Microdata Services

Guidelines for output from the microdata environment of CBS

Version: October 2025

Contents

1. Introduction and background

It is vital that data is properly secured and protected. Within Statistics Netherlands (CBS), our data is secured both physically and legally, and that also applies to the use of data derived from our microdata environment. Security is achieved by shielding the CBS microdata environment, requiring users of the CBS microdata environment to sign a contract with Statistics Netherlands, and applying the principles of statistical data security to prevent the disclosure of individual-level data (i.e. data that could reveal further information about a particular statistical unit). These guidelines relate primarily to statistical security, i.e. preventing accidental disclosure.

All use of the CBS microdata environment is framed by a number of articles in the Statistics Netherlands Act (*Wet op het Centraal Bureau voor de Statistiek*).

Statistics Netherlands Act (translated and abridged)

Section 37

The data collected by Statistics Netherlands is solely intended for statistical purposes and will not be provided to any persons other than those charged with carrying out the duties of Statistics Netherlands.
The data will only be published in ways that preclude the identification of individual persons, households, companies or institutions.

Section 41

Contrary to the provisions of Section 37 the director general may, on request, provide or grant access to a data set to a department, organisation or institution as referred to in the second subsection for the purposes of statistical or academic research, provided appropriate measures have been taken to prevent the identification of individual persons, households, companies or institutions from that data.

Section 42

The director general will only grant a request as referred to in Section 41 if the director general considers that the applicant has taken adequate measures to prevent the data set being used for purposes other than statistical or academic research.

(see appendix 3 for the full text)

In addition to these legal obligations, Statistics Netherlands also needs to be seen by the outside world as a fully reliable organisation. Survey response rates depend to a large extent on the knowledge that individual-level data will be secure. After all, who would offer personal information to an institution that did not handle it with the utmost care?

Some initial comments:

- The prevention of disclosure involves a shared responsibility on the part of Statistics Netherlands and researchers granted permission to work with microdata[1]. The confidentiality statement also makes this clear.

[1] *Microdata: data based on surveys and records on individual entities, such as individuals, companies and institutions.*

- The output checkers assess each output separately for the risk of disclosure. It is the responsibility of the researcher to check whether, in combination with previously released output (regardless of the project or institution concerned), a new output could lead to disclosure. The output checkers do not check this.
- The guidelines apply to datasets on individuals, households, companies and other institutions such as schools and hospitals.
- For a given project, a researcher can only access the files that are necessary to answer the research questions. The outputs provided are subject to purpose limitation and must therefore relate closely to the research question.
- If researchers upload their own files to the CBS microdata environment, the relevant output continues to be subject to the output guidelines of Statistics Netherlands.
- In addition to the output guidelines, there are also publication guidelines for some Statistics Netherlands micro-datasets. These can be found in the documentation report for the relevant datasets. The output checkers do not check the publication guidelines. It is the responsibility of the researcher to ensure that the output is published according to the publication guidelines.

In the following sections, three elements of the guidelines are discussed. Section 2a presents the *confidentiality* guidelines which relate to preventing disclosure. Section 2b focuses on the *readability* of the output, so that it can be checked by output checkers easily. Section 2c sets out the *technical* guidelines.

2. Guidelines

2a. Guidelines relating to disclosure

This section of the guidelines clarifies which output is permitted, and how it is assessed for the risk of disclosure.

The standards applied err on the side of caution; this is a deliberate choice, so that certain security risks can be virtually excluded altogether. If less strict guidelines were applied, the researcher would need to provide significantly more background information with each output to allow for careful review.

If a researcher wants output that does not meet the guidelines, the burden of proof is primarily on the researcher, who needs to demonstrate that there is no disclosure risk. In such cases, explicit permission for the exception must be forthcoming before offering the relevant output.

