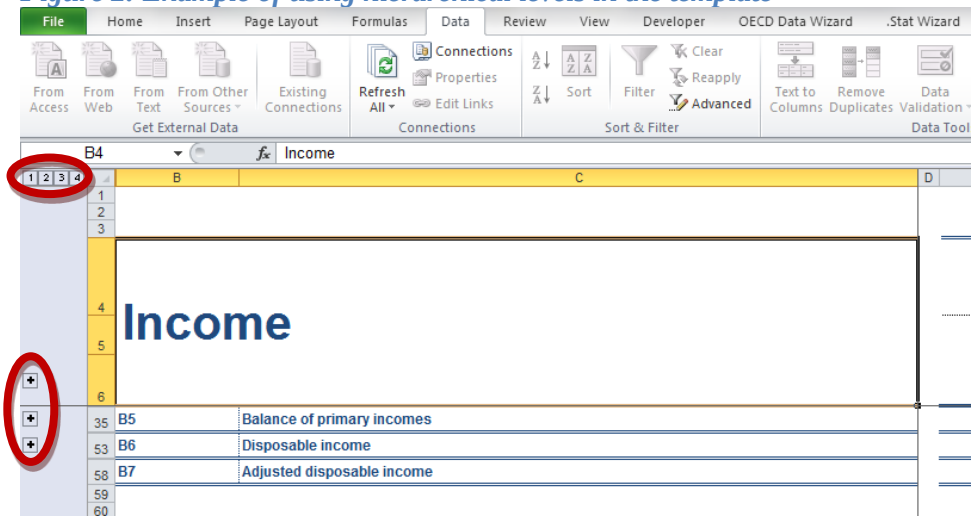
Table 3: First level of disaggregation for the aggregates

Income	
B5	Balance of primary incomes
B6	Disposable income
B7	Adjusted disposable income
Consumption	
P31NC	Final national consumption expenditure
P4	Actual final consumption
Saving	
B8	Gross saving

- The second level contains more detail. It corresponds to the typical level of detail reported internationally for institutional sectors in the SNA. This corresponds to the desirable level of publication once the exercise is considered to have matured beyond the experimental stage.
- The third level contains several transactions (or subdivisions of transactions) that, in most cases, are below the standard level of detail that needs to be provided in international data deliveries. Nonetheless, these transactions have been identified as very useful because they have a direct link with other variables in the exercise or are critical in the micro-macro matching exercise, as for these transactions imputations/allocations based on national accounts information can be significant. However, as not all information is directly needed, part of these extra details is only optional[11]. It is important to stress that these are not meant to be part of the regular publication.

97. In the data sheets, the user has the possibility to group and ungroup transactions by clicking on the numbers (corresponding to the different hierarchical levels) or the pluses ("+") as reported in Figure 2 as an example.

Figure 2: Example of using hierarchical levels in the template



[11] It concerns the detailing for investment income disbursements (D44), net non-life insurance claims minus premiums (D72R-D71P), and net miscellaneous current transfers (D75N).

98. Also, transactions are shaded with different colours which characterise the different hierarchical levels. Please see Figure 3 as an example.

Figure 3: Example of shading used in the template

D4N	**Net property income received / Net property income**
D4R	Property income received
D41R	Interest received
D41R'	Interest received (not adjusted for FISIM)
D41R_FISIM	Adjustment for FISIM (positive sign)
D42R	Distributed income of corporations
D43R	Reinvested earnings on foreign direct investment
D44R	Investment income disbursements
D441R	Investment income attributable to insurance policy holders
D441AR	Property income received attributed to non-life insurance policy holders (*included in current transfers*)
D441BR	Property income received attributed to life insurance policy holders
D442R	Investment income payable on pension entitlements (*included in net social contributions paid*)
D443R	Investment income attributable to collective investment funds share holders
D45R	Rent received
D4P	Property income paid
D41P	Interest paid
D41P'	Interest paid (not adjusted for FISIM)
D41P_FISIM	Adjustment for FISIM (negative sign)
D45P	Rent paid
B5	**Balance of primary incomes**

99. In reporting distributional information at the different levels, one has to make sure that the aggregates are equal to the sum of the underlying components. This need not necessarily be the case if the scaling is done separately for the aggregates and for the underlying components. This is due to the fact that the distribution of micro data, as well as the size of the micro-macro gap may differ per underlying item. Annex 5 provides an example of how this may lead to different results. As aligning at the more detailed level is usually understood to gain the most reliable results, countries are encouraged to derive the aggregates as the sum of the underlying components.

100. A flag indicating the method used for the disaggregation into household groups can be chosen for each transaction in the data tables. Typically, four classes of methods are distinguished:

- A: The transaction values in micro sources are scaled up or down so that their totals match the corresponding totals in national accounts;
- B: Indirect method based on proxies. Missing or unreliable micro information is estimated by using the distribution of a different component as a proxy, therefore assuming that the two are distributed in the same way. Adjustment is made at the micro level before benchmarking aggregates to the national accounts totals;
- C: Indirect method based on external data. Missing or unreliable micro information is estimated using exogenous information (e.g. socio-demographic information) available at the individual and at the household levels and making assumptions (in cases no micro information at all is available) before applying the distribution to SNA totals; and
- D: The national accounts total is distributed among all households at the end of the calculation process in a manner that the inclusion or exclusion of the component does not have an impact on the commonly used distributional indicators.

101. Should it not be possible to allocate a transaction, or should the allocation be possible only for a combination of transactions, data should be reported in the tables without changing the layout of the template (by adding, deleting or merging rows) and a description of the deviation should be provided in

the metadata sheet. Please note, in this regard, that the questionnaire has a locked structure and includes locked cell-fields. This measure has been implemented in order to prevent the user from changing the structure of the Excel template, which would compromise the functions of the questionnaire devices, such as the consistency checks integrated-procedure (see later on).

Sheet 5: Socio-demographic information sheet

102. The Secretariat recommends providing additional socio-demographic information for the various household groups on the number of consumption units and the number of households. This will allow the Secretariat converting the reported totals into the per consumption unit and per household figures. Furthermore, for the income quintile breakdown, the sheet asks for the breakdown of households per quintile into household type[12] and housing status[13], and for the breakdown of persons per quintile into age group[14], gender, labour market status[15], and highest level of education achieved[16].

Sheet 6: Micro source sheet

103. To have more insight in how the alignment to the national accounts totals affects the distributional information, the Secretariat would like to receive information on the underlying micro results per household group (i.e. income quintiles, household type and main source of income) as optional data, if countries are able to derive this information. Countries are encouraged to provide this underlying data for the various items for which the distributional results are based on underlying micro data (indicated with option A in the 'income' and 'consumption & saving' sheets). The classification of households should be similar to the classification in the main data sheets, i.e. households that are classified into the first quintile on the basis of the EG DNA exercise should also be included in the first quintile when providing data for the micro data sheet.

Sheet 7: Metadata sheet

104. The metadata sheet allows reporting of any deviations from the guidelines that might impair the international comparability of the data, or deviations from the previous exercise, which might hamper comparability over time. The sheet is built around a set of questions organised into sections, addressing some key issues characterising the various steps of the proposed procedure. Issues addressed in the metadata sheet include: the choice of reference year, the scope adjustment to the national accounts

[12] Categorizing households into eight groups, 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.

[13] Categorizing households into three groups, i.e. a) rental, b) owner-occupied with mortgage, and c) owner-occupied without mortgage.

[14] Categorizing individuals into six age groups, i.e. a) 0-14, b) 15-24, c) 25-34, d) 35-44, e) 45-64, and f) 65+.

[15] Categorizing individuals according to their main activity, distinguishing: 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 by status. These categories are derived on the basis of the ILO International Classification by Status in Employment (ICSE) 1993.

[16] Categorizing individuals according to the highest level of education achieved. The categories used in the template are derived on the basis of the International Standard Classification of Education (ISCED-2011) which distinguishes nine categories, i.e. 0) less than primary education, 1) primary education, 2) lower secondary education, 3) upper secondary education, 4) post-secondary non-tertiary education, 5) short-cycle tertiary education, 6) bachelor's or equivalent level, 7) master's or equivalent level, 8) doctoral or equivalent level, and 9) not elsewhere classified. In the template, some of these levels are combined to arrive at a smaller level of detail, distinguishing: a) low (corresponding to levels 0-2 of the ISCED-A, 2011), b) middle (3-5), c) high (6-8), and d) not elsewhere classified (9). As individuals that never attended an education program (including small children) are classified in category zero according to the ISCED-A, they have to end up in category 'low' in the EG DNA template.

data, the household definition and classification used, the level of breakdown for income and consumption, the treatment of STiK and the comparability of the data in relation to previous exercises. If the answers with regard to the methodological questions have not changed in comparison with the previous exercise, these can be left empty.

