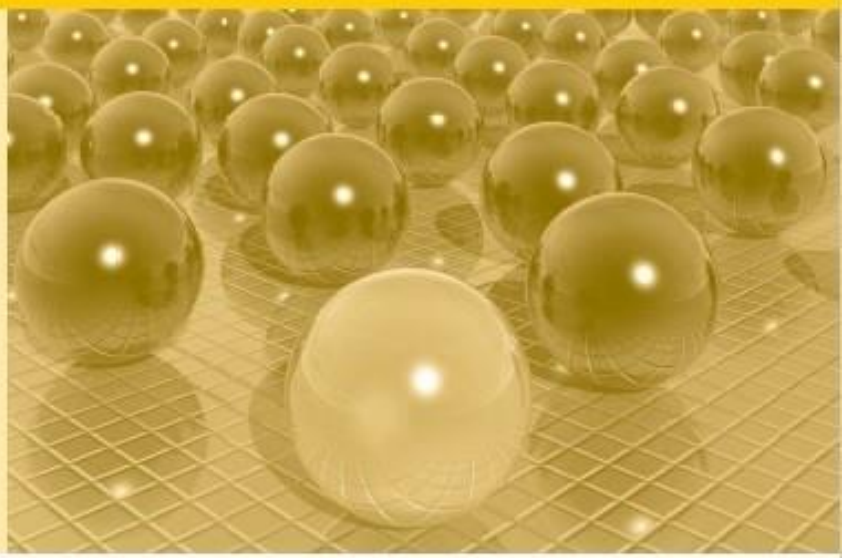
RDC Working Paper

Nr. 54



The Cell-Key-Method at the Research Data Centres of the Statistical Offices of the Federation and the Federal States

Part 2: Effects of the new confidentiality procedure

Patrick Rothe, Volker Güttgemanns, Johannes Rohde, Stefanie Setzer

Published: January 2026

STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER
FORSCHUNGSDATENZENTREN

STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER
FORSCHUNGSDATENZENTREN

Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder
Herstellung: Information und Technik Nordrhein-Westfalen
Telefon 0211 9449-01 • Telefax 0211 9449-8000
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@it.nrw.de

Fachliche Informationen
zu dieser Veröffentlichung:

Statistisches Bundesamt
Forschungsdatenzentrum
Tel.: 0611 75-4220
Fax: 0611 72-3915
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Geschäftsstelle –
Tel.: 0211 9449-2876
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Informationen zum Datenangebot:

Statistisches Bundesamt
Forschungsdatenzentrum
Tel.: 0611 75-4220
Fax: 0611 72-3915
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Geschäftsstelle –
Tel.: 0211 9449-2876
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Erscheinungsfolge: unregelmäßig
Erschienen im Januar 2026

Diese Publikation wird kostenlos als **PDF-Datei** zum Download unter www.forschungsdatenzentrum.de angeboten.

This publication is an English translation of an article originally published in the 03/2024 issue of Wista - Wirtschaft und Statistik. The original article can be found on www.destatis.de

The cell key method at the Research Data Centres of the Statistical Offices of the Federation and the Federal States

Part 2: Effects of the new confidentiality procedure

Patrick Rothe[1], Volker Güttgemanns[2], Johannes Rohde[3], Stefanie Setzer[4],

Abstract:

The cell key method is a disclosure control method that was primarily developed to ensure the confidentiality of frequency tables. Extensive analyses are frequently carried out in the Research Data Centres of the statistical offices of the Federation and the Länder, and the new method can affect the presentation of the results. This article describes the effects that the method has on the quality of results, frequency tables, ratios and time series.

Key Words: confidentiality, stochastic perturbation, post-tabular, ratios, time series

[1] "Fundamental questions of official statistics, Digitalisation, Research Data Centre, Centre of Competence Analses" at the Bayrisches Landesamt für Statistik, e-mail: Patrick.Rothe@statistik.bayern.de
[2] Research Data Centre of the Federal States, main office (formerly), e-mail: Volker.Guettgemanns@it.nrw.de
[3] "Mathematical-statistical methods and experimental statistics" at Information und Technik Nordrhein-Westfahlen, e-mail: Johannes.Rohde@it.nrw.de
[4] Research Data Centre of the Federal Statistical Office Germany, e-mail: stefanie.setzer@destatis.de

Table of contents

I. Introduction

The Statistical Offices of the Federation and the Federal States are introducing a new method to ensure statistical disclosure control for selected statistics: the cell key method (CKM)[5]. In order to ensure confidentiality of the results across all publications, the Research Data Centres (RDC) of official statistics apply this procedure accordingly. The article "The cell key method at the Research Data Centres of the Statistical Offices of the Federation and the Federal States – Part 1: Introduction of the new confidentiality procedure" (Setzer et al. 2024) describes the procedure in detail.