We apply the following confidentiality guidelines:

0. **No individual-level data**: It is not permitted to offer microdata as output. Points to note:
    - output may not include any variables that can be linked to individuals, such as RINPERSOON, RINOBJECTNUMMER, BEID, BRIN.
    - the names of companies, schools, institutions etc. may not be included in the output;
    - figures relating to one institution, even if anonymised, are not permitted**.**

1. **Minimum number of observations:** All tabular and similar output should include **at least ten units** (unweighted) underlying any cell or data point presented. Values of less than ten should be suppressed, and it must not be possible to infer these values using retrospective calculation. Be aware of retrospective calculation using totals for rows and columns, as well as totals for other tabs, different regional distributions, and partly rounded numbers. It is also important to remember that some cell totals are larger than the sum of the different categories because researchers may have omitted values below 10.
   Further notes regarding the above:
    - When data refers to information at the level of individual persons as well as companies/institutions/households, this applies to both the minimum number of persons and the number of companies/institutions/households.
    - Values of 0 should be suppressed, just as values between 1 and 9. Many statistical programs omit categories with no observations from output; remember this when creating output. One exception is zeros, which by definition cannot appear in the selection: an explanation of why the value is zero on logical grounds, and therefore no information is revealed, should be provided in the output form.
    - If the tables contain percentages or weighted values, the unweighted values must be included so that this standard can be checked.
    - A value of less than 10 is acceptable under *missing/unknown* without any further breakdown, because ‘unknown’ is not traceable to specific individuals. The minimum number of observations does apply to the *other* category because this is an aggregation of other (recognisable/traceable) categories.
    - *Maximum and minimum* values are not released because they usually refer to one specific unit. Furthermore, they indicate that lower or higher values do not occur within a given group. Situations where such extreme values are fixed, logically, and therefore pose no disclosure risk, are an exception; for example, the minimum or maximum wage of non-workers, which is zero by definition.
    - *Percentile values* must be at least 10 data points away from the minimum and maximum. This means that the median may be shown for a population of 20 or more

persons or institutions but data on the 10th decile, for example, may only be shown for a population of 100 people or institutions or more.

Additional guidelines for models:

2. **Models:** Models: all modelled output should have **at least 10** degrees of freedom and at least 10 units must have been used to produce the model. Degrees of freedom = the number of observations minus the number of parameters and/or restrictions. Furthermore, models may not be based on a single unit, so a time series based on a single company is not permitted.
    - This is particularly important for regression models, higher order moments (variance, covariance, kurtosis, Skewness), summary and test statistics, estimates (R^2 and variations thereof), estimated variance, information criteria (e.g. AIC, BIC) and individual and group testing and statistics (ego t, F, chi-square, Wald, Hausman)
    - Residuals and graphs/plots of residuals are treated as data regarding individuals and will therefore not be released.
    -

Additional guidelines for frequency tables[2]:

3. **Group disclosure:** In all tabular frequency tables and similar output, individual cells must not contain more than **90% of the total** number of units in the relevant row or column in order to prevent group disclosure. Group disclosure is a situation where some variables in a table define a group of units while other variables in the table reveal information that applies to every member of the group. Even though no individual unit can be identified, confidentiality is breached because the information applies to (almost) every member of the group and therefore the group is identifiable.
    - Group disclosure also occurs when it can be inferred that every person in the group has the same characteristic - because the given standard deviation for that characteristic is 0, for example. Despite the presence of sufficient respondents, information can still be traced back to the individual in such cases.

Additional guidelines for quantitative tables[3]:

4. **Dominance:** Dominance plays a role in output aggregated to (for instance):
    - Companies/institutions
    - Schools
    - Hospitals
    - Housing corporations

    When aggregating to one of these categories, it is important that dominance does not take place. This means that the largest contributor to a cell should not contribute more than 50% to the total for that cell. The researcher must provide these numbers; if they fail to do this, they may be requested to do so.
    - Dominance can also take place when these terms are not used directly. For example, when aggregating by SBI classification, this also involves companies. The same is true for tables of numbers of vacancies with companies underlying, and for certain specialisms with hospitals underlying.

Appendix 1 provides more detail on the above guidelines for each type of output and illustrates them on the basis of examples.

2 Frequency table: table in which you can read how often (frequency) a given observation occurs. The percentage mentioned is stricter than the standard that Statistics Netherlands applies itself because the disclosure risk depends on what is published.
3 Quantitative table: a table in which the cell values are the sum of a variable across all observations. For example, wages and turnover, but also the number of persons employed and number of vacancies (for the latter two, companies are the units of observation).