Sheet 8: Checks sheet

105. In the "Checks" sheet users have the possibility to run the consistency checks integrated-procedure, which detects any inconsistencies in the data (within the transaction hierarchy and between the totals and the various breakdowns into household groups). These checks are based on a two-step approach:

- A first module (activated through the yellow button "*Click here to check reported cells*" in the "Checks" sheet) checks whether all data cells in the data sheets have been filled with either a value, "NAV" or "NAP", and thus no cell has been left empty. An automatic message will pop up, either:
    - to inform the user that all cells in the data sheets have been filled in (and thus the second module can be launched) or
    - to warn the user that some cells (highlighted in yellow within the data sheets) have been left empty and that these should be filled in first. Once all empty cells have been corrected, the user is encouraged to re-run the first module before moving to the second module.
- A second module (activated through the blue button "*Click here to perform consistency checks*" in the "Checks" sheet) checks whether any inconsistency exists in the data (within the transaction hierarchy and between the totals and the various household breakdowns). Any detected discrepancy is automatically reported in the table of the "Checks" sheet, with information on:
    a) "Table": name of the sheet where the discrepancy is detected;
    b) "Aggregate(s) [column(s)]": name of the aggregate/column concerned;
    c) "Transaction(s) [row(s)]": name of the transaction/row concerned and/or the equality not respected;
    d) "Difference": the numeric difference detected;
    e) "Explanation": information on the detected error and on possible solutions for its correction.

106. The information on "Transaction" (point c above) is generated as a hyperlink that re-directs the user to the 'parent' cells (in the data sheets) where the inconsistency is detected. Also, the "Difference" and the "Explanation" cells are shaded in red or green, depending on the magnitude of the detected difference (respectively, if the absolute difference is above 5, or equal or below 5). This means that a green-shaded cell usually refers to a difference that is due to a rounding error and can be deemed acceptable. Cells related to inconsistencies that do not show a difference (e.g. due to misreporting of "NAV"/"NAP") are red-shaded.

107. The user is strongly encouraged to correct any inconsistency before returning the completed template to the OECD. In case of any questions regarding the template and for any issue relating to the integrated consistency checks procedure[17], please contact the Secretariat via EGDNA@OECD.org.

[17] The VBA Excel Macro underlying the consistency checks procedure integrated in the template might not properly work if run outside the Office Excel environment. Please get in contact with the Secretariat in case you are not able to run the procedure.

REFERENCES

Accardo, Jérôme; Bellamy, Vanessa; Consalès, Georges; Fesseau, Maryse; Le Laidier, Sylvie; and Raynaud, Émilie, "Inequalities between households in the national accounts – Breakdown of household accounts", 2009.

Carson, Carol S., "The Underground Economy: An introduction", *Survey of current business*, Volume 64, Number 5, p21-37, 1984.

Cifaldi, Giulia and Neri, Andrea, "Asking income and consumption questions in the same survey: what are the risks?", *Working papers by Banca d'Italia*, 2013.

Coli, Alessandra and Tartamella, Francesca, "Using administrative and survey data to analyse tax evasions from unregistered labour", *IARIW-paper*, 2014.

D'Alessio, Giovanni and Faiella, Ivan, "Non-response behaviour in the bank of Italy's survey of household income and wealth", *Working papers*, number 462, 2002.

D'Alessio, Giovanni and Neri, Andrea, "Income and wealth sample estimates consistent with macro aggregates: some experiments", *Occasional paper Banca D'Italia*, no. 272, 2015.

Eurostat, IMF, OECD, UN, World Bank, "System of National Accounts 2008", New York, 2009.

Fesseau, Maryse and Mattonetti, Maria L., "Distributional measures across household groups in a national accounts framework – Results from an experimental cross-country exercise on household income, consumption and saving", *OECD Statistics Working Paper series*, 2013.

Fessler, Pirmin; Kasy, Maximilian; and Lindner, Peter, "Survey mode effects on income inequality measurement", *IARIW-paper*, 2012.

International Labour Organisation, "Resolution concerning the International Classification of Status in Employment (ICSE), adopted by the Fifteenth International Conference of Labour Statisticians", 1993.

Lohmann, Henning, "Comparability of EU-SILC survey and register data: The relationship among employment, earnings, and poverty", 2010.

Meyer, Bruce D.; Mok, Wallace K.C.; and Sullivan, James X., "The under-reporting of transfers in household surveys: its nature and consequences", 2008.

Mitchell, John, "The distribution of taxes less subsidies on products and production into household groups", *OECD paper*, 2019.

Romanov, Dimitri and Gubman, Yury, "Well-being and measurement error of income reported in a social survey versus income recorded by a tax administration", *IARIW-paper*, 2012.

Sabelhaus, John; Johnson, David; Ash, Stephen; Garner, Thesia; Greenlees, John; Henderson, Steve; and Swanson, David, "Is the consumer expenditure survey representative by income?", 2012.

Törmälehto, Veli-Matti, "High incomes and affluence: Evidence from EU-SILC", *paper prepared for the 2014 EU-SILC International Conference*, held in Lisbon on 16-17 October 2014.

UNESCO Institute for Statistics, "International Standard Classification of Education (ISCED-2011)", Montreal, 2012.

United Nations Economic Commission for Europe, “Canberra Group Handbook on Household Income Statistics – Second Edition 2011”, Geneva, 2011.

Vermeulen, Philip, “How fat is the top tail of the wealth distribution?”, *ECB Working paper series*, N.1692, 2014.

Zwijnenburg, Jorrit; Bournot, Sophie and Giovannelli, Federico, “OECD Expert Group on Disparities in a National Accounts Framework – Results from the 2015 exercise”, *OECD Working Paper,* No. 76, 2017.

Zwijnenburg, Jorrit; Bournot, Sophie; Grahn, David and Guidetti, Emmanuelle, “OECD Expert Group on Disparities in a National Accounts Framework – Results from the 2020 exercise”, *OECD Working Paper,* 2021, forthcoming.

Zwijnenburg, Jorrit, “Further enhancing the work on household distributional data: Techniques for bridging gaps between micro and macro results and nowcasting methodologies for compiling more timely results”, *IARIW-paper*, 2016.

ANNEX 1: Income, consumption and saving: transactions and relationships in the national accounts framework using the associated codes

INCOME

B2R+B3R	**Operating surplus and mixed income**	
B2R	**Operating surplus from actual and imputed rentals**	**= B2R1 + B2R2**
B2R1	Owner occupied dwellings	
B2R2	Leasing of dwellings	
B3R	**Mixed income**	**= B3R1 + B3R2 + B3R3**
B3R1	Own account production	
B3R2	Underground production	
B3R3	Mixed income excluding underground and own account production	
D1R	**Compensation of employees**	**= D11R + D121R + D122R**
D11R	Wages and salaries	
D121R	Employers' actual social contributions	Counterpart in D611
D122R	Employers' imputed social contributions	Counterpart in D612
D4N	**Net property income**	**= D4R – D4P**
D4R	**Property income received**	= D41R + D42R + D43R + D44R + D45R
D41R	Interest received	= D41R' + D41R_FISIM
D41R'	Interest (not adjusted for FISIM)	
D41R_FISIM	Adjustment for FISIM (positive sign)	
D42R	Distributed income of corporations	
D43R	Reinvested earnings on foreign direct investment	
D44R	Investment income disbursements	= D441R + D442R + D443R
D441R	Investment income attributable to insurance policy holders	= D441AR + D441BR
D441AR	*Property income received attributed to non-life insurance policy holders (optional)*	*Included in D71*
D441BR	*Property income received attributed to life insurance policy holders (optional)*	
D442R	*Investment income payable on pension entitlements (optional)*	*Included D614*
D443R	*Investment income attributable to collective investment funds shareholders (optional)*	
D45R	Rent	
D4P	**Property income paid**	= D41P + D45P
D41P	Interest paid	= D41P' + D41P_FISIM
D41P'	Interest (not adjusted for FISIM)	
D41P_FISIM	Adjustment for FISIM (negative sign)	
D45P	Rent	
B5	***Primary income***	***= B2R + B3R + D1R + D4N***
D5P	**Current taxes on income and wealth**	
D61P	**Net social contributions paid**	**D611P + D612P + D613P + D614P - D61xP**
D611P	Employers' actual social contributions	Counterpart in D121R
D612P	Employers' imputed social contributions	Counterpart in D122R
D613P+D614P	Households' social contributions (actual and supplements)	Including D442R
D613P	*Households' actual social contributions*	
D614P	*Households' social contribution supplements*	*Including D442R*
D61xP	Social insurance scheme service charges	
D61R	**Net social contributions received**	**D611R + D612R**
D611R	Employers' actual social contributions	
D612R	Employers' imputed social contributions	
D62P	**Social benefits other than STiK paid**	
D62R	**Social benefits other than STiK received**	
D7N	**Other current transfers (net)**	**= D72R – D71P + D75N**
D72R-D71P	Net non-life insurance claims minus premiums	
D72R	*Non-life insurance claims*	
D71P	*Non-life insurance premiums*	*Including D441AR*
D75N	Net miscellaneous current transfers received – paid	= D75R – D75P
D75R	*Miscellaneous current transfers received*	
D75P	*Miscellaneous current transfers paid*	
D75x	Of which transfers between resident households	
B6	***Disposable income***	***= B5 – D61P + D61R – D62P + D62R + D7N***
D63R	**Social Transfers in Kind**	
D63R1	Education	
D63R2	Health	

