Advances in Methodology and Statistics, 2024, *21*(2), 1–19. https://doi.org/10.51936/trfy1604

Investigating the need for raking on already adjusted sample according to the population characteristics: Case of Croatian senior population survey

Berislav Žmuk

University of Zagreb, Faculty of Economics & Business, Zagreb, Croatia

Abstract

The research paper investigates the impact of raking on survey estimates in the context of a Croatian senior population survey conducted in January 2022. Raking is a statistical technique used to improve the accuracy and precision of survey estimates by adjusting the survey sample to match known characteristics of the population. The main objective of the paper is to determine whether applying raking to a survey is beneficial when the survey sample has already been adjusted to match certain population characteristics. In this case, the survey sample was initially adjusted using quota sampling based on the 2011 census data, considering gender and NUTS2 region as variables. An important factor in this study is that a new census was conducted in 2021, which provided updated population characteristics. The research aims to compare survey point estimates using both the 2011 and 2021 population data. The research conducted raking analyses using a variety of combinations of variables, including age, gender, county, and NUTS2 region. A total of 11 different combinations of these variables were used in the raking analyses. The study also examined the impact of the different number of iterations in the raking process, specifically observing results for 5, 10, and 20 iterations. The research included statistical tests to compare point estimates derived from the survey sample to those derived from raking using different sets of population data. In almost all cases, no statistically significant differences in point estimates were found. However, statistically significant differences did emerge when comparing sample point estimates to raking point estimates based on the 2021 census data, particularly when the number of estimated raking weights was the highest.

Keywords: elderly population, population census, raking method, survey sampling

1. Introduction

According to Groves et al. (2004), a survey can be considered as a systematic method for collecting information from a subset of the target population. The expectation is that the gathered information will be accurate and precise enough to draw reliable conclusions

Email address: bzmuk@efzg.hr
ORCID iD: 0000-0003-3487-1376

about the target population. However, since not all units of the population are observed and inspected, the collected information can only lead to estimates that may deviate to some degree from the true population values. To enhance the accuracy and precision of these estimates, various post-survey approaches and methods have been developed.

The Random Iterative Method (RIM), also known as the Iterative Proportional Fitting Method (IPFM), or simply referred to as "raking", is an approach that utilizes known population values to "correct" or adjust sample values. This correction involves an iterative process of calculating weights to align certain variables in the sample as closely as possible with their corresponding values in the population. The values of other sample variables, for which population values are not known, are adjusted by taking into account these raking weights.

Anderson and Fricker Jr (2015) suggest that raking is an important but often overlooked survey tool. On the other hand, Izrael et al. (2004) point out that almost all survey researchers use raking in their survey analyses. According to Battaglia et al. (2004), raking is designed to reduce various survey-related biases, such as nonresponse, noncoverage, and sampling variability. To initiate the raking process, control totals are typically derived from national censuses (Battaglia et al., 2009). Furthermore, raking is not only intuitive but also easy to implement, thanks to advances in statistical computer program development. Therefore, its widespread use is not surprising.

In general, raking yields significant improvements in the accuracy and precision of estimates. This is particularly evident in cases where sampling is conducted without prior information about the population. However, the question arises as to whether raking can statistically significantly enhance the accuracy and precision of point estimates when the researcher already incorporates knowledge about the population structure in the sampling process.

Therefore, this paper aims to investigate whether there are any improvements in survey point estimates when raking is applied, even when the sample structure is already adjusted to the population structure based on certain characteristics. To address this research question, results from a survey among Croatian seniors serve as the foundation for analysis. The survey employed quota sampling with gender and NUTS2 region as quota variables. However, the values for these quota variables were based on the 2011 population census, whereas the survey took place in January 2022. It is important to note that the initial results of the 2021 population census were published in February 2022, after the survey was conducted.

Consequently, it has been decided to compare point estimates for selected survey variables between the original sample data and point estimates calculated based on raking analyses using certain variables from both the 2011 and 2021 censuses. As a result, two research hypotheses were formulated. The first research hypothesis, H1, posits that in the majority of cases, point estimates for selected survey variables differ statistically significantly between the sample data and raking data based on the 2011 census. Similarly, the second research hypothesis, H2, suggests that, in the majority of cases, point estimates for selected survey variables vary statistically significantly between the sample data and raking data based on the 2021 census. If both research hypotheses are supported, the conclusion will indicate that raking can enhance accuracy and precision, even when quota sampling is applied.

The paper is organized as follows. After a brief introduction, the second section provides a detailed description of the conducted survey on seniors in Croatia. This section also covers the selected variables to be used in the analysis, as well as the methodological approach for the analyses. In the third section, the initial raking results are discussed. Subsequently, a comparison is made between the point estimates from the original sample data and the estimates obtained during the conducted raking analyses for selected survey variables. The fourth section serves as the conclusion of the paper and offers recommendations for further research.

2. Data and methodology

The paper analyses data from a survey conducted among the elderly population in Croatia. This survey was part of the project Senior 2030—Thematic Network for Active Aging Policy in Croatia, conducted in January 2022. The data collection method utilized a computer-assisted telephone interview (CATI) approach, primarily influenced by restrictions related to the Covid-19 pandemic.

The target population for the survey consisted of individuals aged 65 years and older in Croatia. The survey aimed to gain insights into the financial situation of retired individuals and their willingness to rejoin the labour market. It also investigated the general level of activity among retired individuals and their acceptance of technology. Additionally, the survey explored various other topics related to the elderly population.

The survey, in total, comprised 54 questions. However, respondents had the option to answer fewer than 54 questions based on their responses to filter questions. Most of the questions were closed-ended, allowing respondents to select only one answer. The questionnaire also included other question types, including multiple response categorical and open-ended questions. As a result, the survey took approximately 20 to 30 minutes to complete.

To select respondents, the quota random sampling method was employed. The objective was to achieve a sample structure that mirrored the population's demographics based on the 2011 census, considering gender and region (NUTS 2 level). The sampling process occurred in two steps. Firstly, random digit dialing (RDD) was used to select households. If there were multiple eligible individuals in a household, only one person was randomly chosen.

The final sample consists of a total of 701 respondents. The composition of the sample, categorized by gender and NUTS 2-level regions, is presented in Table 1. In addition to these variables, the raking analyses will also consider the variables Age and County. There are several reasons for this choice.

Firstly, it allows examination of the impact of raking on variables that were not initially observed in the process of forming the representative sample based on population shares. Secondly, the Age and County variables have varying numbers of categories, which makes it interesting to investigate how raking variables perform when they have different numbers of categories, thus necessitating the estimation of more or fewer weights.

Lastly, introducing additional raking variables provides the opportunity to combine different variables and compare raking results when one, two, or three raking variables are included in the analyses. It's important to note that the County variable, like NUTS 2, provides location information about respondents, but on a more precise NUTS 3 level. As a result, in the raking analyses, NUTS 2 and County will not be used simultaneously, leading to a total of 11 different combinations of raking variables (1 = NUTS 2; 2 = Age; 3 = Gender; 4 = County; 5 = NUTS 2, Age; 6 = NUTS, Gender; 7 = County, Age; 8 = County, Gender; 9 = Age, Gender; 10 = NUTS 2, Age, Gender; 11 = County, Age, Gender).