2b. Guidelines relating to readability

The principle underlying the following guidelines is that researcher should ensure that the output submitted in combination with the standard output check form is clear and comprehensible for the output checkers.

The output is checked by two (objective) checkers. This ensures the independent verification of output, with the same criteria applied to all researchers. Although the checkers are experienced, they are generally unfamiliar with the content of the projects. The wide range of subjects also means that the checkers do not know the details of all the datasets. This means that it is very important that the researcher provide a clear explanation of the analysis and the variables presented. The researcher can therefore influence the speed and outcome of the output check.

We apply the following guidelines:

1. Output files should be accompanied by a fully and correctly completed **Standard output check form**. This form can be found in the 'shared workspace' (H drive) or under 'metadata' (K drive) in the 'Utilities' folder. It must be included with each output that is submitted for checking.
   A fully completed output form can be found in appendix 2 as an example.
   The form must provide the following information:
   - whether the output complies with the confidentiality guidelines and, if not, why there is no disclosure risk;
   - the purpose of the analysis;
   - the source files used;
   - how the analysis file was compiled (for example, by linking different files or by selecting subpopulations);
   - population on which the output is based;
   - the type of output; e.g. counts, crosstabs, or statistical analysis;
   - the relationship between tables in an output file (where applicable);
   - the relationship between figures and the underlying models in an output file (where applicable);
   - the relationship with output previously submitted (where applicable).
2. Variables in the output must be **documented**. What we mean by this, is that there should be a clear definition of the meaning of the variable and individual values in the case of discrete variables. One possibility is to create a separate document containing this data. This document must be included with each output and can always be supplemented with new variables.
3. Each component of the output must be preceded by a line of text which describes the **population or selection** for that component.
4. In situations involving *dominance* and other situations where the information to be checked involves a risk of disclosure, the underlying tables that cannot be released must be included. In such cases, these must be presented in a **separate zip file** with the name **"NOTOUTPUT_Backgroundinformation.zip"**. In this document must be clearly described which part of the background information is linked to which part of the output. For example, use identical numbers for the tables used in the output and the background document. Of course, this zip file will not be released.

2c. Guidelines on technical requirements

In addition to security and readability, there are also a number of technical requirements for the output to be delivered. These guidelines are designed to ensure that the verification process is efficient and reliable, and to prevent any technical problems when processing files from the CBS microdata environment.
The guidelines below indicate the size, file format and structure of output files. Follow these technical instructions carefully in order to facilitate output control and prevent delays in processing.

1. Size of output: In order to minimise the time needed to check an output, outputs cannot be too large. We apply the following guidelines:
   - The goal is that an output checker will check the output within one hour. The tariffs for output checks are also based on this. If an output is too large, the costs and the time for checking will be increased accordingly. In such cases the researcher will be asked to approve the costs first.
   - Do not submit zip files or zipped folders.
2. Format: Please use the following formats when offering output:
   - Tables in: xlsx, csv or xml format;
   - Scripts in: do, R, py or sps format;
   - Documentation in: txt format;
   - Figures in jpg, png or gif format.
   - We do not allow presentations in .ppt and .pptx format (PowerPoint) and data files with the extension '.RData'. Outputs other than the above - such as '.docx' - are currently allowed, but we may charge more for these (in the future).
   - Graphs are permitted as output, in principle, as long as they are based on safe output. It is important to explain clearly which data each graph contains and to specify the number of observations it is based on. Each data point must be based on at least 10 observations; it is therefore not enough to show only total population numbers.
3. Visibility: It is not permitted to hide output in the output files. It is not permitted to hide columns, rows or tabs in Excel files. Neither is it permitted to present pivot tables for review.
4. Other:
   - The name of the output may not contain more than 100 characters as this could cause problems with exporting the output from the secure CBS microdata environment. Neither can the name of the output submitted begin with a date.
   - More information about types of output and the procedure for submitting output can be found at Submitting output or other export files.
   - Place the output in a zip file in one of the existing export folders. Do not create a folder here yourself, because then the output will not be transferred from the CBS microdata environment to the regular network of Statistics Netherlands.
   - The output is removed from the folder every half hour; after that it is no longer visible to the researcher. If the files are still open at that time, they will be corrupted and we will not be able to open them. It is therefore important to prepare the output outside the export folder and copy it to the export folder when it is ready.