D63R3	Other	
B7	***Adjusted disposable income***	***= B6 + D63R***

CONSUMPTION

CP010	Food and non-alcoholic beverages	
CP020	Alcoholic beverages, tobacco and narcotics	
CP030	Clothing and footwear	
CP040	Housing, water, electricity, gas and other fuels	= CP041 + CP042 + CP043 + CP044 + CP045
CP041	Actual rentals on housing	
CP042	Imputed rentals on housing	
CP043	Maintenance and repair of dwellings	
CP044	Water supply and miscellaneous	
CP045	Electricity, gas and other fuels	
CP050	Furnishings, household equipment and routine household maintenance	
CP060	Health	= CP061 + CP062 + CP063
CP061	Medical products, appliances and equipment	
CP062	Out-patient services	
CP063	Hospital services	
CP070	Transport	= CP071 + CP072 + CP073
CP071	Purchases of vehicles	
CP072	Operation of personal transport equipment	
CP073	Transport services	
CP080	Communication	
CP090	Recreation and culture	
CP100	Education	
CP110	Restaurants and hotels	
CP120	Miscellaneous goods and services	= CP12x + CP1261 + CP125
CP12x	Miscellaneous (less FISIM, less insurance)	
CP1261	FISIM	
CP125	Insurance expenditures (life and non-life)	
P31DC	Final domestic consumption expenditure	= CP010 + CP020 + CP030 + CP040 + CP050 + CP060 + CP070 + CP080 + CP090 + CP100 + CP110 + CP120
P33-P34	Adjustment for expenditures by resident households abroad minus expenditures by non-resident households on the territory	
P33	Resident household expenditure abroad	
P34	Non-resident household expenditure on the territory	
P31NC	Final national consumption expenditure	= P31DC + P33 - P34
D63	Social Transfers in Kind	
P4	Actual final consumption	= P3 + D63

SAVING

D8	Change in net equity of households in pension funds	On basis of the parts of D61P (+), D62R (-) and CP125 (-) that relate to employment-related pensions.
B8	Saving	= B6+D8-P3=B7+D8-P4

MEMORANDUM ITEMS

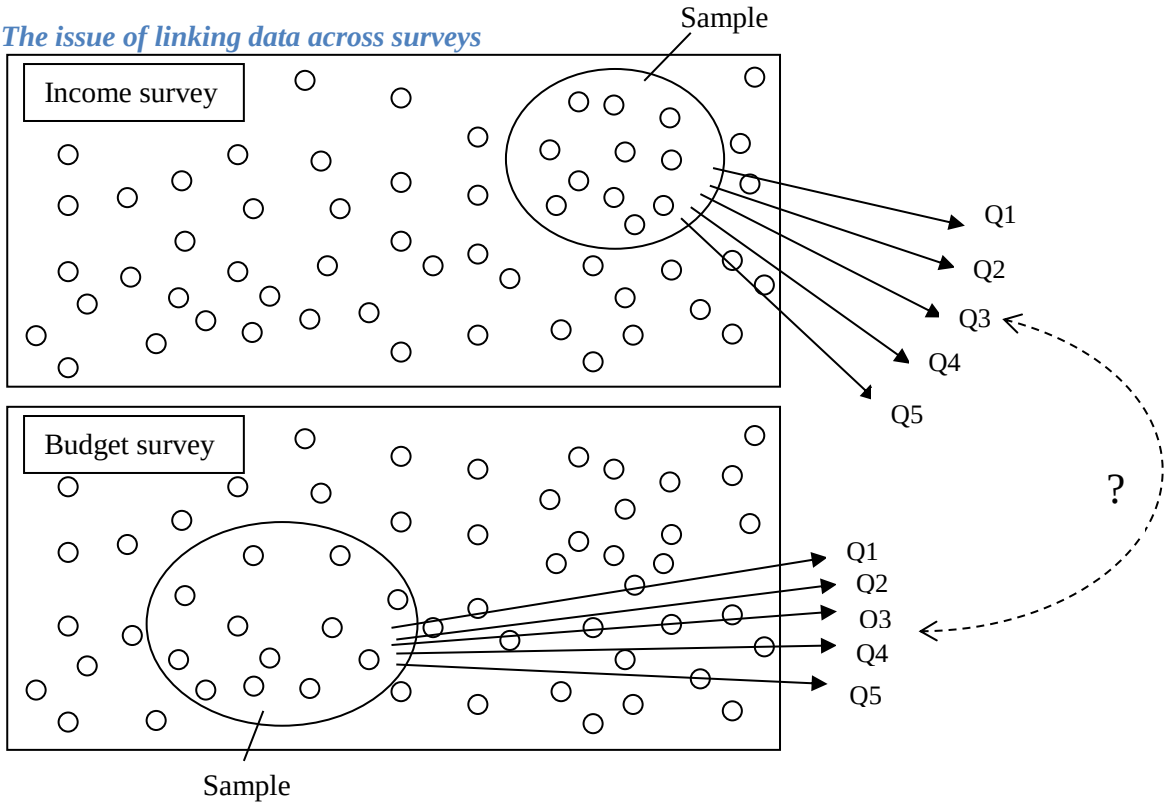
D2-D3	Taxes less subsidies on production and imports	=D2-D3
D2	Taxes on production and imports	=D21+D29
D21	Taxes on products	
D29	Other taxes on production	
D3	Subsidies	=D31+D39
D31	Subsidies on products	
D39	Other subsidies on production	
D21-D31	Taxes less subsidies on products	=D21-D31
D29-D39	Other taxes less subsidies on production	=D29-D39

ANNEX 2: Linking data across different data sources

An important step in the compilation approach is the linking of data across various data sets to construct coherent data on income, consumption and saving for the various quintiles. In many cases, data from different micro data sources are used, obtained via micro data surveys or administrative data, and the way in which they are combined may seriously affect the overall results. Linking, for example, middle income consumption data to low income disposable incomes, will lead to saving ratios for the first quintile that are too low. On the other hand, linking low income consumption to high income disposable income will lead to flawed results for higher income quintiles.

Sometimes the various data sources may describe exactly the same households, in which case it will be easy to link the data, but in many cases, it will concern different samples of households. The question then arises how these data are linked to create complete sets of accounts, to arrive at coherent distributional results for income, consumption and saving. Figure A2.1 provides a simplified example of the issue, showing a country that uses different surveys for its income and its consumption items. As it concerns surveys, different individuals may be selected in the samples. Furthermore, the sample may differ in size.

Figure A2.1: The issue of linking data across surveys



In order to arrive at reliable distributional results for income, consumption and saving, data from both data sets need to be classified into quintiles (or other household groups) in a coherent way, making sure that similar households end up in the same quintiles. Three methods have been distinguished to achieve this:

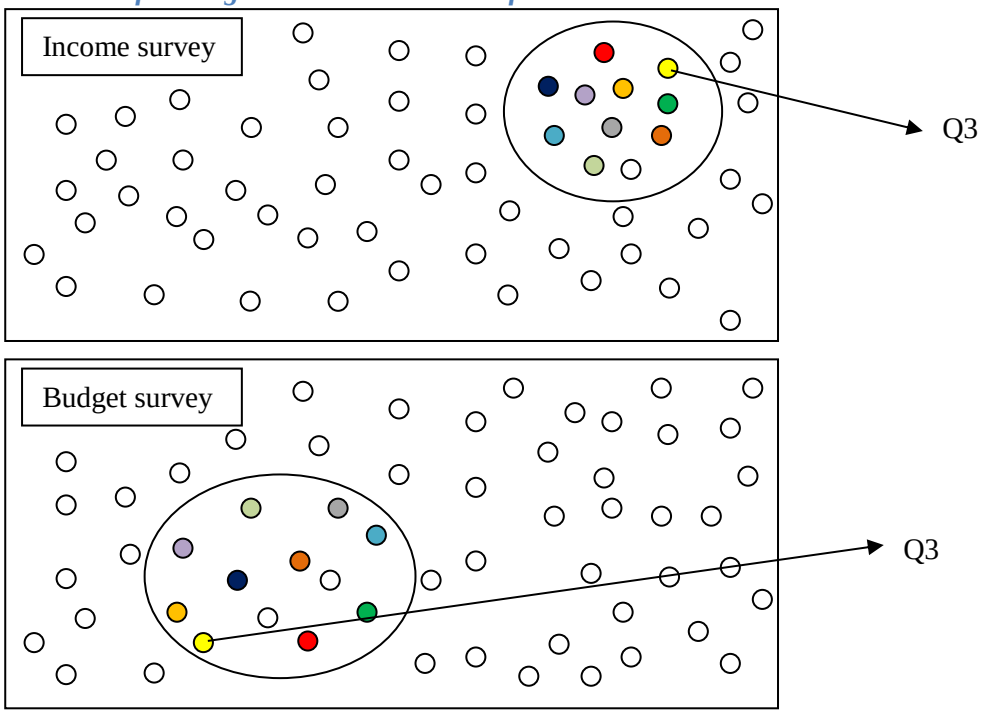
a) Link records at the start of the process on the basis of matching micro data characteristics. One can then follow the step-by-step approach and allocate records to the quintiles on the basis of their adjusted disposable income per consumption unit.

b) Impute a 'national accounts aligned disposable income' in the various data sets on the basis of common characteristics, and allocate records in these data sets according to this imputed variable.
c) Separately process the data sets and allocate records to quintiles on the basis of an income variable available in the various data sets; this implies linking data at an aggregated level at the end of the process.

A. Linking records on the basis of micro data characteristics

In the first approach, records from different data sources are linked on the basis of characteristics from the micro data results. This can be done on the basis of unique identifiers (for instance social security numbers) or on the basis of common characteristics, such as income level, and/or household characteristics (such as household size, age of the head of the household, education level and/or region). The more detail that can be used in the linking, the more accurate the ultimate linkages will be. After the linkage, results are processed according to the step-by-step approach explained in Figure A2.1, and allocated to the relevant quintiles on the basis of the disposable income per consumption unit as derived from the income results. As records have been linked at the start of the process, alignments and imputations in step 3 do not affect the linkages. Records from the other data sources will automatically follow the records they have been linked to initially. Figure A2.2 presents an example of this approach.

Figure A2.2: The issue of linking records on the basis of micro data characteristics



In Figure A2.2, matching households in both sample surveys have been given the same colours. As can be seen from the figure, it concerns different households, but they have been linked on the basis of common characteristics, for example household type and income level. In this way, the micro data can be linked across the two surveys to construct full sets of account per household (type). As an example, let's assume that the yellow record had an income level on the basis of the micro survey results that would have classified it into the second quintile, but that after alignment to the national accounts totals and imputation for missing items would be classified in the third quintile. This means that the record in the budget survey that was linked to the yellow record would now also be classified in the third quintile.

In this way, by linking all records in the various data sets to records from the income survey, data from all surveys can be allocated to the relevant quintiles.

Pros and cons of this approach

The main advantage of this approach is that households are linked on the basis of their survey results and that one can analyse the plausibility of the combined original micro data at the start of the process. If such complete records at the micro level indeed show implausible results for combinations of income and consumption data, edits may be performed before further processing the data. This editing could be done by correcting either income or consumption results, or by changing some of the characteristics that are at the basis of the linking. This approach also allows for incorporating information on the gaps between the micro aggregates and the macro results, when editing the micro data. Instead of applying a proportional allocation to close the gaps for the various items, one could thus edit those items at the micro level for which the gaps between the micro data and the relevant national accounts data are most significant. Doing so, one could also decide to allocate part of the gaps to specific (groups of) households.

The downside of this approach is that it requires identical information across the various data sources in terms of concepts and in terms of reporting. If the concepts that underlie the linkage differ, this may lead to incorrect matches. Furthermore, if the responses in one survey are deemed to be liable to reporting errors, whereas the responses in the other are not, this may also lead to incorrect matches. The latter will be particularly relevant in combing survey data with administrative data. In that case, it may be preferred to first edit the micro results before linking them across the various data sets, or to link the records on the basis of other information, not directly linked to their income levels.

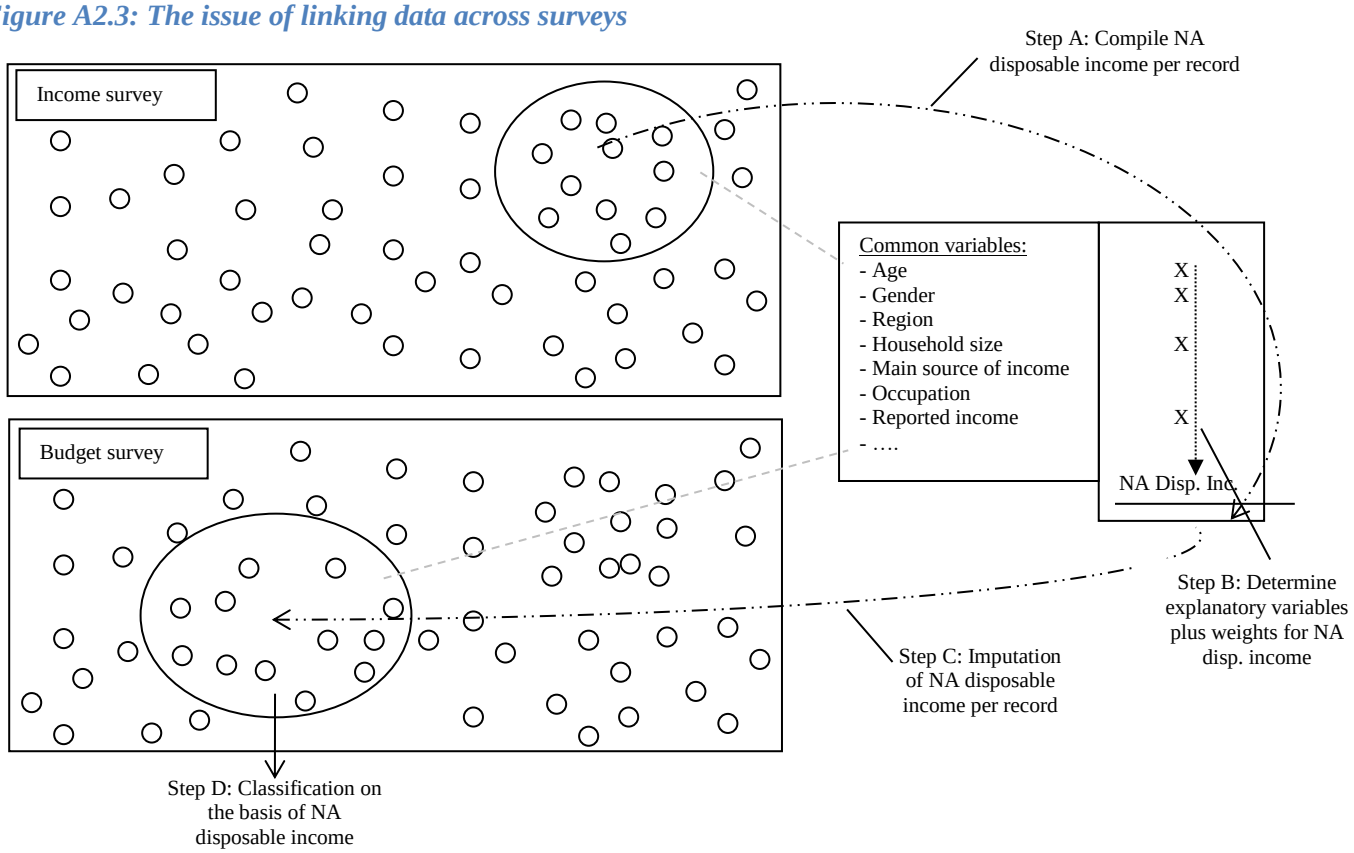
Another downside of this approach is that it may be quite complicated and resource demanding to match all results and to process any necessary adjustments that result from matching the data at a micro level. In this respect, most likely not all records will have a match in the other data sets. Even though certain margins can be applied to broaden the scope for matches, it will still be the case that certain outliers are found in the data sets for which no match can be found. In the example in Figure A2.2, this concerns one record in the income survey and two records in the budget survey. In these cases, statistical matching techniques may be needed to impute matching records in the other surveys. Another solution would be to ignore these outliers in compiling the distributional results, but this is not recommended.