Table 1. Structure of the target population in the sample, as per the 2011 and 2021 Censuses, in percentages

Variable	Share in the sample/population		
	S	C_1	C_2
NUTS2			
Adriatic Croatia	34.09	34.11	35.63
City of Zagreb	17.97	18.03	18.27
Northern Croatia	18.83	18.67	19.10
Pannonian Croatia	29.10	29.20	27.00
Age			
65–69	35.38	26.63	32.12
70–74	28.10	28.00	26.29
75–79	20.26	23.14	16.90
80–84	12.13	14.25	14.12
85 and more	4.14	7.99	10.57
Gender			
Female	61.06	60.96	58.40
Male	38.94	39.04	41.60
County			
City of Zagreb	18.83	18.03	18.27
County of Bjelovar-Bilogora	2.71	2.90	2.72
County of Dubrovnik-Neretva	3.42	2.88	2.99
County of Istria	6.56	4.94	5.43
County of Karlovac	4.56	3.59	3.16
County of Koprivnica-Križevci	2.57	2.70	2.56
County of Krapina-Zagorje	3.00	3.09	2.88
County of Lika-Senj	0.86	1.66	1.29
County of Međimurje	2.00	2.34	2.49
County of Osijek-Baranja	6.70	6.79	6.53
County of Požega-Slavonia	0.86	1.85	1.68
County of Primorje-Gorski kotar	9.84	7.38	7.88
County of Sisak-Moslavina	4.99	4.44	3.97
County of Slavonski Brod-Posavina	3.57	3.67	3.39
County of Split-Dalmatia	7.13	9.95	10.60
County of Šibenik-Knin	2.43	3.15	3.04
County of Varaždin	4.14	3.90	3.81
County of Virovitica-Podravina	2.43	1.92	1.78

Continued on next page

Table 1. Structure of the target population in the sample, as per the 2011 and 2021 Censuses, in percentages (Continued)

Variable	Share in the sample/population		
	S	C_1	C_2
County of Vukovar-Sirmium	3.28	4.05	3.76
County of Zadar	3.85	4.16	4.41
County of Zagreb	6.28	6.64	7.35

Legend: S = Survey, C_1 = Census 2011, C_2 = Census 2021.

As mentioned, the survey was conducted in January 2022, utilizing data about the structure of the target population from the 2011 census. However, a more recent census was conducted in Croatia from September to November 2021 (Croatian Bureau of Statistics [CBS], 2021). The reason for using data from the older census is the unavailability of data from the 2021 census. In fact, no official census data were accessible until the survey commenced in January 2022. The initial basic census results only became available at the very end of February 2022 (CBS, 2022).

Therefore, in the raking analyses, data from both the 2011 and 2021 censuses will be used. It is expected that the weights calculated in the raking process using the 2011 census data for variables NUTS2 and Gender will be lower than the weights calculated using the 2021 census data. This expectation is rooted in the fact that the sample's structure mirrors the target population's structure in 2011 based on these variables. However, for a more accurate reflection of the present situation, it's reasonable to incorporate up-to-date information. Hence, in this case, data from the 2021 census will also be used.

According to the 2011 census data, the target population consisted of 758 633 individuals. However, based on the 2021 census, the target population had increased to 868 638 persons. This represents a population growth of 110 005 individuals or 14.05 % over the span of 10 years.

This increase also had a notable impact on the changes in the structure of individuals aged 65 years and older concerning their age, gender, and location. In Table 1, the structures of the target population in 2011 and 2021 are provided alongside the sample structure for convenient comparison. As anticipated, the differences in structure are most minimal between the sample and the 2011 census for variables NUTS2 and Gender, which were initially used to create the sample. However, for the other two variables, Age and County, which will be employed in the raking analyses, the differences in structures between the sample and the 2011 census are more pronounced.

As mentioned earlier, the survey questionnaire comprised a total of 54 questions. However, for the purposes of conducting raking analyses in this paper, only eight questions will be examined. The objective of the paper is not to perform raking analyses on all survey questions, but rather to observe the effects of raking and compare the results of raking analyses for different scenarios. Therefore, the variables selected for raking analyses have unique characteristics compared to other observed variables. These differences include the number of categories (two or more), variable scale types (qualitative or quantitative), the type of observed estimates (proportion or arithmetic mean), and the amount of data collected for each variable (ranging from 82 to 701 responses). The complete list of observed variables for the raking analyses is provided in Table 2.

The variable, "Physical activity", served as the foundational question to determine

Table 2. Observed variables from the survey

Variable	Variable categories	n
Physical activity	Yes; No	701
Internet use	Daily use; Weekly use; Monthly use; No Internet use	701
Work activity reason	Financial reasons; Non-financial reasons	123
Willingness to work again	Yes; No	578
Willingness to work again reason	Financial reasons; Non-financial reasons	82
Feeling of loneliness	I never feel lonely; Several times a year; Several times a month; Several times a week; Daily	701
Missing funds for life	I have enough resources to meet my basic living needs; Up to 500 HRK; Between 501 and 1,000 HRK; Between 1,001 and 1,500 HRK; Between 1,501 and 2,000 HRK; Between 2,001 and 3,000 HRK; Between 3,001 and 4,000 HRK; Between 4,001 and 5,000 HRK; More than 5,000 HRK	701
Work experience	0; 1; 2; ...; 47; 48	701

whether a respondent had engaged in any recent physical activity. The variable, "Internet use", categorizes individuals aged 65 years and older into four distinct levels of Internet usage.

While the first two observed variables were examined among the full sample of 701 respondents, the following three variables were explored in a reduced sample due to different responses to filter questions. The question "Work activity reason" was answered only by respondents who had worked after their retirement, totaling 123 individuals in the sample. Conversely, the question "Willingness to work again" pertained to 578 respondents who did not work after their retirement. The question related to the "Willingness to work again reason" was posed solely to respondents who expressed a willingness to work but had not done so after retirement, resulting in 82 individuals in the sample.

The variable "Feeling of loneliness" evaluates the level of loneliness among respondents, categorizing them into five distinct categories. "Missing funds for life" is another variable, encompassing even more categories (nine in total) when compared to the "Feeling of loneliness" variable. In contrast, the variable "Work experience" is distinct from the previous ones; it is a discrete quantitative variable measured on a ratio scale, ranging from 0 to 48 years in the case of the observed sample.

The raking estimator is one of the most commonly used forms of post-survey weight adjustment, applied after sample selection and data collection (Valliant et al., 2013). Raking aims to enhance the accuracy and precision of survey estimates by leveraging known information about the target population (Heeringa et al., 2013). To compute weights using the raking approach, several steps must be followed (Biffignandi & Bethlehem, 2021).