Appendix 1: Examples of confidentiality guidelines

The following guidelines are described in detail:
General guidelines:

0. No microdata: It is not permitted to offer microdata as output.
1. Minimum number of observations: All output should have **at least 10 units** (unweighted) underlying any cell or data point.

Additional guidelines for models:

2. Models: All modelled output should have at least 10 degrees of freedom; degrees of freedom = (number of observations) -/- (number of parameters) -/- (other restrictions of the model). In addition to that, residuals and graphs/plots of residuals will not be released.

Additional guidelines frequency tables[4]:

3. Group disclosure: In all tabular and similar output no cell can contain more than 90% of the total number of units in its row or column to prevent group disclosure.

Additional guidelines for magnitude tables[5]:

4. Dominance: In all magnitude tables and similar quantitative data, the largest contributor to a cell should not contribute more than 50% of the total value for that cell.

Obviously, the first guideline of 'no microdata' applie sto all kinds of outputs. The table below shows the different types of output and the rules that apply to each of them.

Type of statistic	Type output	Confidentiality guideline(s) that apply
Descriptive statistics	Frequency tables	1, 3
	Magnitude tables and percentiles	1, 3, 4
	Maxima, minima	Do not release
	Mode, median	1, 3
	Means, indices, ratios, indicators	1, 3, 4
	Concentration ratios	1, 3, 4
	Higher moment of distributions (incl. variance, covariance, kurtosis, skewness)	1, 2
	Graphs: pictorial representations of data	1,2,3,4
Correlations and regression analyses	Lineair regression coefficients	1, 2
	Non-lineair regression coefficients	1, 2
	Estimation residuals	Do not release
	Summary and test statistics from estimates (R^2, χ^2 etc)	1, 2
	Correlation coefficients	1

1. Minimum number of observations

Every cell in a frequency or crosstable must contain al least 10 observations. Example 1a shows a crosstable with many small numbers which means that there is a disclosure risk. After looking at this table, a resident of Zuiderwijk aan Zee know that the only family in that town with 20 children is in the lowest income class.

[4] *Frequency table: table showing how often a given observation occurs (i.e. its frequency).*
[5] *Quantitative table: a table in which the cell values are the sum of a variable across all observations. For example, wages and turnover, but also the number of persons employed and number of vacancies (for the latter two, companies are the units of observation).*

In this case, you would have three options in order to mitigate the disclosure risk: merge the last two categories of the number of children (1b) or suppressing cells (1c) or rounding to quintiles (1d). When removing cells, you should make sure that enough cells are suppressed to make it impossible to infer the values cells from other cells.

Example 1a
Table: Households in Zuiderwijk aan Zee, by number of children and income class.

		Income class			
		Low	**Middle**	**High**	**Total**
Number of children	**1-3**	32	64	84	180
	4-10	28	48	28	104
	11-20	10	16	10	36
	>20	1	0	0	1
	Total	71	128	122	321

Example 1b
Table: Households in Zuiderwijk aan Zee, by number of children and income class.

		Income class			
		Low	**Middle**	**High**	**Total**
Number of children	**1-3**	32	64	84	180
	4-10	28	48	28	104
	>10	11	16	10	37
	Total	71	128	122	321

Example 1c
Table: Households in Zuiderwijk aan Zee, by income class and number of children.
This example shows how cells that contained numbers that involved a disclosure risk have been suppressed. Because of this, no information on the family with more than 20 children will now be released.

		Income class			
		Low	**Middle**	**High**	**Total**
Number of children	**1-3**	32	64	84	180
	4-10	28	48	28	104
	11-20	x	x	x	x
	>20	x	x	x	x
	Total	71	128	122	321

Example 1d
Table: Households in Zuiderwijk aan Zee, by number of children and income class.
Because of the risk of group disclosure, this would not be allowed even if the low income class was rounded to 5 and middle/high income was 0. Coincidentally, in this case it is just one fictional family.