It may also be the case that samples vary in size. In those cases, one record from a smaller sample may need to be linked to multiple records in the larger sample. It is understood that this affects the weighting of the data, particularly if the number of times records are used in the matching is not in line with their relative weights in the overall aggregates. As a consequence, the weights may need to be refined before compiling the aggregated results. This will most likely also be necessary if records have been imputed to create matching records for outliers in the other data sets. Initial weights will then probably have to be updated too.

B. Allocating records on the basis on an imputed income variable

The second option to arrive at a coherent distribution of households across quintiles for all data sets is to impute a disposable income in all data sets on the basis of common characteristics, via which households can be classified consistently into quintiles. This method is similar to the previous one in that households are matched on the basis of similar characteristics, but it differs in the sense that records are not matched at an individual level and that the matching is done on the basis of an imputed disposable income item. As records are not linked individually, the various steps in the methodology can be processed independently and at the final stage households can be classified on the basis of this imputed disposable income. Figure A2.3 presents a simplified example of how this technique would work.

Figure A2.3: The issue of linking data across surveys



Looking at the steps in this approach, first, in the income survey results, an income level has to be created according to national accounts definitions and in line with the national accounts totals (step A). This requires linking and aligning the relevant items from the micro survey to national accounts, and imputing for any missing items. As a result, one arrives at a 'NA aligned disposable income' per record. Subsequently, a regression analysis can be run on the basis of common variables in the various data sets to find explanatory variables to explain these disposable income levels (step B). As these variables will be used to impute an 'NA aligned disposable income' in all data sets, it is important to look at common characteristics available in all data sets. This may include 'age', 'gender', 'region', 'household size', 'main source of income', 'occupation', 'income' et cetera. The regression analysis will lead to a model that can be used to assign NA aligned disposable income levels to micro records in the other micro data sets (step C). In the final step, households in the other data sets can be classified into income quintiles on the basis of these imputed income levels (step D). The latter may be done using quintile boundaries defined on the basis of the imputed income results in the respective datasets or boundaries determined on the basis of the income part of the exercise. In the former case, one can make sure that the five quintiles consist of 20% of the households according to the results of the specific data set. However, income levels may then deviate from the ones used for the classification of households in the income part, also implying that households with similar characteristics would not necessarily end up in the same quintiles across all data sets. In the second option, the boundaries will match those used for allocating households in the income part (probably leading to a better match between income and consumption results), but as this may lead to different numbers of households per quintile for the consumption part, the weighting of the underlying micro data may have to be adjusted. In adjusting the weights, one has to make sure that all quintiles consist of 20% of the households and that the sum of the quintiles still adds up to the national accounts totals.

Pros and cons of this approach

The merit of this approach is that households with similar characteristics will be assigned similar income levels, and that it does not require a one-on-one matching of individual records. As long as the inputs in the various data sets are comparable, it will lead to a good match at an aggregated level, as a consequence of which explicit imputations for outliers[18], or linking multiple records from a smaller sample to records from a larger sample, are not necessary.

Just as was the case with the micro linking in the previous approach, this approach depends on the comparability of the reported characteristics that are used to impute disposable income. If these characteristics differ due to conceptual differences or due to specific reporting errors in one of the datasets, this may lead to incorrect imputation results. That is why one should always check which variables can indeed be considered as equivalent variables and whether some editing is required first.

Another downside is that it may not be that straightforward to run the regression, also depending on the number of common variables and the stability of the results. In respect of the latter, it would be interesting to look at the stability of the parameters over time, especially when compiling distributional results for time series' purposes. Furthermore, for countries applying this technique, it would be interesting to know which characteristics are used in the model and what the parameters look like.

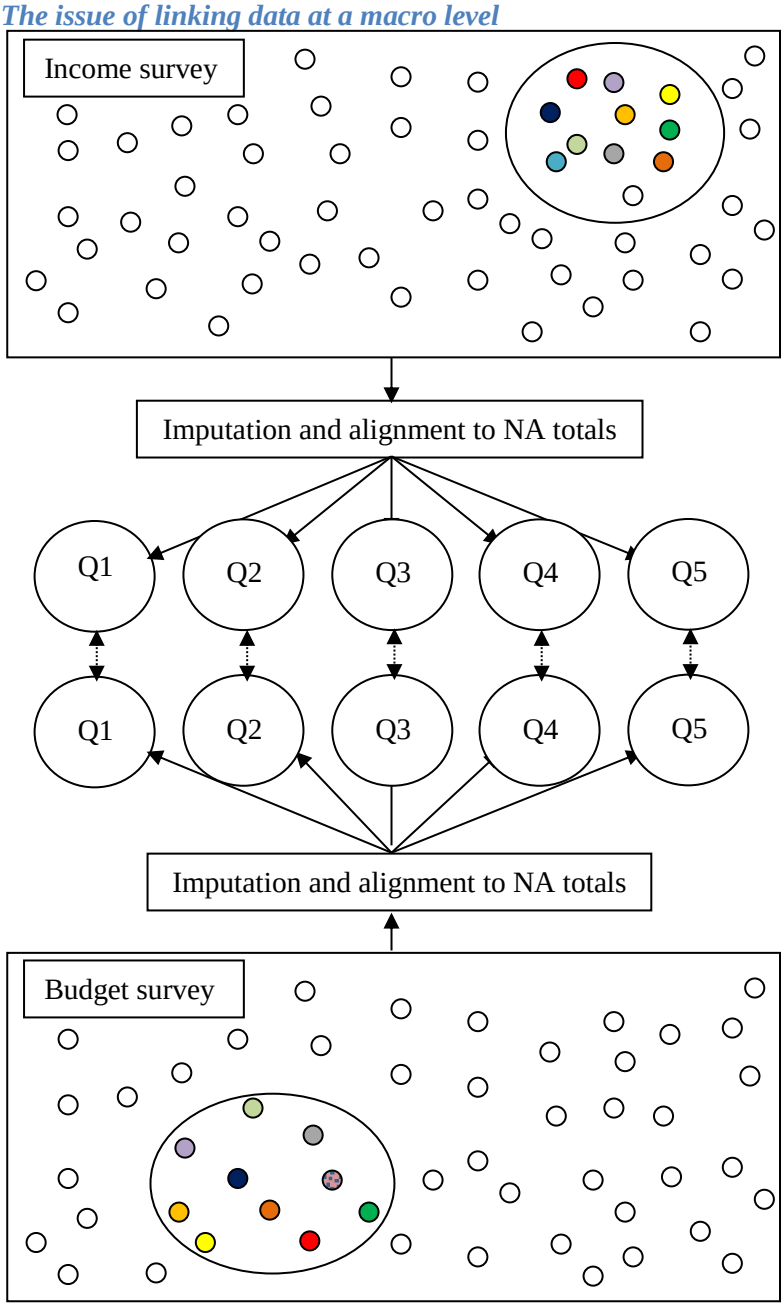
Another issue may arise if countries want to apply the income boundaries derived from the income part of the exercise for the breakdown of the consumption results. In that case, some rather complicated re-weighting may be needed to arrive at consistent results. On the other hand, if boundaries are derived on the basis of the income results from the consumption dataset, a downside may be that the consumption levels may end up not to be in line with the income levels for the same quintiles. It would be interesting to see to what extent the income boundaries for the five quintiles for the income and consumption part diverge.

C. Linking aggregated results at the end of the process

Data from different data sources can also be linked at an aggregated level. In that case, the various steps will be processed separately for the various data sets, and distributional results will only be linked in a final step, i.e. after alignment of the results to the national accounts totals and imputation for missing items. Figure A2.4 presents an example of how this would work. Survey data for both income and consumption are aligned separately and then combined at an aggregated level, most likely on the basis of their income levels.

[18] Of course, it will still be relevant to compare outliers across the various datasets to see whether in some of them some information may be missing.

Figure A2.4: The issue of linking data at a macro level



Pros and cons of this approach

The benefit of this approach is that it is less complicated and probably less time-consuming than the other two approaches. The downside is that it may lead to less reliable results as the plausibility of the results can only be checked at an aggregated level. Moreover, imputations and alignments may lead to incorrect linking of data across the two data sets. For example, a household that is represented in both the income survey and the budget survey may end up in different quintiles, if the income results from micro data sources on income are adjusted due to the alignment to the national accounts totals. If the alignment and imputation only lead to a proportional increase of all income levels and does not change the ranking of the various households, this may not be that problematic, but if they alter the ranking, this would probably lead to incorrect linking at the aggregated level.