In the first step, initial weights with values of 1 are introduced. Subsequently, the weights are adjusted for the first raking variable to make the weighted sample representative according to that variable. For example, in the sample, there are 39 % males and 61 % females. However, in the population, according to the 2011 Census, there are 42 % males and 58 % females. Therefore, males in the sample should be given a higher weight, above 1, whereas females will receive a lower weight, below 1. The same process is then repeated for each additional raking variable, which may disrupt the balance established for previous raking variables. Once this is done for all raking variables, the adjustment process begins again with the first raking variable. This iterative process continues until the weights no longer change significantly.

After the final weights are estimated, they are multiplied and associated with specific respondents based on their characteristics related to the raking variables. A detailed description of the raking procedure can be found in Kalton and Flores-Cervantes (2003) and Kolenikov (2014).

In raking, there is a possibility that the process may not converge. Therefore, it is advisable to introduce a tolerance for changes in weights. However, for the purpose of comparing estimation results in this research, final weights will be determined after a specific number of iterations. Consequently, the analysis will involve the examination of results after 5, 10, and 20 iterations. The raking process will be conducted using Stata software.

During the raking process, there is a chance of generating extremely low or high weight values. To ensure direct comparability of raking results, such extreme weights will not be trimmed.

In the next phase of the analysis, the estimated weights in various scenarios will be compared, and the differences will be discussed. Following the observation of the weights, point estimates for the selected eight variables will be calculated and compared. This comparison will be made between the sample point estimates and the raking point estimates at the variable category level. Additionally, differences between raking point estimates based on census 2011 and census 2021 data will also be examined.

The first seven variables are measured on a nominal categorical scale, whereas only the last variable, "Work experience", is measured on a ratio scale. Consequently, for the first seven variables, it will be assessed whether there is a statistically significant difference in the estimated proportions for each variable category between the original sample results and the raking results based on the 2011 census data, and between the original sample results and the raking results based on the 2021 census data. This will involve performing statistical tests for the equality of two proportions. In all the conducted tests, the null hypothesis will posit that there is no difference in the compared proportions, while the alternative hypothesis will suggest that there is some statistically significant difference. In this case, the test statistics will be calculated as follows:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

where $\hat{p}_1$ and $\hat{p}_2$ represent the two proportions being compared, and $\sigma_{\hat{p}_1 - \hat{p}_2}$ is the standard error calculated as

$$\sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1 - 1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2 - 1}},$$

with n_1 and n_2 representing sample sizes that may vary depending on the observed variable (refer to Table 2).

For the variable "Work experience" the difference in means will be examined. The null hypothesis will state that there is no difference between the two means being compared, while the alternative hypothesis will assume the presence of some statistically significant difference in the observed means. The test statistic is calculated as follows:

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sigma_{\overline{x}_1 - \overline{x}_2}},$$

where $\overline{x}_1$ and $\overline{x}_2$ represent the two averages being compared, and $\sigma_{\overline{x}_1-\overline{x}_2}$ is the standard error calculated as

$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1}+\frac{\hat{\sigma}_2^2}{n_2}},$$

with $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ representing estimated variances, and n_1 and n_2 denoting sample sizes.

3. Results and discussion

The raking process was conducted as explained in the previous chapter. Given that raking was carried out while considering 11 different combinations of raking variables, two different censuses, and three iteration levels for eight variables from the questionnaire, a total of 2448 weights were estimated. Table 3 presents basic descriptive statistics of the estimated weights.

The results presented in Table 3 confirmed the previously mentioned assumption that raking weights are smaller when data from the 2011 census are considered. In cases where variables Gender and NUTS2 were used as raking variables, the raking weights tend to be smaller. This can be attributed to the fact that these two variables were initially used to establish a particular respondent structure in the sample. On the other hand, raking weights tend to be, on average, higher and exhibit greater variation when census 2021 data were employed. The only exception to this trend is the case in which variables County and Gender were used in the raking, although the difference appears to be minor. The highest variation in weight values occurred when 2021 census data were used in conjunction with variables County, Age, and Gender as raking variables.

According to the results in Table 3, it can be concluded that the statistics for weight estimates are consistent across different numbers of iterations. This holds true in cases where only one variable was applied as a raking variable. However, when multiple variables are used as raking variables, this statement does not hold. Differences between weights for varying numbers of iterations would become apparent if the results in Table 3 were presented with more decimal places (five or more). Unfortunately, due to technical limitations of the paper, it was not possible to include such precision in Table 3. Nonetheless, these minor differences in the weight estimates for different numbers of iterations reveal that an equilibrium state can be reached quite rapidly. This reaffirms the earlier decision not to observe more than 20 iterations, as the gain in precision would be minimal while the computation time would significantly increase.

Figure 1 compares the number of weights and weight averages separately for raking conducted based on census 2011 data and census 2021 data. This figure illustrates the situation for the case of 5 conducted iterations in the raking analysis. As previously concluded, based on the results from Table 3, the weight averages for the observed iterations are very similar. Therefore, the relationship between the number of weights and weight averages in the case of 10 and 20 conducted raking iterations is quite similar to what is presented in Figure 1 for 5 iterations.

According to Figure 1, it appears that the increase in the number of weights follows an increase in average weights. However, this increase is negligible when using 2011 census data, whereas it becomes more pronounced when observing 2021 census data. The conducted linear regression, which estimated weights based on 2011 census data, suggests that, on average, for every increase in weight, the average weight should increase by 3.54E-04. This increase is so small that it is not statistically significant (p value = 0.2248). On the other hand, the linear regression model, estimating weights based on 2021 census data, indicates that,

Table 3. Descriptive statistics of weights based on various raking variables and combinations, different numbers of raking iterations, and data from 2011 and 2021