		Income class			
		Low	**Middle**	**High**	**Total**
Number of children	**1-3**	30	65	85	180
	4-10	30	50	30	105
	11-20	10	15	10	35
	>20	0	0	0	0
	Total	70	130	120	320

2. Models

In the examples below, we have given two examples of model outputs: one with output from Stata, and one from R. For Stata, we check the number of degrees of freedom noted under 'Residual' (in the case of Example 2a, there are 72), and not the degrees of freedom listed under the model (which simply indicates how many parameters are included). In R, too, the number of degrees of freedom to be checked is under 'Residuals' (in the case of Example 2b, there are 48).

The residuals can not be included in the output, because they present information on individual observations. They can be viewed as a graph when calculating the regression. That graph may be viewed in the CBS microdata environment, but not included as output. In the example, we see that R provides standard information for this, including a minimum and maximum. This information should therefore be removed.

In complex regressions, degrees of freedom are not always present. For this reason, guideline 1 always remains in effect: a minimum number of ten observations for the underlying data on which the model bases its estimates. For omitted variables, an explanation must also be included as to why they have been omitted, because this could be the result of a non-occurring variable inside the population selection, which could lead to a number of 0 being derived.

Example 2a: Model results from Stata

```
. sysuse auto
(1978 automobile data)

. reg price length
```

Source	SS	df	MS	Number of obs	=	74
				F(1, 72)	=	16.50
Model	118425867	1	118425867	Prob > F	=	0.0001
Residual	516639529	72	7175549.01	R-squared	=	0.1865
				Adj R-squared	=	0.1752
Total	635065396	73	8699525.97	Root MSE	=	2678.7

price	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
length	57.20224	14.08047	4.06	0.000	29.13332	85.27115
_cons	-4584.899	2664.437	-1.72	0.090	-9896.357	726.559

Example 2b: Model results from R

```
Call:
lm(formula = y ~ x1 + x2, data = ols_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9384 -1.6412 -0.4262  1.4503  5.6626

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.34110    0.66670   59.01   <2e-16 ***
x1           0.51066    0.03054   16.72   <2e-16 ***
x2                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.2 on 48 degrees of freedom
Multiple R-squared:  0.8535,    Adjusted R-squared:  0.8504
F-statistic: 279.6 on 1 and 48 DF,  p-value: < 2.2e-16
```

3. Group disclosure

We can take a table on the drug use among adolescents in 'Zuiderwijk aan Zee' as an example.

Table 3: Residents of Zuiderwijk aan Zee, by age and drug use.

Drug use Age	None	Only soft drugs	Hard drugs (possibly including soft drugs)	Total
<15 years	78	12	13	103
15 – 17 years	-	367	12	381
18 – 21 years	-	28	389	417
Total	80	407	414	901

The low frequency in the column 'none' for the age groups above 15 years means that there is disclosure: it shows that almost everyone in this group is using drugs, despite the fact that data has been suppressed. Looking at this table, the parent of a 19-year-old boy in this village can therefore be almost certain that his son is using drugs. In this table, then, the column 'None' for the category '<15 years' should also be left blank because the residual group for the other categories is smaller than 10 respondents.

4. Dominance

As an example of a violation of the dominance rule, we will look at firms' investments by region and activity.

Table 4: Investment by companies (mln euro) to region and activity.
Selection: All companies in the Netherlands in the period X

	Region 1	Region 2	Region 3	Total
Activity 1	78	12	9	99
Activity 2	4	367	12	383
Activity 3	10	28	389	427
Total	92	407	410	909

In this example, background tables like those shown below are required in order to check the dominance rule. This background information must be offered in a **separate file** with the name 'NONOUTPUT'.

Table 4a shows that this output complies with guideline 1: All output should have **at least 10 units** (unweighted) underlying any cell or data point. Table 4c (which is derived from table 4b) shows that one cell (Activity 3, Region 3) does not meet the dominance rule. Together, all the companies in this cell invest 389 mln in activity 3. The company that made the largest investment (234 mln) is good for 60% of the total for this cell. This is greater than 50%, meaning that this company dominates this cell. This means that the risk of disclosure too high and this data should therefore not be released in its current form.