The protective effect of the cell key method arises from a post-tabular perturbation of all frequency counts. In the perturbed table, it is ultimately no longer possible to see which values were perturbed and which are still equal to the original value. This approach represents a big difference to the cell blocking that has been predominantly used so far, in which reported values always correspond to the original value, but particularly small numbers of cases are blocked. The effects of the new procedure on the quality of results, frequency count tables, ratios and time series are presented in this article. Chapter 2 describes in detail the effects on frequency count tables, while Chapter 3 deals with the effects that the cell key method can have on ratios. The application of the cell key method can also have an impact on time series, as shown in Chapter 4. The article concludes with a brief conclusion on the use of the new confidentiality procedure in the Research Data Centres.

II. Effects on frequency count tables

1. Short overview

Two main effects are particularly noteworthy for frequency count tables that were perturbed using the cell key method:

- Tables kept confidential using the cell key method are always consistent. If logically identical table fields occur in different tables (for example in a univariate case number table and as the marginal value of a cross table), the same perturbed frequency count is always stated. This is ensured by a predetermined transition matrix and the record key, which is permanently assigned to the individual survey units.
- Tables kept confidential using the cell key method are not additive. Since inner and marginal table cells are perturbed independently, the perturbed inner cells do not add (or only add randomly) to the perturbed marginal totals. Restoring additivity would be possible in theory but

[5] The cell key method was developed by the Australian Bureau of Statistcs (Fraser/Wooton, 2005).

would affect the consistency and quality of the results and also increase the required computing time, which is why this is abstained from.

2. Effects in detail

Additivity and consistency of the confidential table results represent two essential requirements for statistical disclosure control methods. These two properties are of particular importance, especially in the context of using the cell key method.

Additivity in a table is given if the inner table cells can be added to the total shown in the corresponding marginal cell – which is normal for an unedited table and is generally expected by users. However, certain statistical confidentiality procedures that come from the family of data manipulation methods result in this property being violated. This means that the sum of the added inner table cells does not necessarily correspond to the marginal value shown in the table. This described effect ("non-additivity") also occurs using the cell key method and arises because each table cell is subjected to the data-changing algorithm separately – i.e. independently of all other table cells. Inner fields are therefore treated in the same way as the intermediate or total sums of the table. This separate procedure usually causes a loss of the additivity property.

At first glance, this effect can seem irritating to users. However, this approach ultimately leads to better results concerning data quality and information preservation: The separate perturbation of each table cell prevents deviations from adding up across a table row or column. Thus, the marginal values do not differ greatly from the original values. Otherwise and in extreme cases, a modification by the amount of the maximum deviation multiplied by the number of involved table cells could not be ruled out for marginal totals. However, the independent perturbation of all table cells, including table cells with subtotals or totals, means that the modified value can never deviate further from the original value than the previously set maximum value allows.[6]

However, this advantage in terms of data quality comes at the price of the discrepancy between the added totals of the individual table cells and the subtotals and totals shown in the table. This difference can be significant under certain circumstances. Therefore, users should refrain from carrying out calculations with the information from result tables that were kept confidential using the cell key method, but rather use the total cells shown in the table.

Non-additivity is particularly noticeable when only a few table cells contribute to a marginal total and this can be determined very quickly using simple mental arithmetic. An example of this is the breakdown of the variable "gender" according to the three characteristics "female", "male" and "diverse" in a

[6] Further considerations concering the non-additivity can be found in Höhne/höninger (2018).

fictitious table: The summation of the frequency counts in the three inner fields is very easy, although it is very likely that the sum calculated by adding the three inner fields differs from the total shown in the confidential table. However, in larger tables with a larger number of contributing columns or rows, this problem is not immediately noticeable.

Consistency, on the other hand, refers to the property of a specific table cell to always show the identical content value - regardless of the specific table in which this combination of characteristics is shown. Not all confidentiality procedures can guarantee this, as it is essential to ensure that the same content table cell is treated identically in all cases by the chosen method. The cell key method is able to fulfill this property due to its design. This is ensured by aggregating the record keys that contribute to an individual table cell into the cell key, which gives the procedure its name. Since this turns out to be identical when the same combination of feature carriers is concerned, the data modification is always identical. This applies regardless of the table or publication method the corresponding table field is published in.

In addition to greater user-friendliness (e.g. independence from the time of the results retrieval) this feature also goes hand in hand with increased protection of the data and the parameters used to perturb it. In contrast, if the behavior was not consistent, the underlying original information could be inferred - at least approximately - from a larger number of result retrievals, since the positive and negative deviations from the resulting mean average balance out in an unbiased procedure.

Figure 1 is based on an example from the article “The cell key method at the Research Data Centres of the Statistical Offices of the Federation and the Federal States – Part 1: Introduction of the new confidentiality procedure” (Setzer and others, 2024) and shows the modifications of the marginal cells that can be caused by the cell key method.