In linking the data an aggregated level, the same question arises as with the second approach, i.e., which income boundaries should be applied for the quintile breakdown for consumption. The income boundaries that have been defined on the basis of the aligned NA results from the micro data on income can be used, or one could opt for boundaries on the basis of the income variables available from the budget survey. As the former may include several imputed items and alignments to the national account totals that may not be captured in the reported income levels in the budget survey, it would probably be best to use the latter and allocate 20 percent of the households to each of the quintiles on the basis of the reported disposable income per consumption unit in the budget survey. In that regard, it would also be interesting to see how these income levels deviate from the ones in the income surveys, and how the alignment of the consumption results would relate to the income results. If the first approach is used, income boundaries from the income part will be used to allocate households on the consumption part. This will probably not lead to an equal allocation of households across the quintiles, as a consequence of which weights will have to be adjusted. It would be interesting to see for countries that apply this technique to what extent they have to make these kinds of adjustments.

ANNEX 3: Allocating micro-macro gaps to underlying households[19]

This annex describes possible reasons that may cause the differences between the micro results and the adjusted national accounts totals. The possible reasons are related to the first three steps of the step-by-step approach, focusing on the quality of the data and of the assumptions used in the exercise:

Step 1: Adjustment of the national accounts totals:

A. The quality of the national accounts totals
B. The quality of the adjustments to the national account totals

Step 2: Linking micro data source variables to the national accounts variables:

C. Assumptions regarding the conceptual and classification differences

Step 3: Imputation for missing elements and aligning data to national accounts totals

D. The quality of the correction for the underground economy and illegal activities
E. The quality of the micro data – Estimation errors
F. The quality of the micro data – Measurement errors

The reasons for the gaps are discussed below, in accordance with the above categorisation.

A. The quality of the national accounts total

A first possible reason for the gap between the micro and macro results may be quality issues related to the national account totals. The national accounts totals are the product of a balancing framework in which data from various data sources are combined and confronted. Often, source data need to be adjusted to arrive at consistency and comprehensiveness. In that process, choices have to be made that may cause differences from the direct data sources. The quality of the data that are used in the system and the strength of the assumptions made in the balancing process will determine the quality of the final results. Gaps between micro and macro data may point to possible quality issues in this process. If gaps between micro and macro data appear, it needs to be discussed whether these can be linked back to choices that have been made in the compilation process of the national accounts totals.

B. The adjustments to the NA total

Furthermore, in the compilation process to arrive at distributional results, national accounts totals have to be adjusted to exclude NPISHs, people living in non-private dwellings, and consumption expenditure by non-residents. In some cases, specific information will be available to make these adjustments, but in other cases, these adjustments will have to be based on rather rough assumptions. Gaps between the micro and macro results may thus be due to quality issues in making these adjustments. In discussing whether these adjustments may indeed be at the cause of (part of the) gaps between the micro and macro results, the quality of the data underlying these adjustments and the strength of the underlying assumptions have to be assessed.

C. Conceptual differences and classification issues

Looking at the second step of the procedure, gaps may also appear due to conceptual differences and classification issues between micro and macro data. Sometimes the definition of the national accounts

[19] This annex is derived on the basis of a paper prepared for the 2016 IARIW conference. For more information, see Zwijnenburg (2016).

may vary from the one used in the micro survey or in the administrative data source, and (part of the) transactions may be classified differently. For instance, wages and salaries paid while on sick or maternity leave may be recorded as wages and salaries in micro sources while they are classified as social benefits in the national accounts. Furthermore, the income received by a sleeping or silent partner participating in an unincorporated enterprise is typically considered as property income by micro sources but as mixed income in national accounts. In those cases, the micro components should be reallocated or adjustments need to be made to match the SNA classification and definitions.

Also, the time of recording may differ between the national accounts totals and the micro results. The latter often focus on a certain point in time (e.g. end of the quarter or end of the year), whereas the national accounts focus on covering transactions within a certain time frame. This may also give rise to the need for adjustments, for instance to correct for changes in the population or for specific economic events that may have occurred during the period.

In analysing possible reasons for the gaps between micro and macro results, it is important to acknowledge the various adjustments that are made in the process to correct for conceptual and classification differences, and to check the validity and quality of these adjustments. If these adjustments are based on actual data, it is assumed they will be of better quality then when they are based on assumptions.

D. The correction for the underground economy, illegal activities and other missing elements

The gaps may also be related to the third step in the compilation process. First, imputations are made for the underground economy, illegal activities and other elements that may be missing from the micro data. The underground economy and illegal activities are usually not covered by micro data. Here, the underground economy relates to activities that in principle are legal, but are deliberately concealed from public authorities. Illegal activities are forbidden by law or become illegal when carried out by unauthorised persons. As they represent economic significance, they are included in the national accounts and therefore may partly explain the gap between the micro data and the national accounts totals.

In addition to the underground economy and illegal activities, some other elements may be missing from the micro data as well. This may not only concern specific items (such as imputed items like FISIM), but may also relate to specific groups. Administrative data may for instance exclude some groups of households that fall below certain thresholds or that are exempted from tax, and survey data may for instance lack information on unincorporated enterprises, which are included in the household sector in the national accounts.

If data on specific groups or items are missing, imputations are needed to align the data with the national accounts definitions. When no adjustments have been made so far, this may be one of the reasons for the gaps between the micro and macro data. If adjustments have already been made, one will need to look at the quality of these adjustments to see whether these are accurate or may explain part of the remaining gap between the micro and macro data. The quality of the adjustments will depend on the available information underlying the adjustment or the assumptions that have been used for this purpose.

Estimations for the underground economy, illegal activities and other missing elements are usually made within the national accounts division. These estimates are based on information regarding the coverage of the micro data sources and on information that may be available on specific missing activities. Usually, part of these estimates is also based on assumptions. Therefore, in checking possible reasons for the gaps between micro and macro data, it is important to analyse the reliability of these specific estimates. Specific information from the estimation methods in national accounts may also be helpful in allocating the gaps to specific household groups.

E. The quality of the micro data – Estimation errors (sampling and coverage)

Micro estimates may also be subject to quality issues. This usually concerns estimation and measurement errors. The first category consists of sampling and coverage errors, related to the weighting of the sample data to arrive at estimates for the target population. Measurement errors concern the accuracy of the micro data reported in the surveys.

Estimation errors affect the extrapolation from the data from the survey samples to the target population. These relate to the sample size, the representativeness of the sample and the magnitude of the non-response. The errors related to the sample size are referred to as the standard sampling error, implying that the smaller the survey sample, the larger the sampling error, as less data underlie the ultimate estimates. The other two issues are referred to as coverage errors. These occur in the case of the sampling frame being different from the target population and in the case of selective non-response. All these aspects may lead to higher variability of the weighted estimates and therefore to possible gaps with the macro results. Especially survey data may suffer from estimation errors. Administrative data sources tend to have broad coverage and are therefore less prone to these kinds of errors. However, it has to be realised that administrative data may also suffer from non-random under-coverage, especially in the case of reporting thresholds.

In case of gaps between micro and macro estimates, it may be advisable to analyse to what extent estimation errors may be responsible for these gaps. As administrative data are assumed to have a broad coverage, they are often used to check whether survey data suffer from some kind of estimation error. For this purpose, also data from other data sources are often used. Vermeulen (2014), for example, uses the Forbes list of extremely wealthy to check for the coverage of the very wealthy people in wealth survey micro data. Coli and Tartamella (2014) describe how they used the confrontation of data from various data sources to detect non-registered workers. Other research has looked at the distribution of households in the tails of the datasets (see Törmälehto (2014)).

F. The quality of the micro data – Measurement errors

Errors may also occur due to mistakes in the data reported in the surveys. These are referred to as measurement errors, and may relate to item non-response or the reporting of incorrect data. Measurement errors may be caused by misinterpretation of the questions, problems with recalling the exact numbers or willingly providing wrong information. Fessler et al. (2012) point out that it may also be affected by the data collection method. Meyer, Mok and Sullivan (2008) show that a lot of statistics have to deal with measurement error and that these kinds of errors have increased over time, at least for some specific items. Particularly questions on income are usually understood to be relatively sensitive and prone to higher non-response rates or larger measurement errors.

To check the validity of micro results, some countries have confronted data from micro surveys with administrative data or with data from other data sources. For instance, when France shifted from an interview to a register-based income recording, they experienced a substantial increase in the share of property income for the top 5 per cent from 2006 to 2007 (see Törmälehto (2014)), implying an underreporting for and/or under-coverage of the high-income class in the survey samples. In confronting data from different data sources, one of course has to be aware that other data sources may also cope with measurement errors. However, it is assumed that, apart from the underground economy and illegal activities, most administrative data sources are less prone to measurement errors, due to the fact that these data are often derived from administrative systems and do not involve a lot of manual processing. Furthermore, the legal sanctions that are associated with the accuracy of tax reporting also make them less vulnerable to measurement errors.