				C_1				C_2			
n_R	Raking variable	n_I	n_W	M	SD	min	max	M	SD	min	max
1	Age	*	10	1.20	0.44	0.75	1.93	1.28	0.72	0.83	2.56
	County	*	42	1.11	0.36	0.75	2.16	1.07	0.30	0.69	1.96
	Gender	*	4	1.00	0.00	1.00	1.00	1.01	0.08	0.96	1.07
	NUTS2	*	8	1.00	0.01	0.99	1.00	1.00	0.05	0.93	1.05
2	Age, Gender	5	20	1.20	0.42	0.75	1.94	1.29	0.70	0.80	2.73
		10	20	1.20	0.42	0.75	1.94	1.29	0.70	0.80	2.73
		20	20	1.20	0.42	0.75	1.94	1.29	0.70	0.80	2.73
	County, Age	5	182	1.22	0.48	0.54	2.75	1.20	0.58	0.57	3.40
		10	182	1.22	0.48	0.54	2.75	1.20	0.58	0.57	3.40
		20	182	1.22	0.48	0.54	2.75	1.20	0.58	0.57	3.40
	County, Gender	5	84	1.10	0.36	0.74	2.19	1.08	0.30	0.68	2.05
		10	84	1.10	0.36	0.74	2.19	1.08	0.30	0.68	2.05
		20	84	1.10	0.36	0.74	2.19	1.08	0.30	0.68	2.05
	NUTS2, Gender	5	16	1.00	0.01	0.99	1.01	1.01	0.07	0.89	1.11
		10	16	1.00	0.01	0.99	1.01	1.01	0.07	0.89	1.11
		20	16	1.00	0.01	0.99	1.01	1.01	0.07	0.89	1.11
	NUTS2, Age	5	40	1.20	0.41	0.71	1.99	1.28	0.67	0.78	2.69
		10	40	1.20	0.41	0.71	1.99	1.28	0.67	0.78	2.69
		20	40	1.20	0.41	0.71	1.99	1.28	0.67	0.78	2.69
3	County, Age, Gender	5	332	1.17	0.44	0.54	2.77	2.16	5.02	0.01	41.34
		10	332	1.17	0.44	0.54	2.77	2.17	5.06	0.01	41.62
		20	332	1.17	0.44	0.54	2.77	2.17	5.06	0.01	41.62
	NUTS2, Age, Gender	5	78	1.18	0.39	0.71	2.00	1.26	0.63	0.75	2.81
		10	78	1.18	0.39	0.71	2.00	1.26	0.63	0.75	2.81
		20	78	1.18	0.39	0.71	2.00	1.26	0.63	0.75	2.81

* Since raking was conducted according to just one variable, the final weights were estimated in the first step after the initial iteration.

Legend: C_1 = Census 2011, C_2 = Census 2021, n_R = no. of raking variables, n_I = no. of raking iterations, n_W = no. of weights

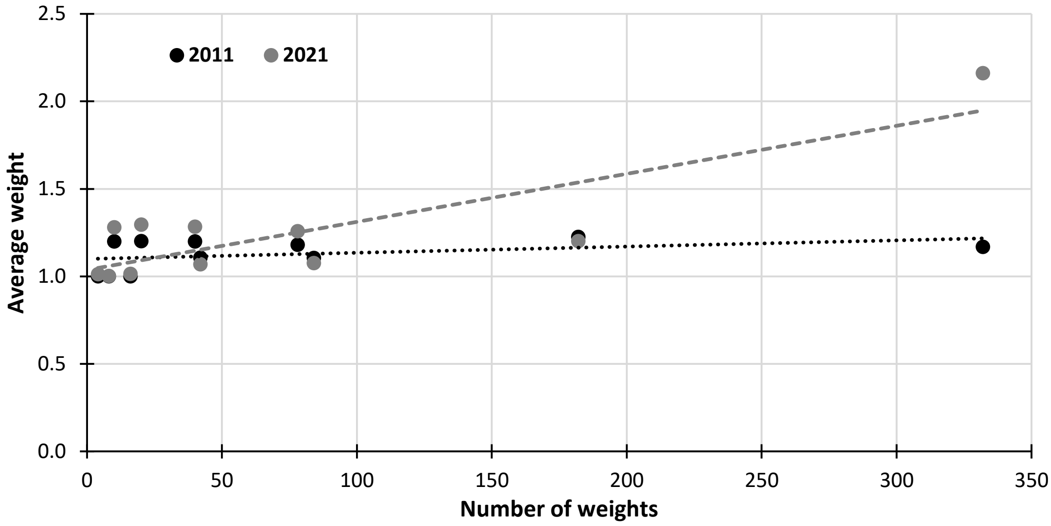

Figure 1. Average 2011 and 2021 weight comparison for different number of weights, number of raking iterations = 5

on average, for every increase in weight, the average weight should increase by 2.75E-03. While this increase is also relatively small, it appears to be statistically significant (p value = 0.0012).

These results indicate that weight averages tend to remain at a consistent level regardless of the number of weights, provided that the variables used in the raking analysis are instrumental in determining the respondent's structure within the sample. The sample's structure is established based on variables Gender and NUTS2 and their values from the 2011 census. Consequently, an increase in the number of weights did not impact the weight averages in this scenario.

However, when using 2021 census data, a statistically significant positive correlation emerged between the number of weights and the weight averages. Therefore, the larger the disparity between the sample structure and the structure defined by the utilized raking variables, the more pronounced the effect of an increase in the number of weights on weight averages becomes.

After observing the estimated weights at a general level, Figure 2 and Figure 3 focus on point estimates and standard errors for the first selected questionnaire question, "Physical Activity", under different conditions: when various variables were included in the raking process and when 5 iterations were performed. Figure 2 presents point estimates for the percentage of older individuals who are physically inactive. Since the variable "Physical Activity" is binary, point estimates for the percentage of physically active older individuals can be easily calculated as 1 minus the estimated percentage of physically inactive older individuals. Meanwhile, the standard errors remain the same for both cases (the percentage of physically inactive and physically active individuals).

The sample proportion of physically inactive older individuals is 0.097. According to Figure 2, the 2011 census estimates tend to be not only higher than this value but also higher than the 2021 census estimates for the majority of raking variables. However, in the last conducted raking analysis, which was based on the variables County, Age, and Gender, the 2021 census estimate yielded significantly different values from the sample estimate and the 2011 census estimate. As illustrated in Figure 3, this difference is accompanied by higher standard errors compared to all other raking analyses, where the standard errors are either at the sample level or slightly above it.

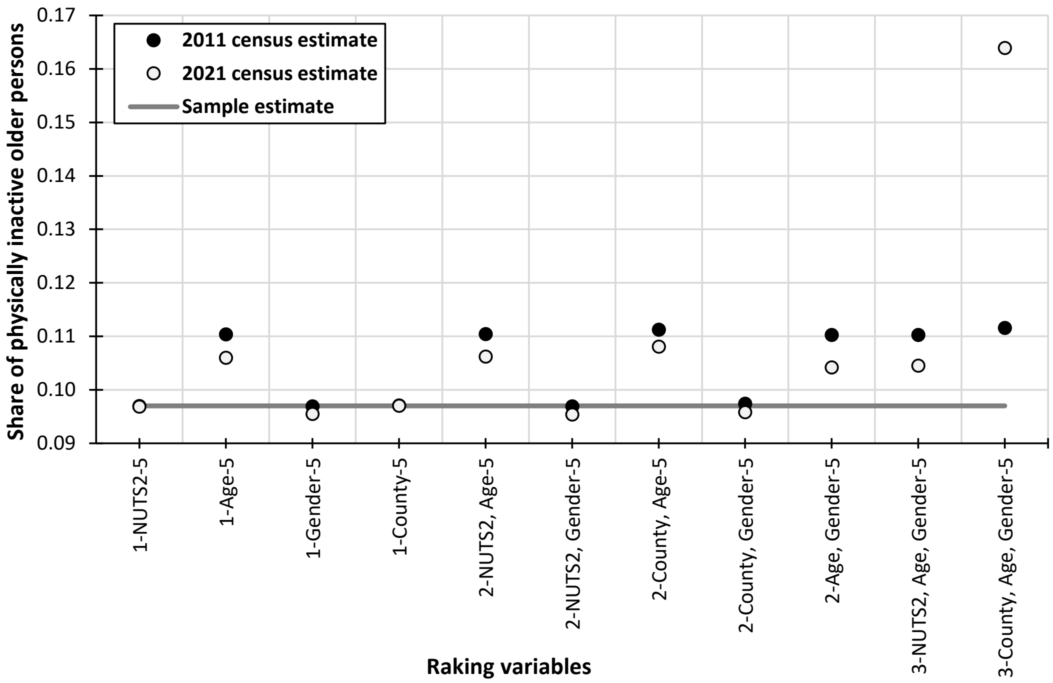

Figure 2. Point estimates of variable "Physical activity" using different raking variables, number of raking iterations = 5, n = 701 respondents