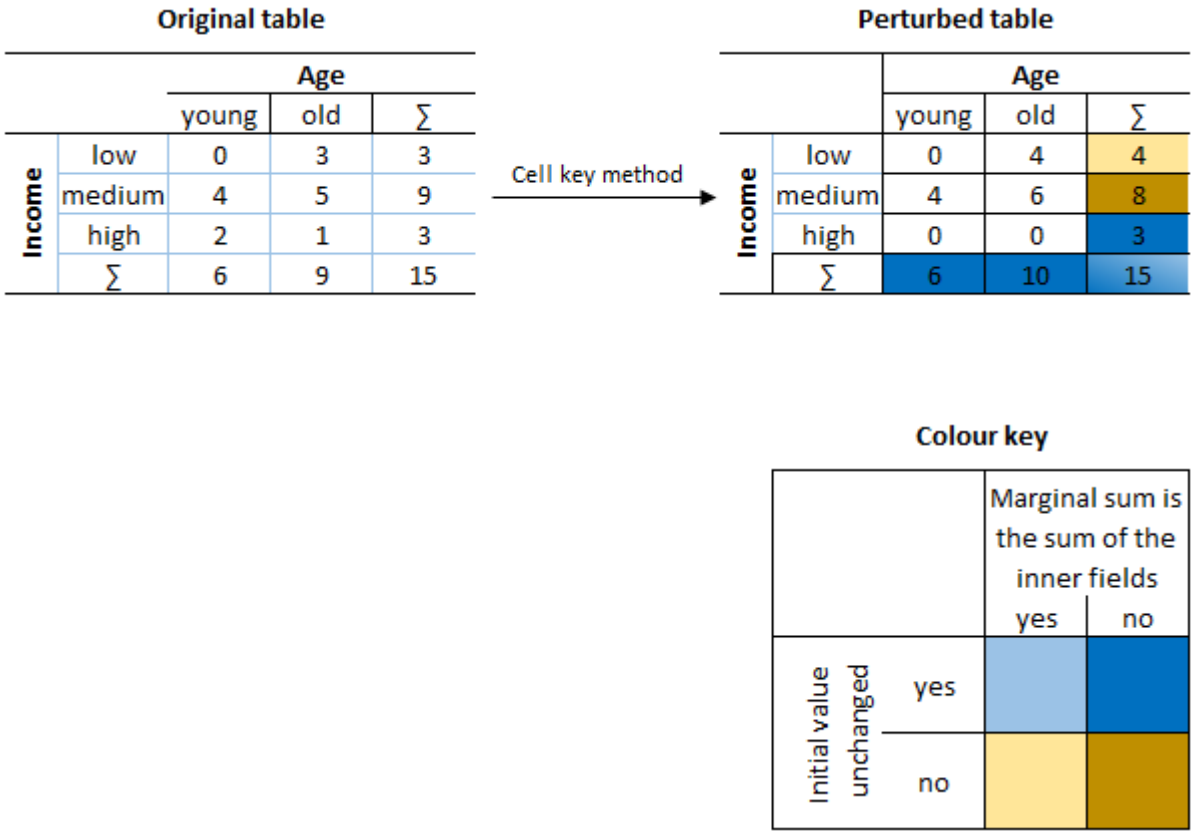


Figure 1: Effects of the cell key method on the additivity of a frequency count table

3. Further quality criteria

A confidentiality procedure should always be designed in a way that ensures the balance between the protection of the respondents' information and the preservation of the information potential of statistics. In addition to the aspects of consistency and additivity already explained, Höhne/Höninger (2018) name three further fundamental objectives of secrecy procedures, with different procedures prioritizing the individual goals differently. Overview 2 presents these objectives and addresses their fulfillment by the cell key method and the cell blocking that is currently mainly used in Research Data Centres.

Objective	Cell key method	Cell blocking
1. No implausible values	The design of the transition matrix can ensure that no implausible results are generated in frequency count tables (original frequency count 0 is not changed). However, when calculating key figures, implausibilities can arise, especially if the frequency count is small, for example if the numerator and denominator of ratios are changed in opposite directions or the sign changes in trend analyses. However, these undesirable effects can be avoided by taking appropriate measures.	The frequency count will not be changed, but will be blocked if the number of cases is too low. Therefore, implausible case numbers cannot arise. This also applies to the calculation of key figures and time series.
2. Protection of individual data	When designing the transition matrix, it is ensured that the modifications are large enough to ensure the protection of the individual data.	Frequency counts below a set minimum frequency count are blocked; counter-blocking prevents re-calculations. The protection of individual data is therefore ensured if the procedure is used correctly.
3. Acceptance of results	When designing the transition matrix, it is considered that the modifications are (in accordance with item 2) small enough to be accepted by the majority of users.	Depending on the size of the (sub)population under consideration and the level of detail of the analysed characteristics, some users consider the blocks to be too restrictive.
4. Results are consistent	Logically identical table cells are always perturbed with the same value.	Blocking patterns are also implemented across tables so that modifications are consistent when the cell blocking procedure is implemented correctly. However, with frequently used statistics and many result tables, overlooked connections between table cells can cause mistakes.
5. Results are additive	Due to the independent perturbation of the individual frequency counts, tables that are kept confidential are not additive.[1]	Additivity is retained in tables as long as there are no blocks.