In checking the plausibility of the micro data, some countries have also confronted income and consumption data on a micro-economic level. Accardo et al. (2009) and Sabelhaus et al. (2012) describe the situation that households declared incomes that were too low compared with their consumption expenditure (excluding exceptional purchases) without indicating that they had to significantly reduce their assets or increase their liabilities. In those cases, the incomes of these households were adjusted upwards to bring them in line with the level of their consumption expenditure. Data on wealth may also be taken into consideration, to confront the savings ratio from the financial and from the income and consumption side. Other research has checked the micro data on the basis of other micro data within the same survey, for instance by confronting micro data within groups that share the same characteristics, and then look at outliers. Cifaldi and Neri (2013) describe how this was applied to check the plausibility of the reported income for self-employed in Italy. Finally, D'Alessio and Neri (2015) describe the use of interviewers' score (an assessment by the interviewer of the reliability of respondents' answers to the various questions) to correct for part of the gaps between micro and macro data.

Allocation of the gaps to the relevant households

After the gaps have been attributed to probable causes, the related estimates have to be allocated to the relevant households or household groups. As the allocation will differ per cause, this detailed approach may indeed lead to more accurate results. For all causes that concern micro data underlying the distributional results, specific solutions have to be found. Looking at the reasons that were presented in the previous section, this relates to categories C to F. Categories A and B only concern the adjusted national accounts totals with which the micro data have to be aligned, so they only affect the distributional results in an indirect way.

It obviously depends on the cause, what kind of information is available to properly allocate the relevant amounts. Sometimes the micro data itself may already show what kind of information is missing, for instance in the case that the combination of income and consumption data show implausible results for certain households. In other cases, countries may rely on results from recent research into specific causes, for instance the confrontation of survey data with administrative data or with alternative data sources (see also the examples in the previous section). Alternatively, the allocations may need to be based on (rather rough) assumptions. Although this will be suboptimal in comparison with the other two options, it is assumed that it leads to more reliable results as compared to simply applying a proportional allocation. Literature provides a lot of information that may be used to come up with a proper allocation for each of the underlying causes. When this information is used in combination with the findings from the analysis of the micro data used in the exercise, it leads to more transparent and better informed adjustments per underlying cause.

When looking at literature, relevant information can be found on various causes. In addition to the literature that was already mentioned in the previous section, Carson (1984), for example, provides information on which household types are most likely to be involved in the underground economy or illegal activities. She shows that it is mostly linked to younger people, in higher and lower (rather than middle) income groups, self-employed and people with higher education. With regard to estimation errors, literature often points to the difficulty of capturing high-income households in surveys, partly due to the high concentration of income in the top percentiles and the reluctance of the highest income households to participate in surveys (see D'Alessio and Faiella (2002) and D'Alessio and Neri (2015)). Finally, with regard to measurement errors, literature often shows that lower-income households tend to over-report their income, and that the opposite is true for higher-income households (Lohmann (2010) and Romanov and Gubman (2012)). The latter is confirmed by research by D'Alessio and Faiella (2002) and Sabelhaus et al. (2012). Studies show that these errors are also often correlated with characteristics as age, educational level and social status. Of course, it depends on the items, the data sources and the

country characteristics to what extent these results from literature will apply to specific country cases, but it may provide insights in how to approach the allocation question and may help in coming up with the most appropriate allocation for the specific purposes.

The allocation of the amounts to the underlying households should ideally be done at the level of the micro statistics, i.e. by making adjustments to the survey data, applying imputations at the micro level, or by adjusting the survey weights to arrive at the relevant aggregates. This will lead to improved micro data that underlie the new distributional measures and will make sure that the quintile classification is re-adjusted on the basis of these improved data. However, as this may be time-consuming, especially in the case that the process of compiling the micro results has already been finalised, an alternative is to allocate the amounts at a meso-economic level. In that case, the quintile allocation on the basis of the 'unadjusted' micro data is taken as starting point and the amounts that have been attributed to the various causes are allocated to the quintiles. It is clear that the distributional results on the basis of this meso approach will not be as accurate as in the case of processing the corrections on the micro-economic level. However, for the purpose of a quick overview of the possible impact of the allocation of the amounts related to the various causes for micro-macro-gaps, it may provide quick insights.

ANNEX 4: The 2010 health care cost by age groups (on a per capita basis, as a percentage of GDP per capita)

Table A4.1: Public expenditures per age group, per capita 2010

As a percentage of GDP per capita

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95+
Australia	1.7	1.1	0.9	1.1	1.2	1.4	1.6	1.7	1.8	2.0	2.2	2.6	3.3	4.1	5.0	6.1	7.1	7.8	7.9	7.1
Austria	1.8	1.2	1.0	1.1	1.3	1.6	1.7	1.8	2.0	2.1	2.4	2.9	3.5	4.4	5.4	6.6	7.7	8.4	8.6	7.6
Canada	1.4	0.9	0.8	0.9	1.1	1.2	1.3	1.4	1.5	1.7	1.9	2.2	2.7	3.4	4.2	5.1	5.9	6.5	6.6	5.9
Denmark	1.7	1.1	1.0	1.1	1.3	1.4	1.6	1.7	1.8	2.0	2.2	2.6	3.3	4.1	5.0	6.1	7.1	7.8	7.9	7.1
France	1.2	0.8	0.7	0.7	0.9	1.0	1.1	1.2	1.3	1.4	1.6	1.9	2.3	2.9	3.5	4.3	5.0	5.5	5.6	5.0
Germany	1.7	1.1	1.0	1.1	1.3	1.5	1.6	1.7	1.8	2.0	2.3	2.7	3.3	4.1	5.1	6.1	7.2	7.9	8.0	7.1
Israel	1.7	1.1	0.9	1.1	1.3	1.4	1.6	1.7	1.8	2.0	2.2	2.6	3.2	4.0	5.0	6.0	7.0	7.7	7.8	7.0
Italy	2.2	1.4	1.2	1.4	1.6	1.9	2.1	2.2	2.4	2.6	2.9	3.5	4.3	5.4	6.6	8.0	9.3	10.3	10.4	9.3
Japan	2.1	1.4	1.2	1.3	1.6	1.8	2.0	2.1	2.3	2.5	2.8	3.3	4.1	5.1	6.3	7.6	8.8	9.7	9.9	8.8
Korea	1.3	0.8	0.7	0.8	1.0	1.1	1.2	1.3	1.4	1.5	1.7	2.1	2.5	3.2	3.9	4.7	5.5	6.1	6.2	5.5
Mexico	1.6	1.0	0.9	1.0	1.2	1.3	1.5	1.6	1.7	1.8	2.1	2.5	3.0	3.8	4.7	5.7	6.6	7.3	7.4	6.6
Netherlands	1.3	0.8	0.7	0.8	0.9	1.1	1.2	1.3	1.4	1.5	1.7	2.0	2.5	3.1	3.8	4.6	5.4	5.9	6.0	5.4
New Zealand	2.2	1.4	1.2	1.4	1.6	1.9	2.1	2.3	2.4	2.6	2.9	3.5	4.3	5.3	6.6	8.0	9.3	10.2	10.4	9.2
Portugal	2.1	1.3	1.2	1.3	1.5	1.7	1.9	2.1	2.2	2.4	2.7	3.2	3.9	4.9	6.1	7.4	8.6	9.4	9.6	8.5
Slovenia	1.2	0.8	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.6	1.9	2.3	2.9	3.6	4.4	5.1	5.6	5.7	5.1
Sweden	0.3	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.6	0.8	1.0	1.2	1.4	1.5	1.6	1.4
Switzerland	1.7	1.1	1.0	1.1	1.3	1.4	1.6	1.7	1.8	2.0	2.2	2.7	3.3	4.1	5.1	6.1	7.1	7.8	8.0	7.1
Turkey	3.8	2.4	2.1	2.3	2.8	3.2	3.5	3.8	4.0	4.4	4.9	5.8	7.2	9.0	11.1	13.4	15.6	17.2	17.5	15.6
United Kingdom	1.7	1.1	0.9	1.0	1.2	1.4	1.6	1.7	1.8	1.9	2.2	2.6	3.2	3.9	4.9	5.9	6.9	7.6	7.7	6.9
United States	0.5	0.3	0.3	0.3	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.8	1.0	1.3	1.6	2.0	2.3	2.5	2.5	2.3

Source: OECD-ECO directorate – Estimated survivor's cost curve.