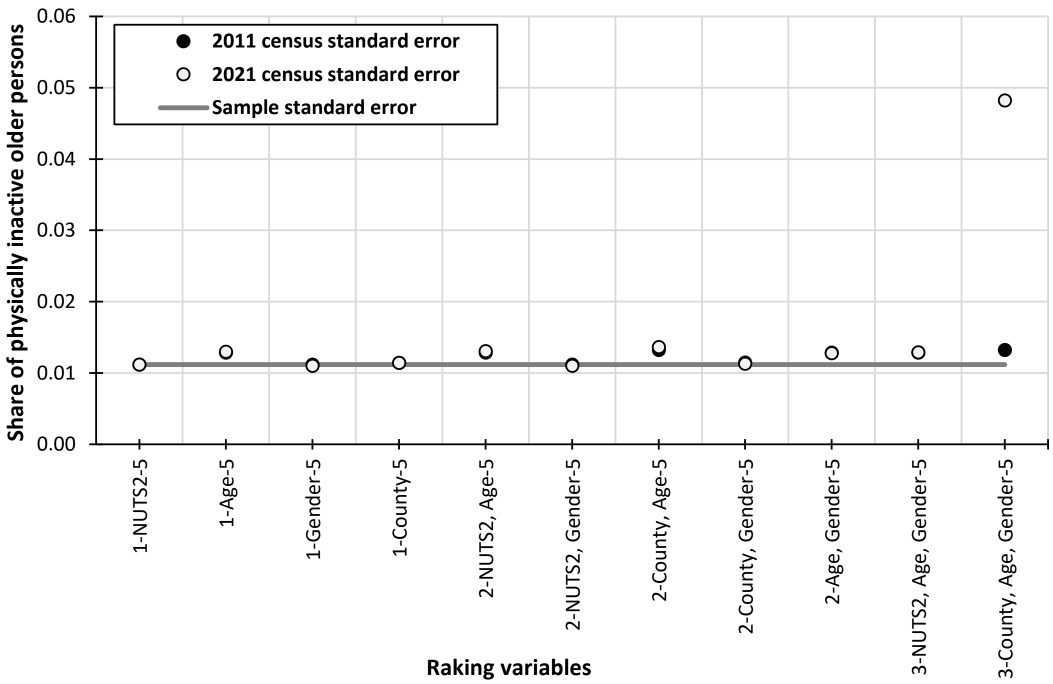

Figure 3. Standard errors of variable "Physical activity" using different raking variables, number of raking iterations = 5, n = 701 respondents

As mentioned earlier, the results for the proportion of physically active older individuals can be easily calculated based on the provided results in Figure 2 and Figure 3. Furthermore, the results of the raking analysis with 10 and 20 iterations align with the presented results. Given the paper's length limitations, these additional results will not be separately displayed.

The variable "Physical Activity" is a binary variable, making the presentation of results straightforward. However, the variable "Internet Use" comprises four levels, which complicates the graphical presentation and comparison of raking results. In Figure 4, estimates for all four levels of the "Internet Use" variable are presented for the sample, the 2011 census, and the 2021 census. A noticeable pattern, similar to that seen in the "Physical Activity" variable in Figure 2, is observed. In cases where the raking variables included only NUTS2, only Gender, NUTS2 and Gender, and County and Gender, the sample, 2011 census, and 2021 census estimates are very similar. Conversely, in the raking analyses where the raking variables were County, Age, and Gender, the 2021 census estimates are notably different from the sample and 2011 census estimates when compared to raking analyses with other included raking variables.

In Table 4, point estimates of the observed variables based on the original sample are displayed. Furthermore, point estimates calculated by considering raking analyses conducted based on 2011 and 2021 census data are also presented. It's worth noting that Table 4 focuses on the case where raking variables were Age and Gender, and the raking results are based on 5 iterations.

Table 4. Point estimates of the observed variables based on original sample and raking, considering 2011 and 2021 census data, with raking variables Age and Gender, number of raking iterations = 5

		Estimations			Test statistics	
Variable/Category	n	S	C_{11}	C_{21}	S/C_{11}	S/C_{21}
Physical activity						
Yes	633	0.9030	0.8897	0.8958	0.8140	0.4493
No	68	0.0970	0.1103	0.1042	−0.8140	−0.4493
Internet use						
Daily use	345	0.4922	0.4542	0.4670	1.4218	0.9431
Weekly use	48	0.0685	0.0646	0.0649	0.2895	0.2710
Monthly use	11	0.0157	0.0156	0.0145	0.0139	0.1880
No Internet use	297	0.4237	0.4655	0.4537	−1.5776	−1.1320
Work activity reason						
Financial reasons	39	0.3171	0.3227	0.3216	−0.0947	−0.0766
Non-financial reasons	84	0.6829	0.6773	0.6784	0.0947	0.0766
Willingness to work again						
Yes	82	0.1419	0.1220	0.1339	0.9985	0.3912
No	496	0.8581	0.8780	0.8661	−0.9985	−0.3912
Willingness to work again reason						
Financial reasons	30	0.3659	0.3562	0.3586	0.1281	0.0955

Continued on next page

Table 4. Point estimates of the observed variables based on original sample and raking, considering 2011 and 2021 census data, with raking variables Age and Gender, number of raking iterations = 5 (Continued)

		Estimations			Test statistics	
Variable/Category	n	S	C_{11}	C_{21}	S/C_{11}	S/C_{21}
Non-financial reasons	52	0.6341	0.6438	0.6414	−0.1281	−0.0955
Feeling of loneliness						
I never feel lonely	445	0.6348	0.6317	0.6269	0.1203	0.3060
Several times a year	44	0.0628	0.0568	0.0571	0.4699	0.4429
Several times a month	66	0.0942	0.0949	0.0978	−0.0487	−0.2296
Several times a week	56	0.0799	0.0804	0.0798	−0.0383	0.0061
Daily	90	0.1284	0.1361	0.1384	−0.4272	−0.5495
Missing funds for life						
I have enough resources to meet my basic living needs	362	0.5164	0.5256	0.5276	−0.3454	−0.4191
Up to 500 HRK	31	0.0442	0.0448	0.0438	−0.0500	0.0416
Between 501 and 1,000 HRK	77	0.1098	0.1172	0.1156	−0.4322	−0.3388
Between 1,001 and 1,500 HRK	71	0.1013	0.0949	0.0933	0.4001	0.5013
Between 1,501 and 2,000 HRK	74	0.1056	0.1039	0.1040	0.0990	0.0948
Between 2,001 and 3,000 HRK	57	0.0813	0.0757	0.0743	0.3894	0.4907
Between 3,001 and 4,000 HRK	18	0.0257	0.0226	0.0238	0.3789	0.2273
Between 4,001 and 5,000 HRK	5	0.0071	0.0071	0.0073	−0.0017	−0.0319
More than 5,000 HRK	6	0.0086	0.0081	0.0104	0.0870	−0.3483
Work experience						
Average	701	32.7760	32.5837	32.8013	0.3301	−0.0430

Legend: S = Sample, C_1 = Census 2011, C_2 = Census 2021, n = sample size.