Table 4a: Number of contributors in each cell
This table refers to table 4

	Region 1	Region 2	Region 3	Total
Activity 1	14	10	12	36
Activity 2	13	15	11	39
Activity 3	20	21	23	64
Total	47	46	46	139

Table 4b: Size of the largest contributer (mln euro)
This table refers to table 4

	Region 1	Region 2	Region 3	Total
Activity 1	34	4	2	
Activity 2	1	167	5	
Activity 3	2	10	234	
Total				

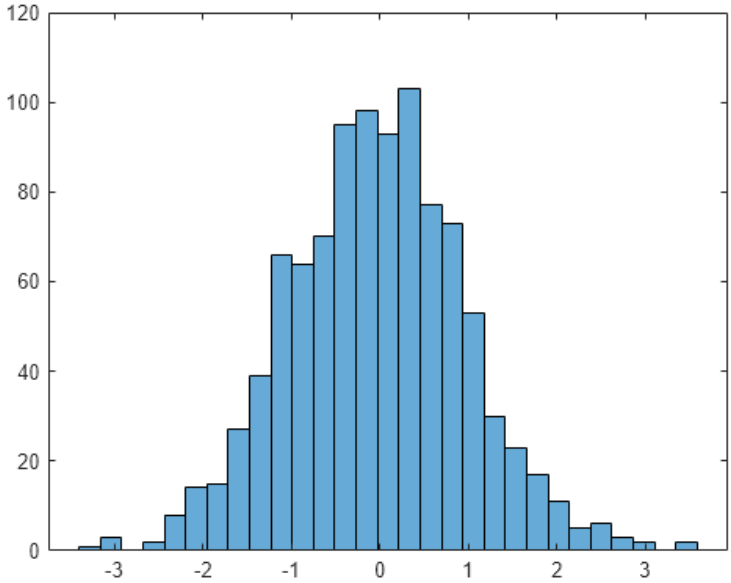
Table 4c: Contribution of largest contributor in %
This table refers to table 4

	Region 1	Region 2	Region 3	Total
Activity 1	44%	33%	22%	
Activity 2	25%	46%	42%	
Activity 3	20%	36%	60%	
Total				

5. Supplementary: Figures

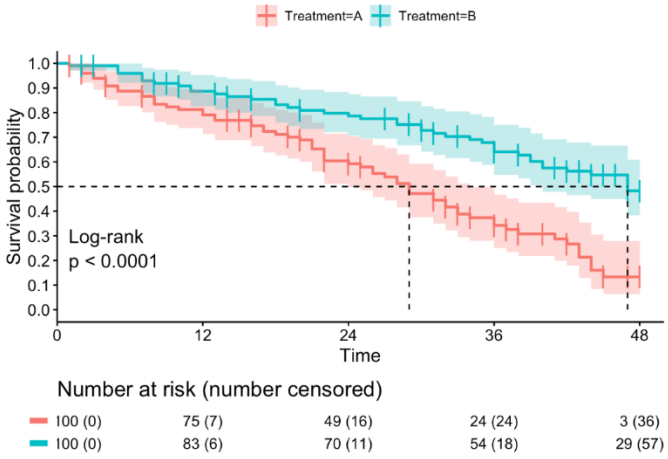
Researchers can offer many types of figures, which the data points should include at least 10 underlying observations. Below we discuss one commonly used figure (histogram) and one that is more difficult figure to interpret (Kaplan Meier plot).

Figure 1: Histogram



For a histogram (Figure 1), we wish to receive the corresponding number for each 'bin'. For the figure above, it is currently possible to see that N=0 for some of the bins, which means it does not meet the guidelines. The scale is also small enough that you could read the frequencies corresponding to the other values with fairly high degree of precision, even for numbers under 10.

Figure 2: Kaplan Meier



A Kaplan Meier plot (Figure 2) allows values of 0 when they represent the beginning of a period, because this is 0 by definition and therefore implies no risk of disclosure. However, it is important to be careful that, even in a population with enough observations, the difference scores do not result in disclosure; intermediate numbers lower than 3 are not allowed here for this reason.