For more details information on the method please read the Working paper No 477: "Projecting OECD Health and Long-Term Care Expenditures: What Are the Main Drivers?"
http://www.oecd.org/tax/publicfinanceandfiscalpolicy/36085940.pdf

Table A4.2: Public expenditures per age group, per capita 2010

Expressed in 2005 constant PPP US$

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95+
Australia	692.1	442.7	386.1	428.5	507.2	585.1	646.9	693.6	738.5	802.1	907.6	1076.2	1322.3	1649.2	2043.9	2472.6	2876.5	3166.2	3218.0	2868.5
Austria	736.7	471.2	410.7	455.7	539.1	621.9	687.5	737.2	785.1	852.9	965.4	1145.0	1407.1	1755.1	2175.1	2631.4	3061.1	3369.4	3424.6	3052.9
Canada	556.6	356.0	311.0	346.0	410.1	473.4	523.4	560.9	596.7	647.4	731.7	866.7	1064.3	1327.0	1644.4	1989.4	2314.5	2547.8	2589.3	2307.4
Denmark	671.9	429.7	375.1	416.7	493.5	569.5	629.6	674.9	718.3	779.8	882.0	1045.4	1284.1	1601.3	1984.4	2400.8	2793.0	3074.4	3124.6	2784.9
France	406.1	259.8	226.7	251.9	298.3	344.3	380.6	408.0	434.3	471.4	533.1	631.9	776.2	967.9	1199.5	1451.2	1688.2	1858.3	1888.7	1683.3
Germany	640.5	409.7	357.7	397.7	471.1	543.7	601.1	644.3	685.6	744.1	841.3	997.0	1224.5	1526.8	1892.1	2289.1	2663.1	2931.4	2979.3	2655.1
Israel	481.3	307.8	269.8	301.2	357.7	413.3	456.9	489.4	520.0	563.2	635.3	751.4	921.8	1148.7	1423.3	1722.0	2003.7	2205.8	2241.6	1996.4
Italy	710.1	454.3	394.2	434.9	512.9	590.8	653.2	701.1	748.1	814.9	924.9	1099.6	1353.4	1689.4	2094.1	2533.1	2946.1	3242.5	3296.0	2940.6
Japan	716.5	458.3	399.8	443.9	525.5	606.3	670.3	718.6	765.0	830.8	939.9	1114.3	1369.1	1707.4	2116.0	2559.9	2978.0	3278.0	3331.6	2969.6
Korea	383.9	245.5	214.3	238.2	282.1	325.6	359.9	385.9	410.6	445.7	504.0	597.3	733.7	914.9	1133.8	1371.7	1595.8	1756.6	1785.3	1591.1
Mexico	241.9	154.7	135.0	149.8	177.4	204.7	226.3	242.6	258.3	280.5	317.3	376.2	462.2	576.4	714.3	864.2	1005.3	1106.6	1124.7	1002.5
Netherlands	547.9	350.5	305.2	338.2	399.9	461.2	509.8	546.8	582.5	633.2	717.1	850.9	1046.1	1305.0	1617.3	1956.6	2276.0	2505.2	2546.3	2270.3
New Zealand	665.7	425.7	372.4	414.9	492.1	568.2	628.2	673.1	715.7	776.0	876.4	1037.6	1273.6	1587.6	1967.3	2380.2	2769.3	3048.5	3098.1	2760.2
Poland	517.5	331.0	288.8	320.8	379.8	438.3	484.5	519.5	553.0	600.4	679.1	805.1	989.0	1233.4	1528.5	1849.2	2151.3	2368.0	2406.7	2145.1
Portugal	528.3	337.9	294.8	327.4	387.6	447.3	494.5	530.1	564.3	612.8	693.2	821.8	1009.6	1259.1	1560.3	1887.7	2196.0	2417.2	2456.8	2189.8
Slovenia	330.7	211.5	184.2	204.0	241.2	278.1	307.5	329.8	351.4	382.0	432.7	513.5	631.3	787.5	976.0	1180.8	1373.5	1511.8	1536.6	1370.1
Spain	1037.8	663.9	578.6	641.7	759.2	875.8	968.2	1038.2	1105.7	1201.3	1359.8	1612.9	1982.2	2472.4	3064.1	3706.9	4312.2	4746.5	4824.3	4300.8
Sweden	130.2	83.3	72.6	80.6	95.3	109.9	121.5	130.3	138.8	150.8	170.7	202.4	248.8	310.3	384.5	465.2	541.1	595.6	605.4	539.7
Switzerland	800.9	512.3	446.8	495.9	587.0	677.2	748.7	802.8	854.7	928.2	1050.3	1245.3	1530.1	1908.3	2365.0	2861.1	3328.4	3663.7	3723.6	3319.2
Turkey	575.6	368.2	321.0	356.3	421.7	486.4	537.8	576.6	614.0	666.9	754.7	894.9	1099.7	1371.5	1699.8	2056.3	2392.2	2633.1	2676.2	2385.6
United Kingdom	594.1	380.0	331.9	369.1	437.3	504.8	558.0	598.1	636.4	690.6	780.7	925.0	1136.0	1416.4	1755.2	2123.5	2470.5	2719.5	2763.8	2463.0
United States	254.6	162.9	141.7	156.9	185.4	213.8	236.3	253.5	270.1	293.8	332.9	395.2	485.9	606.3	751.4	909.0	1057.3	1163.8	1182.9	1054.8

Source: OECD-ECO directorate – Estimated survivor's cost curve.

ANNEX 5: Example how scaling at different aggregation levels may lead to different results

The example assumes an aggregate item that consists of two underlying sub-items with different allocations across quintiles according to the micro data, and with different micro-macro gaps. Table A5.1 shows the initial situation.

Table A5.1: Distribution across quintiles according to micro data

	Q1	Q2	Q3	Q4	Q5	Micro total	Macro total
Sub-item 1	20	20	20	20	20	100	200
Sub-item 2	0	0	0	0	100	100	160
TOTAL	20	20	20	20	120	200	360

As a gap exists between the micro and the NA total, this needs to be bridged in the process. However, one has to bear in mind that if this is done on the basis of proportionally adjusting the micro data, it will make a difference whether this is done at the detailed level or at the aggregated level. Tables A5.2 and A5.3 present the results of the alignment according to applying it at the detailed and at the aggregated level, respectively.

Table A5.2: Distribution after alignment at the detailed level

	Q1	Q2	Q3	Q4	Q5		Macro total
Sub-item 1	40	40	40	40	40		200
Sub-item 2	0	0	0	0	160		160
TOTAL	40	40	40	40	200		360

Table A5.3: Distribution after alignment at the aggregated level

	Q1	Q2	Q3	Q4	Q5		Macro total
Sub-item 1	-	-	-	-	-		-
Sub-item 2	-	-	-	-	-		-
TOTAL	36	36	36	36	216		360

The tables show that as the distribution and the micro-macro gaps differ for the two items, the alignment at the detailed level will lead to different results than applying the alignment at the aggregated level. As it is assumed that compilation at the more detailed level will lead to more accurate results, countries are encouraged to apply the calculations at these more detailed levels and derive the aggregates on the basis of that.

ANNEX 6: Examples of correct and incorrect reporting of missing data and zero values

Example 1: Missing numeric value, single-option solution

Wrong: a cell is empty where the value of "10" should have been reported.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	10
D122R	Employers' imputed social contributions (*counterpart in D612*)	

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R holds.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	10
D122R	Employers' imputed social contributions (*counterpart in D612*)	10

Wrong: a cell is empty where the value of "10" should have been reported.

D1R	**Compensation of employees**	
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAP
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAP

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R holds.

D1R	**Compensation of employees**	10
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAP
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAP

Example 2: Missing non-numeric value, single-option solution

Wrong: a cell is empty where "NAV" should have been reported.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R cannot hold because of "NAV" entries.

D1R	**Compensation of employees**	NAV
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAV
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Example 3: Missing numeric or non-numeric value, double-option solution.

Wrong: a cell is empty where either a numeric value greater than "10" or "NAV" (according to country-specific circumstances) should have been reported.

D1R	**Compensation of employees**	
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAV
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Correct: all the cells are filled in and the identity D1=D11R+D121R+D122R cannot hold because of "NAV" entries.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAV
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R cannot hold because of "NAV" entries.

D1R	**Compensation of employees**	NAV
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	NAV
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Example 4: Misplaced non-numeric value and missing zero or non-numeric value, double-option solution.

Wrong: a cell reports a wrong value where either "NAP" or the value of "0" (according to country-specific circumstances) should have been reported.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAV

Wrong: a cell is empty where either "NAP" or the value of "0" (according to country-specific circumstances) should have been reported.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R holds.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	0

Correct: all the cells are filled in properly and the identity D1=D11R+D121R+D122R holds.

D1R	**Compensation of employees**	30
D11R	Wages and salaries	10
D121R	Employers"actual social contributions (*counterpart in D611)*	20
D122R	Employers' imputed social contributions (*counterpart in D612*)	NAP