However, the primary focus of Table 4 is not on the point estimates but on examining whether there is a statistically significant difference between them. Due to technical restrictions, Table 4 only provides calculated statistics for testing the significance of the point estimate differences. In all cases, it has been demonstrated that the null hypotheses, at the significance level of 0.05, stating that the difference in point estimates is equal to zero, cannot be rejected. Therefore, it can be concluded that there is no statistically significant difference in point estimates in all observed cases and combinations.

Whereas Table 4 presents only one example, Figure 5 displays the test statistics of conducted statistical tests, which examine differences in observed variables for different raking variables and the number of iterations. According to Figure 5, where the presence of statistically significant differences in point estimates between the sample and raking results based on the 2011 census data is tested, it is evident that almost all dots or test statistics fall within the range of −1.96 and 1.96. As a result, it can be concluded that there are no statistically significant differences in point estimates when using the original sample data compared to the data corrected in the raking process based on the 2011 census data, at the

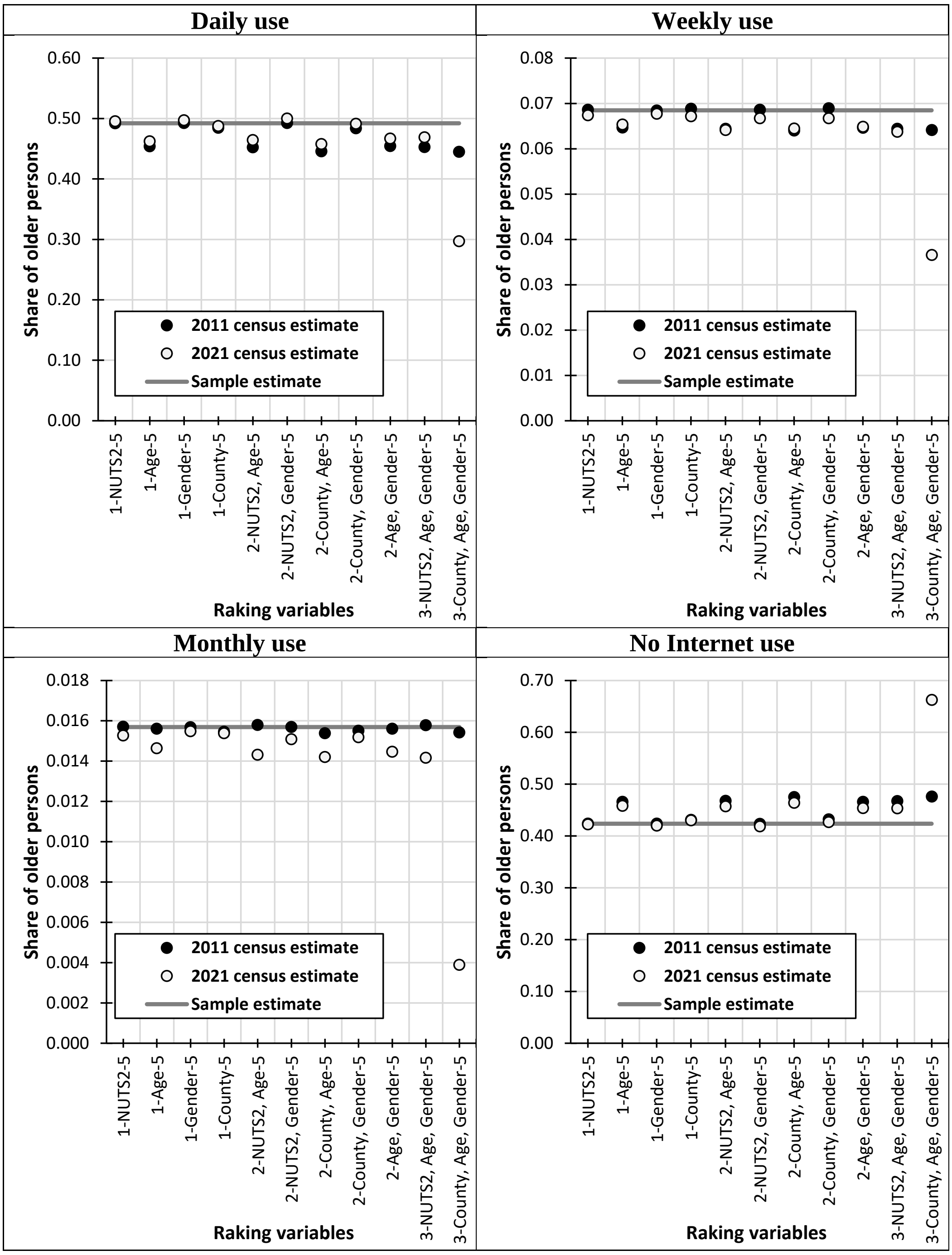

Figure 4. Point estimates of variable "Internet use" using different raking variables, number of raking iterations = 5, n = 701 respondents

significance level of 0.05. However, there are three marginal cases in which the test statistics equaled -1.9604, resulting in a corresponding p value of 0.0499. This was observed for the variable category "No Internet use" within the "Internet use" variable, and it occurred in all three tested numbers of applied raking iterations (5, 10, and 20). In summary, it can be concluded that there are no statistically significant differences in point estimates between the sample data and raking data based on the 2011 census data across all observed variables and their items, for all three observed numbers of raking iterations, and for all 11 combinations of variables used in the raking analysis. Therefore, it can be inferred that, in this case, the raking analysis was not necessary, as the original sample data already yielded virtually identical results.

However, as mentioned earlier, the survey was conducted in January 2022. Basing the analysis on data from more than 10 years ago is not ideal. Therefore, the test statistics shown in Figure 6 indicate the differences between the original sample point estimates and the point estimates calculated in the raking analysis using 2021 census data.

According to Figure 6, there are significantly more differences in point estimates when comparing sample point estimates to those calculated in the raking analyses based on 2021 census data, compared to the analyses using 2011 census data. The first statistically significant difference in point estimates can be found in the variable "Willingness to work again" and its categories "Yes" ($z = -8.97$) and "No" ($z = 8.97$) when NUTS2 and Age variables were used as raking variables in the analysis with 10 iterations. However, this difference might have occurred because raking was successful for other variables while leaving this variable unbalanced during this step. Notably, in the raking analyses with 5 and 20 iterations, this difference did not appear to be statistically significant.