Objective fully achieved Objective partly achieves Objective not achieved

[1] Additivity could be restored by an algorithm that changes the perturbed values so that the inner table cells sum equals the marginal cells. However, this approach would have two decisive disadvantages: on the one hand, establishing additivity requires additional computing power and extends the time required to create the table; on the other hand, the results would no longer be consistent, which would significantly affect the quality of the results.

Figure 2: Comparison of cell key method and cell blocking concerning the achievement of non-disclosure objectives

On the question of acceptance of the results (item 3): Overall, internal test calculations by the statistical offices have shown that the loss of information caused by the application of the cell key method is small if suitable parameters are chosen (Rohde and others, 2021). It should be noted that with small (original) frequency counts, perturbations with a small value in absolute terms can already have a strong effect: For example, if the original frequency count 3 is perturbed with the value + 6 (+ 200%), the relative deviation is significantly higher than for the perturbation of a higher frequency count with the same perturbation value (e.g. + 0.6% for the original frequency count 1,000). The generation of tables with

many small frequency counts should therefore be avoided, if possible, by not carrying out evaluations at an unnecessarily deep regional or subject level.

III. Ratios

Ratios are mathematical relationships between statistical quantities that represent a meaningful connection. In general, three categories of ratios can be distinguished (Bourier, 2011):

- Structural numbers: These relate a partial size to the total size and are often referred to as proportion values. Example: gender quotas.
- Relational numbers: Here, different variables are related to one another that have a factual connection. Example: gross domestic product per inhabitant.
- Measurable numbers: These relate quantities that are similar in content, but at different times, periods or regions. Example: change in gross domestic product in comparison to the previous year.

Calculating ratios is a common method in scientific research. Thereby, it is important to observe the applicable confidentiality regulations, especially in the context of research projects in the Research Data Centres. Against this background, a fundamental requirement is to ensure confidentiality in accordance with subject-specific concepts in order to ensure a consistent approach across different analyses and publications. This may also require special statistical regulations for dealing with ratios, which must be considered by the Research Data Centres and their users.

For users of the Research Data Centres, it is crucial to understand the statistical disclosure control methods and its effects in order to be able to assess the quality of the calculated ratios themselves. The following section explains the effects of the cell key method on ratios and on their significance.

1. Effects on ratios

The basis for the confidentiality of ratios using the cell key method is the modification of the numerators and denominators according to their cell keys. This implies that, especially with small frequency counts, the calculated ratio can differ significantly from the value of the non-perturbed ratio. These inaccuracies decrease with larger frequency counts. For this reason, it is not recommended to calculate ratios that are based on only a few cases.

2. Handling of ratios

Ratios are often calculated from the perturbed frequency counts of the numerator and denominator, which is here referred to as the "a posteriori ratio".[7] Data-altering statistical disclosure control procedures like the cell key method can therefore lead to undesired effects:

- Inaccuracy: When using the cell key method, the direction of modifications in the frequency counts can randomly be in opposite directions, i.e. the numerator and denominator are changed by a large positive or negative value (or vice versa). This can lead to significant deviations between the original and the a posteriori ratio, especially for small frequency counts. However, this effect diminishes with increasing frequency counts.
- Implausibility: Another undesired effect can occur if, for example, after the modification, the originally smaller numerator is larger than the modified denominator. In this case, a posteriori proportion values of over 100% would occur.
- Change of statement: Especially when analysing time series, the application of the cell key method can lead to trend distortions or even a trend reversal (see chapter 4).

While the perturbation of frequency counts shows a fixed variance, Enderle and others (2018) note that this does not apply to ratios calculated based on two perturbed values. The original frequency counts from which the ratio is calculated play a major role in the extent of the deviation of the actual ratio value $R := X/Y$ from the perturbed ratio $\hat{R} := \hat{X}/\hat{Y}$: the bigger the frequency counts X and Y, the smaller the deviation of the ratio. To illustrate the problem of small frequency counts, Enderle and others (2018) give the following example: The relatively small original frequency counts $x = 4$ and $y = 4$ are related to each other. The following applies to the actual ratio: $R = 4/4$, i.e. 100%. With a maximum deviation of $D = 3$, the perturbed frequency count could take on the values $\hat{x} = x + D = 7$ and $\hat{y} = y - D = 1$. The perturbed ratio value would then be $\hat{R} = 7/1$, i.e. 700%.