Appendix 2: Standard output description form[6]

Name *[name of user]*	Ben Slim
Telephone number *[telephone number in case of questions regarding the output]*	06-12345678
Institution *[name of institution]*	University of The Hague
Date *[date on which the output was produced]*	dd-mm-yyyy
Project *[contract number + name of research]*	1234 - The flow of benefit claimants into the labour market
This output consists entirely of documentation *[syntax or text file] If so, the rest of the form does not need to be completed.*	No
This output is a modified version of previous rejected output	Yes, out from *dd-mm-yyyy* / No
This output fully meets the output guidelines	Yes
This output does not fully meet the output guidelines I have the following (detailed) explanation for this:	No
Data *[used data files]*	BAANKENMERKEN, BIJSTANDUITKERINGTAB AOTOTUITKERINGTAB, WUSUITKERING1ATAB
Research data file *[description of the composed research data file]*	The research data file consists of all WW, AO and entitled to social assistance to dd-mm-yyyy, paired with the personal characteristics of the same year and job-related data from yyyy.
Purpose and purpose limitation *[purpose of output and how the output helps to address the research question of the project]*	The goal is to examine how many and what kind of benefit recipients transfer to a job within one year. This also answers our research question.
Relationship to earlier output *[description of the relation with earlier delivered output]*	In the previous output, I only looked at unemployment benefits. Now I will carry out the same procedure for a wider package of benefits.
Content of the output *[description of the output, for example there will be some frequencies or crosstabs or regression analysis[7] in this output]*	An excel file with several contingency tables: tables by type of benefit and personal characteristics
I have clearly indicated any selections in the output near to the analyses where these apply to. *[description of the selections made]*	Persons in GBA receiving an unemployment benefit (WW), AO or income support on dd-mm-yyyy. If an additional selection has been made, this is shown in the relevant table.

[6] If this form is not fully completed or added to the output, no check will be carried out on the output. If the form has been completed in full and is submitted by 05:00 a.m., the (regular) output or documentation will be verified within three working days. Output light will be checked the same day.

[7] If the output contains results from models, please clearly mention the underlying number of observations.

Appendix 3: Statistics Netherlands Act

Below are the main sections of the Statistics Netherlands Act that are relevant to offering output.

Section 37

1. The data received by the director general in connection with the performance of his/her duties to implement this act will be used solely for statistical purposes.
2. The data referred to in the first subsection will not be provided to any persons other than those charged with carrying out the duties of the Statistics Netherlands.
3. The data referred to in the first subsection shall only be published in such a way that no recognisable data can be derived from it regarding an individual person, household, company or institution, unless, in the case of data relating to a company or institution, there are good reasons to assume that the company or institution concerned will not object to its publication.

Section 41

1. Contrary to the provisions of Section 37, the director general may, upon request, provide or grant access to a data set to a department, organisation or institution as referred to in the second subsection for the purposes of statistical or academic research, whereby appropriate measures have been taken to prevent the identification of individual persons, households, companies or institutions from that data.

2. A data set as referred to in the first subsection may be provided to or made accessible to:

- **a.** a university (as defined by the Higher Education and Research Act);
- **b.** an organisation or institution for academic research established by law;
- **c.** planning offices established by or by virtue of the law;
- **d.** the Community statistical agency and national statistical agencies of the member states of the European Union;
- **e.** research departments of ministries and other departments, organisations and institutions.

Section 42

The director general will only grant a request as referred to in Section 41 if the director general considers that the applicant has taken adequate measures to prevent the data set being used for purposes other than statistical or academic research.

Section 42a

1. Contrary to the provisions of Section 37 the director general may, upon request, provide data for the purposes of academic research in the field of health care, that Statistics Netherlands has collected for statistical research on the grounds of Section 12a of the Burials and Cremation Act from persons who were involved in a scientific study.

2. The director general will only agree to a request as referred to in the first subsection if the person concerned has expressly consented to such a provision or, if the person concerned has not expressly given permission for the data to be provided, in so far as the person requesting the data adequately demonstrates that:

a. requesting consent during the life of the person concerned was not reasonably possible or could not reasonably be required;
b. it has not been shown that during his/her life the person concerned objected to the processing of his personal data for the purpose of academic research;
c. the research serves the general or public interest;
d. the research cannot be carried out without the relevant data;
e. the research meets any further requirements that can reasonably be imposed on it.

3. The director general may attach further conditions to the provision of data as referred to in the first subsection.

4. Section 41 (2) and Section 42 shall apply mutatis mutandis.

5. Our Minister will adopt further rules, in consultation with Our Minister of Health, Welfare and Sport, after hearing the Data Protection Board, with respect to the power referred to in the first subsection.