On the other hand, in the raking analyses where raking variables were County, Age, and Gender, it is evident that there are more statistically significant differences in point estimates between the sample point estimates and point estimates calculated in the raking analyses based on the 2021 census data. This difference was consistent across all three raking iterations (5, 10, and 20). Specifically, statistically significant differences appeared for the following variables and their categories:

- "Physical activity" (categories: "Yes"; "No");
- "Internet use" (categories: "Daily use"; "Weekly use"; "Monthly use"; "No Internet use");
- "Willingness to work again" (categories: "Yes"; "No");
- "Feeling of loneliness" (categories: "I never feel lonely"; "Several times a year"; "Several times a month"; "Daily");
- "Missing funds for life" (categories: "I have enough resources to meet my basic living needs"; "Up to 500 HRK"; "Between 1,001 and 1,500 HRK"; "Between 2,001 and 3,000 HRK"; "Between 3,001 and 4,000 HRK").

In summary, significant differences in point estimates between the original sample and raking data were observed for five out of eight observed variables when the 2021 census data were used.

When examining the characteristics of the raking analyses that led to such a high number of statistically significant differences, it's apparent that this particular analysis required the estimation of the most weights (332) compared to the other conducted raking analyses. The number of estimated weights was nearly half of the total sample size (701). Therefore, it's not surprising that the differences in weight values were the most pronounced when compared to all the other conducted raking analyses.

To create a sample of respondents, their structure was initially determined based on

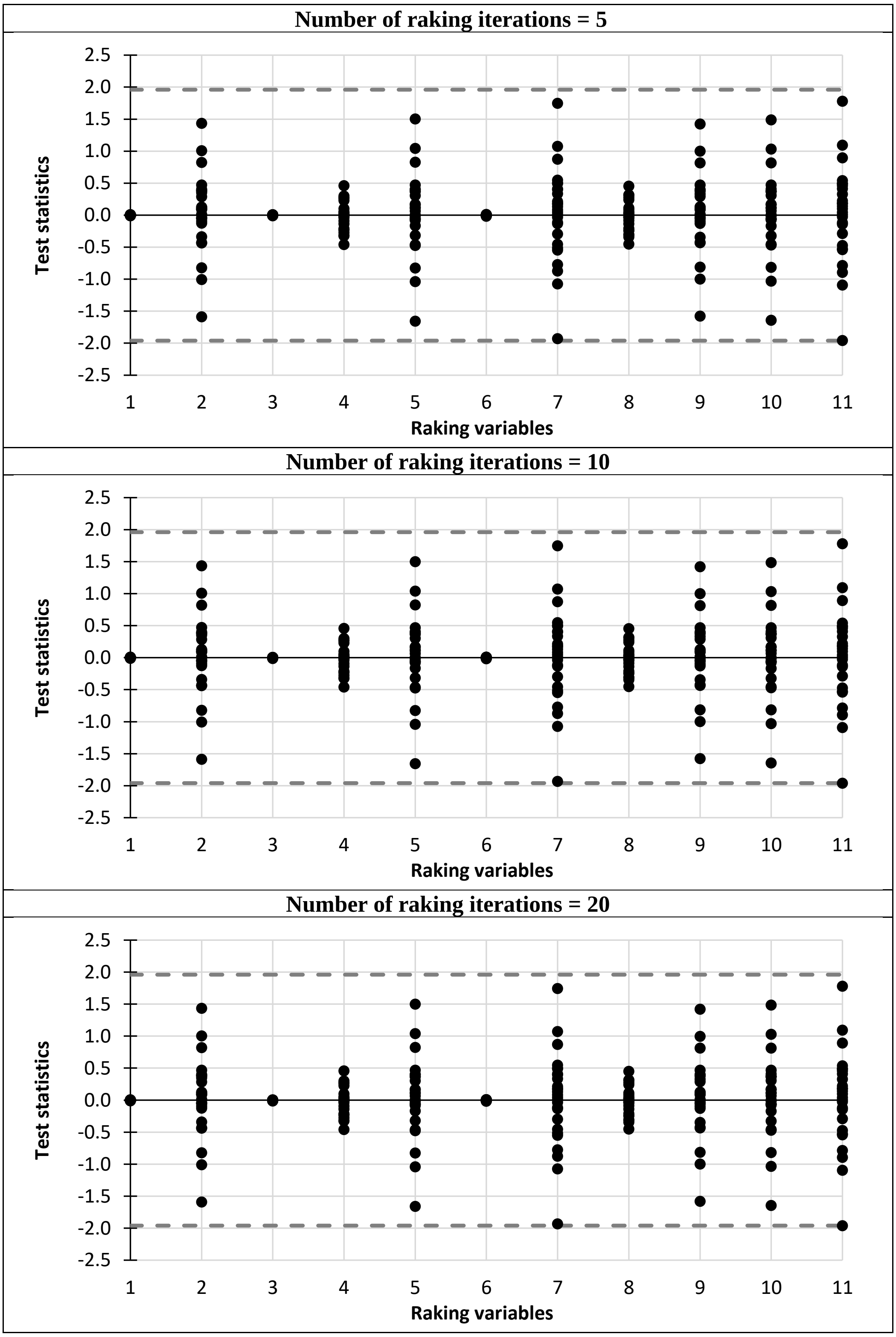

Figure 5. Test statistics for the conducted statistical tests of differences between two proportions and between two averages, sample vs 2011. Each dot represents one test statistic. Raking variables: 1 = NUTS2, 2 = Age, 3 = Gender, 4 = County, 5 = {NUTS2, Age}, 6 = {NUTS, Gender}, 7 = {County, Age}, 8 = {County, Gender}, 9 = {Age, Gender}, 10 = {NUTS2, Age, Gender}, 11 = {County, Age, Gender}.