3. Assessment of a ratio's quality

Scientists must be able to assess the quality of a share value. This requires information about the distribution of random deviations from the original size, although the details of these deviations cannot be disclosed. To assess the quality of share values, the relative standard deviation can be used as a measure of the scatter around the original frequency count.[8]

[7] Further suggestions for techniques for using stochastic perturbation on ratios that are formed as a quotient from two sums of values can be found in Gießing (2013).
[8] Instead of the absolute standard deviation, the relative standard deviation is used as a measure of quality as this relates the random scatter around the original number of cases to the size of the original value.

Assume that q_{xy} represents the probability that a modified proportion value $\tilde{v} := \left(\widetilde{\frac{x}{y}}\right)$ is based on the original numerator $x \in \mathcal{D}_x$ and the original denominator $y \in \mathcal{D}_y$ and $d(v)_{xy}$ represents the deviations between the modified proportion value $\tilde{v}$ and the possible original proportion value v then the (absolute) standard deviation for the modified share value $\tilde{v}$ can be approximated as follows:

$$\sigma(\tilde{v}) = \sqrt{\sum_{\substack{x\in\mathcal{D}_x \\ y\in\mathcal{D}_y}} q_{xy} \cdot d(v)_{xy}^2}$$

The corresponding Relative Root Mean Square Error (RRMSE) or the approximated relative standard deviation (also called coefficient of variation) for the modified proportion value $\tilde{v}$ is then obtained according to

$$k(\tilde{v}) = \frac{\sigma(\tilde{v})}{\tilde{v}}.$$

For averages or ratios in which the sum of values in the numerator and/or denominator (modified if necessary) is included in the calculation, the described method requires some adjustments. An alternative approach is to determine the relative standard deviation of v using the transition probabilities.

For higher education statistics, the effects of different parameterisations on the quality of ratios were examined. The result shows that the choice of the probability of stay and the maximum deviation from the original frequency count are not the decisive factors for the quality of ratios. Rather, the quality depends primarily on the size of the underlying frequency counts that are included in the calculation. Accordingly, small frequency counts in particular affect the quality and significance of the ratios. Sufficient statistical significance be only be guaranteed when the base numbers in the numerator and denominator reach a certain size (Enderle/Vollmar, 2019).

4. Recommendations concerning the calculation of ratios

Without knowledge of the unmodified ratios, it is not possible for users to assess the quality of a ratio. The size of the relative standard deviation or RRMSE depends primarily on the size of the included frequency count. It is therefore generally recommended to only calculate ratios based on sufficiently large sample sizes in the numerator and denominator as the variance in these cases tends to zero anyway.

If the calculation of individual ratios based on small frequency counts is essential for a research question, users of the Research Data Centres can ask their supporting RDC location for support.

IV. Time series

Time series are valuable instruments for displaying temporal progressions and developments. Key figures are calculated based on points in time or time periods. This allows analysis of absolute or relative changes over time.

As with the calculation of ratios, the applicable confidentiality regulations and specific regulations for a given statistics must also be observed when looking at time series.

1. Effects on time series

The use of the cell key method can have negative effects on the analysis of time series. This applies both when comparing two survey years kept confidential using the cell key method as well as when considering two survey years that are kept confidential with different methods, i.e. when comparing the survey years before and after the introduction of the cell key method. Especially if there are very small differences in the unmodified frequency counts in the two compared observation periods, differences between the survey years could be increased or decreased if the values under consideration are modified in opposite directions. In extreme cases, even a trend reversal is possible. Very weak changes over time should therefore be interpreted with caution when using the cell key method.

2. Handling of time series

Time series are based on data from different time periods and used to analyse the development of issues over time. The individual data points for the different reporting periods form the basis for corresponding considerations.

Generation of differences from modified data points can lead to less certain results. This is particularly true when a difference is calculated based on small frequency counts.

When calculating differences, specific challenges must be considered when maintaining confidentiality using the cell key method:

I. Handling differences when both data points come from CKM data sets

When it comes to the difference between two randomly modified data points, similar challenges can arise as with balances (see section 4.3). This is particularly true when the frequency count is small and when the individual points show strong, opposite modifications. As a consequence, applying the cell key method could theoretically cause changes in trend strength and even reverse trends.

There are (statistic-specific) solutions to these challenges. In the 2022 census, a method was chosen that avoids a sign change. In population statistics and higher education statistics, however, the difference is formed from the two perturbed data points. From a methodological point of view, this leads

to a doubling of the variance specified in the CKM parameters. If X_1 and X_2 are two modified data points with identical variance V specified as a CKM parameter, their difference is:

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) = 2V$$

Compared to the individual absolute frequency count, this increases the risk of a larger deviation from the original difference. However, if the time series (or the individual data points) are based on sufficiently large frequency counts, the effect of distorting the trend strength becomes negligible.

In order to minimize the risk of a significant change in the trend strength, the Research Data Centres recommend that analyses in scientific publications based on time series to only be carried out on the basis of a sufficiently large frequency count. The sample size limit depends on the used statistics and the considered characteristics. The RDC location responsible for the respective statistic can provide advice on this.

II. Handling of differences when the data points come from data sets using different confidentiality methods

If the confidentiality procedure changes between two reporting times, for example by switching from cell blocking to the cell key method, this leads to a methodological break in the time series. In this case, the Research Data Centres have stipulated that no differences may be calculated at the point of break, i.e. where the confidentiality procedure is changed. In consequence, the time series is interrupted at the breaking point. Comparisons between the reporting dates of the changeover are therefore not possible. However, some specialist statistics, such as the census, have decided to subsequently ensure confidentiality for earlier years of their statistics using the cell key method. Corresponding comparisons based on the two sets of data kept confidential using the cell key method are then possible again. The information as to whether this was implemented for a specific statistic can be found in the corresponding metadata reports provided by the Research Data Centres.

3. Handling of balances

In terms of the fundamental problem, dealing with balances is very similar to dealing with time series. If a balance, i.e. the difference between two modified frequency counts, is calculated (for example the migration balance in population statistics), the combination of two stochastic variables leads to greater uncertainty in the result. This happens because the variance of the difference is doubled compared to the existing variance when modifying individual frequency counts (see section 4.2). This can distort the significance of the balance when the frequency count is small. There is also the theoretical possibility that when using the cell key method, the sign of the balance changes if both original frequency counts are modified in opposite directions and the frequency counts are both small and close to each other.

However, the relative effect of such distortions due to possible large deviations from the original difference decreases with larger underlying sample sizes - similar to ratios and time series differences.

The Research Data Centres therefore recommend that when balancing, corresponding evaluations should only be carried out on the basis of sufficiently high frequency counts. If there are many small frequency counts, the evaluation should, if possible, be carried out at a higher regional or subject-specific aggregation level (for example, districts instead of municipalities or 4-digit economic sectors instead of 5-digit ones). Weakly populated categories should also be grouped as well.

V. Conclusion

Compared to traditional confidentiality methods, especially the widespread cell blocking, the cell key method has some advantages in terms of balancing the loss of information and the risk of detection. However, this article has shown in detail that data users and Research Data Centres also have to take new special features into account when evaluating and interpreting the results that are kept confidential using the cell key method.

Since the cell key method is designed to keep frequency count tables confidential, problems can arise with further analyses based on ratios, time series and balances. Those can be avoided if evaluations are always carried out on the basis of sufficiently large frequency counts. If problems or uncertainties arise when accessing data via the Research Data Centres, users can contact their supporting RDC location at any time.

IV. References

Bourier, Günther. Beschreibende Statistik. Praxisorientierte Einführung – Mit Aufgaben und Lösungen. Wiesbaden 2011, page 19 ff. DOI: 10.1007/978-3-8349-6556-1

Enderle, Tobias/Giessing, Sarah/Tent, Reinhard. Designing Confidentiality on the Fly Methodology – Three Aspects. In: Domingo-Ferrer, Josep/Montes, Francisco (Herausgeber). Privacy in Statistical Databases. LNCS (Lecture Notes in Computer Science). 2018. Edition 11126, page 28 ff. DOI: 10.1007/978-3-319-99771-1_3

Enderle, Tobias/Vollmar, Meike. Geheimhaltung in der Hochschulstatistik. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, page 87 ff.

Fraser, Bruce/Wooton, Janice. A proposed method for confidentialising tabular output to protect against differencing. Work session on statistical data confidentiality. Supporting paper. Genf 2005. [Access on 30.04.2024]. Available at: unece.org

Giessing, Sarah. What shall we do with the ratios? Work session on statistical data confidentiality. Supporting paper. Ottawa 2013. [Access on 7.05.2024]. Available at: unece.org

Höhne, Jörg/Höninger, Julia. Die Cell-Key-Methode – ein Geheimhaltungsverfahren. In: Zeitschrift für amtliche Statistik Berlin Brandenburg. Edition 3+4/2018, page 14 ff. [Access on 30.04.2024]. Available at: www.statistischebibliothek.de

Marley, Jennifer K./Leaver, Victoria L. A Method for Confidentialising User-Defined Tables: Statistical Properties and a Risk-Utility Analysis. In: Proceedings of 58th World Statistical Congress. 2011. [Access on 30.04.2024]. Available at: 2011.isiproceedings.org

Rohde, Johannes/Seifert, Christiane/Gießing, Sarah/Setzer, Stefanie (unter Mitarbeit von Breitenfeld, Jörg/Brings, Stefan/Höhne, Jörg/Höninger, Julia/Rothe, Patrick/Schedding-Kleis, Ulrike). Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens. Version 1.1 vom 23.04.2021. Internal document of the Statistical Offices of the Federation and the Federal States.