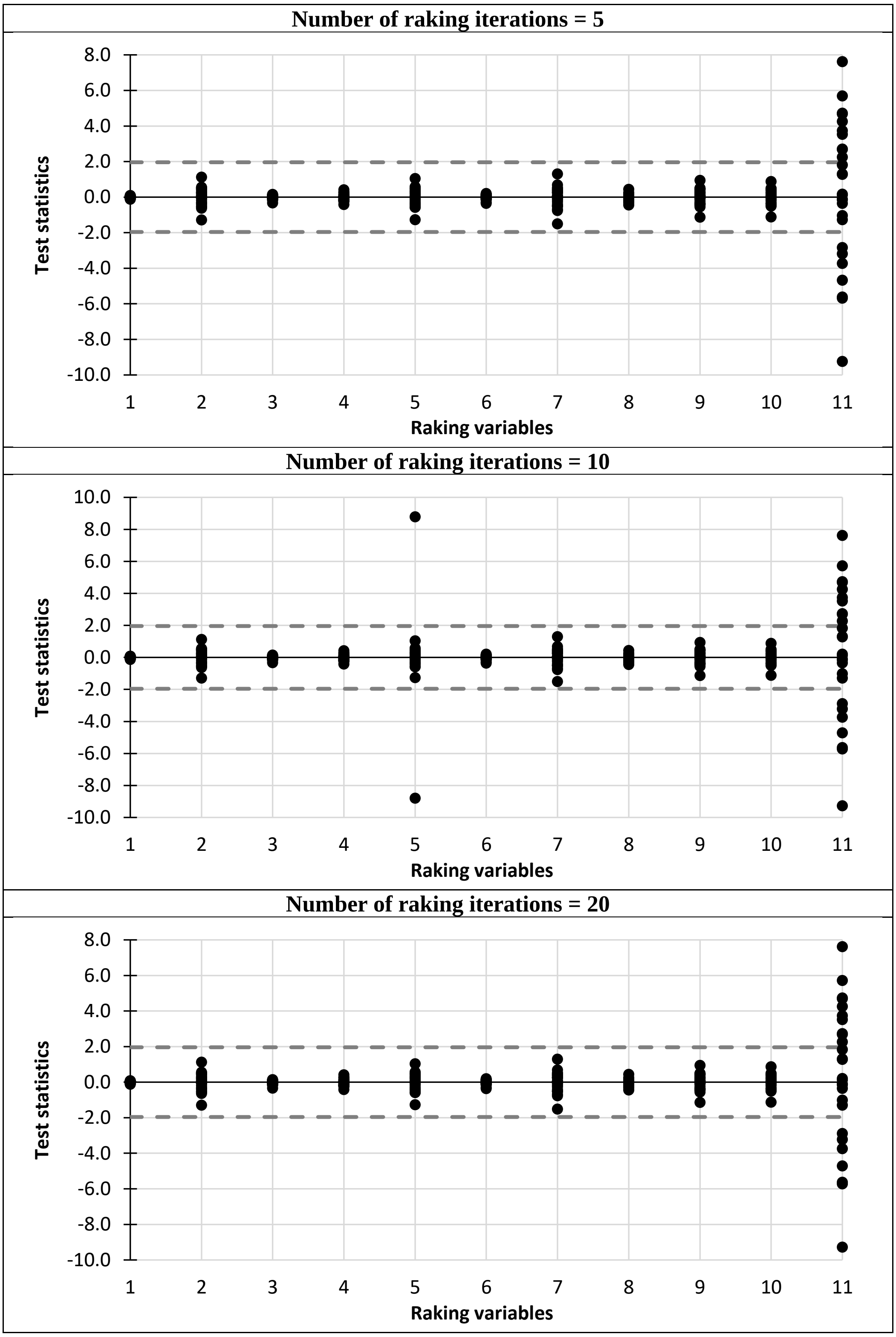

Figure 6. Test statistics for the conducted statistical tests of differences between two proportions and between two averages, sample vs 2021. Each dot represents one test statistic. Raking variables: 1 = NUTS2, 2 = Age, 3 = Gender, 4 = County, 5 = {NUTS2, Age}, 6 = {NUTS, Gender}, 7 = {County, Age}, 8 = {County, Gender}, 9 = {Age, Gender}, 10 = {NUTS2, Age, Gender}, 11 = {County, Age, Gender}.

the 2011 census data for variables Gender and NUTS 2. Subsequently, the raking analysis included different variables, specifically County, Age, and Gender. Additionally, the structure from the 2021 census was considered. These various factors could account for the emergence of statistically significant differences in point estimates.

4. Conclusions

While raking is a valuable method for enhancing the accuracy and precision of survey estimates, the conducted analyses have shown that in the case of quota sampling, raking is less effective in improving the estimates. The analysis results lead to rejection of both research hypotheses. To be precise, in only a very small number of cases (56 out of a total of 1782, or 3.14 %), there was a statistically significant difference between the point estimates of the original sample and the point estimates after conducting raking analyses.

Of those 56 cases, three were observed when comparing the original sample data with the raking data based on the 2011 census. Notably, more statistically significant differences in point estimates were detected when comparing the original sample data with data from raking analyses based on the 2021 census. However, this is still far from the majority, as posited in the second research hypothesis (53 cases out of 891, or 5.95 %). These results also shed light on why raking required only a small number of iterations to reach equilibrium in the sample for the selected raking variables.

This paper presents a preliminary examination of raking within the context of quota sampling, focusing on two census variables. The primary analysis encompassed various characteristics of sampling variables, different numbers of raking iterations, diverse raking variables, and distinct census data used in the raking analyses. However, additional investigations are possible. Firstly, in this study, it was decided not to trim raking weights. Additionally, only point estimates were considered in this paper. The inclusion of confidence intervals may offer valuable supplementary information. Lastly, the paper employed three different numbers of raking iterations as a starting point. In future research, it is advisable to establish a specific tolerance level to determine the appropriate number of raking iterations.

Funding

This research has been conducted within the project Senior 2030—Thematic Network for Active Ageing Policy in Croatia funded by the European Social Fund.

References

Anderson, L., & Fricker Jr, R. D. (2015). Raking: An important and often overlooked survey analysis tool. *Phalanx*, *48*(3), 36–42.

Battaglia, M. P., Hoaglin, D. C., & Frankel, M. R. (2009). Practical considerations in raking survey data. *Survey Practice*, *2*(5), 1–11. https://doi.org/10.29115/SP-2009-0019

Battaglia, M. P., Izrael, D., Hoaglin, D. C., & Frankel, M. R. (2004). Tips and tricks for raking survey data (a.k.a. sample balancing). *Annual Conference of the American Association for Public Opinion Research*, 4740–4745.

Biffignandi, S., & Bethlehem, J. (2021). *Handbook of web surveys* (2nd ed.). Wiley.

Croatian Bureau of Statistics. (2021). *Važnost točnog prijevoda u međunarodnim popisima stanovništva* [The importance of accurate translation in international censuses]. Retrieved December 27, 2024, from https://popis2021.hr/popisni-upitnik-engleski.html

Croatian Bureau of Statistics. (2022). *Popis stanovništva, kućanstava i stanova 2021 (Prvi rezultati)* [Census of population, households and dwellings in 2021 (First results)] [Data Set]. Retrieved December 27, 2024, from https://podaci.dzs.hr/media/0y5d0lzh/popis-2021-prvi-rezultati.xlsx

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Wiley.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2013). *Applied survey data analysis*. Chapman & Hall.

Izrael, D., Hoaglin, D. C., & Battaglia, M. P. (2004). To rake or not to rake is not the question anymore with the enhanced raking macro. *Annual SAS® Users Group International Conference*, 1–11. https://support.sas.com/resources/papers/proceedings/proceedings/sugi29/207-29.pdf

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, *19*(2), 81–97.

Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, *14*(1), 22–59. https://doi.org/10.1177/1536867X1401400104

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. Springer.