Setzer, Stefanie/Rohde, Johannes/Güttgemanns, Volker/Rothe, Patrick. Die Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder - Teil 1: Vorstellung des neuen Geheimhaltungsverfahrens. In: WISTA Wirtschaft und Statistik. Edition 3/2024, page 31 ff. English translation available as RDC working paper no. 53 at www.forschungsdatenzentrum.de

Up to now, the following RDC working papers were published:

Arbeitspapier Nr. 53: The cell key method at the Research Data Centres of the Statistical Offices of the Federation and the Federal States. Part 1: Introduction of the new confidentiality procedure, S. Setzer, J. Rohde, V. Güttgemanns, P. Rothe, Januar 2026

Arbeitspapier Nr. 52: Möglichkeiten zur Verknüpfung mehrerer Mikrozensus-Jahre zu einem Mikrozensus-Panel (2012–2015), S. Brocker, H. Mühlenfeld, November 2020

Arbeitspapier Nr. 51: Record-Linkage bei fehlenden Betriebsnummern im Produzierenden Gewerbe, L. Möll, Juli 2020

Arbeitspapier Nr. 50: Statistische Geheimhaltung - Der Schutz vertraulicher Daten in der amtlichen Statistik, P. Rothe, Februar 2019

Arbeitspapier Nr. 49: Formal, faktisch oder absolut nachgefragt? Die Auswirkungen der Entgeltumstellung auf die Entwicklung der Nachfrage in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, R. Voshage, U. Drafz, H. Habla, M. Klumpe, C. Meisdrock, K. Nowak, S. Raab, A. Richter, M. Rößner, K. Schmidtke, November 2015

Arbeitspapier Nr. 48: Bereitstellung der Daten des Zensus 2011 über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, S. Raab, C. Meisdrock, Oktober 2015

Arbeitspapier Nr. 47: FiRe – Ein Schritt zur Teilautomatisierung der Geheimhaltungsprüfung, J. Pohlisch, J. Höninger, R., Voshage, Dezember 2014

Arbeitspapier Nr. 46: Entwicklung des Mikrosimulationsmodells EITDsim, G. Struch, April 2013

Arbeitspapier Nr. 45: Ein Mikrosimulationsmodell zur Berechnung der einkommensteuerlichen Bemessungsgrundlage und der Steuerzahlung auf Basis des Taxpayer-Panels, T. P., Schmidt, April 2013

Arbeitspapier Nr. 44: Das SAS-Makro newvar. Entwicklung und Anwendung eines Hilfsinstruments zur effizienten Erstellung neuer Variablen in der DRG-Statistik, T. Hochgürtel, T. Lösch, März 2012

Arbeitspapier Nr. 43: Average wage, qualification of the workforce and export performance in German enterprises: Evidence from KombiFiD data, J. Wagner, Januar 2012

Arbeitspapier Nr. 42: The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study, J. Wagner, Dezember 2011

Arbeitspapier Nr. 41: How to define an enterprise and assign trade declarations to the right one: Exploration of German traders' micro transaction data, C. Stirböck, August 2011

Arbeitspapier Nr. 40: Definition von nutzerseitigen Kriterien für Datenstrukturfiles, J. Höninger/M. Rosemann/R. Voshage, Juli 2011

Arbeitspapier Nr. 39: Improvement of data access – The long way to remote data access in Germany, M. Brandt/M. Zwick, Juni 2011

Arbeitspapier Nr. 38: Decentralised Access to Confidential Microdata in Europe, M. Brandt/P. Eilsberger/M. Zwick, Juni 2011

Arbeitspapier Nr. 37: Masking Micro Data with Stochastic Noise, J. Höhne/J. Höninger, Mai 2011

Arbeitspapier Nr. 36: Enthüllungsrisiko beim Remote Access: Die Schwerpunkteigenschaft der Regressionsgerade, A. Vogel, April 2011

Arbeitspapier Nr. 35: Temporary agency work and firm performance, S. Nielen/A. Schiersch, April 2011

Arbeitspapier Nr. 34: Harmonisation of statistical confidentiality in the Federal Republic of Germany, M. Brandt/A. Crößmann/C. Gürke, März 2011

Arbeitspapier Nr. 33: Remote Access. Eine Welt ohne Mikrodaten ??, G. Ronning/P. Bleninger/ J. Drechsler/C. Gürke, Februar 2011

Arbeitspapier Nr. 32: Compiling a Harmonized Database from Germany´s 1978 to 2003 Sample Surveys of Income and Expenditure. T. Bönke/C. Schröder/C. Werdt, Mai 2010

Arbeitspapier Nr. 31: The Research Potential of New Types of Enterprise Data based on Surveys from Official Statistics in Germany., J. Wagner, Oktober 2009

Arbeitspapier Nr. 30: Geschlechterspezifische Einkommensunterschiede bei Selbstständigen im Vergleich zu abhängig Beschäftigten – Ein empirischer Vergleich auf der Grundlage steuerstatistischer Mikrodaten, P. Eilsberger/M. Zwick, Januar 2008

Arbeitspapier Nr. 29: Reichtum in Niedersachsen und anderen Bundesländern –Ergebnisse der Steuergeschäftsstatistik 2003 für Selbstständige (Freie Berufe und Unternehmer) und abhängig Beschäftigte, P. Böhm/J. Merz, November 2008

Arbeitspapier Nr. 28: Exports and Productivity in the German Business Services Sector. First Evidence from the Turnover Tax Statistics Panel, A. Vogel, Juli 2009

Arbeitspapier Nr. 27: Künstler in den Daten der amtlichen Statistik, C. Haak, August 2008

Arbeitspapier Nr. 26: Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany, B. Fitzenberger/K. Kohn/A. C. Lembcke, August 2008

Arbeitspapier Nr. 25: German engineering firms during the 1990's. How efficient are export champions?, A. Schiersch, Juli 2008

Arbeitspapier Nr. 24: Zum Einkommensreichtum Älterer in Deutschland – Neue Reichtumskennzahlen und Ergebnisse aus der Lohn- und Einkommensteuerstatistik (FAST 2001), P. Böhm/J. Merz, Februar 2008

Arbeitspapier Nr. 23: Neue Datenangebote in den Forschungsdatenzentren. Betriebs- und Unternehmensdaten im Längsschnitt, M. Brandt/D. Oberschachtsiek/R. Pohl, November 2007

Arbeitspapier Nr. 22: Stichprobendaten von Versicherten der gesetzlichen Krankenversicherung – Grundlage und Struktur des Datenmaterials, P. Lugert, Dezember 2007

Arbeitspapier Nr. 21: KombiFid – Kombinierte Firmendaten für Deutschland, S. Bender/J. Wagner/ M. Zwick, November 2007

Arbeitspapier Nr. 20: Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, U. Kaiser/ J. Wagner, Juni 2007

Arbeitspapier Nr. 18: Die Gehalts- und Lohnstrukturerhebung: Methodik, Datenzugang und Forschungspotential, H.-P. Hafner/ R. Lenz, Mai 2007

Arbeitspapier Nr. 17: Anonymisation of Linked Employer Employee Datasets. Theoretical Thoughts and an Application to the German Structure of Earnings Survey, H.-P. Hafner/R. Lenz, Dezember 2006

Arbeitspapier Nr. 16: Die europäische Union – Integration von unten oder Eliteprojekt? Eine Sekundäranalyse von Mikrodaten der amtlichen Statistik, R. Nauenburg, November 2006

Arbeitspapier Nr. 15: Keeping in Touch – A Benefit of Public Holidays Using German Time Use diary Data, J. Merz/L. Osberg, November 2006

Arbeitspapier Nr. 14: Zur Konzeption eines Taxpayer-Panels für Deutschland, D. Vorgrimler/C. Gräb/S. Kriete-Dodds, November 2006

Arbeitspapier Nr. 13: Anonymisierte Daten der amtlichen Steuerstatistik, D. Vorgrimler, September 2006

Arbeitspapier Nr. 12: Mikrosimulation in der Betriebswirtschaftlichen Steuerlehre, R. Maiterth, August 2006

Arbeitspapier Nr. 11: Der Anteil der freien Berufe und der Gewerbetreibenden an der Gemeindefinanzierung, M. Zwick, September 2006

Arbeitspapier Nr. 10: Konstruktion und Bewertung eines ökonomischen Einkommens aus der Faktisch Anonymisierten Lohn- und Einkommensteuerstatistik, T. Bönke/F. Neher/C. Schröder, August 2006

Arbeitspapier Nr. 9: Anonymising business micro data – results of a German project, R. Lenz/M. Rosemann/D. Vorgrimler/R. Sturm, Juni 2006

Arbeitspapier Nr. 8: Scientific analyses using the Continuing Vocational Training Survey 2000, R. Lenz/H.-P. Hafner/D. Schmidt, Juni 2006

Arbeitspapier Nr. 7: A standard for the release of microdata, R. Lenz/ D. Vorgrimler/M. Scheffler, Juni 2006

Arbeitspapier Nr. 6: Measuring the disclosure protection of micro aggregated business microdata, R. Lenz, Juni 2006

Arbeitspapier Nr. 5: De facto anonymised microdata file on income tax statistics 1998, J. Merz/ D. Vorgrimler/M. Zwick, Oktober 2005

Arbeitspapier Nr. 4: Matching German turnover tax statistics, R. Lenz/D. Vorgrimler, Juni 2005

Arbeitspapier Nr. 3: The research data centres of the Federal Statistical Office and the statistical offices of the Länder,*S.* Zühlke/M. Zwick/S. Scharnhorst/T. Wende, März 2005

Arbeitspapier Nr. 2: Eine kommunale Einkommen- und Körperschaftsteuer als Alternative zur deutschen Gewerbesteuer: Eine empirische Analyse für ausgewählte Gemeinden, R. Maiterth/M. Zwick, April 2005

Arbeitspapier Nr. 1: Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik, H. Müller, März 2005