Raking Method as a Tool for Improving Representativeness in Non-Probability Studies

Víctor Juan Vera-Ponce[1,2,*], Fiorella E. Zuzunaga-Montoya[3], Nataly Mayely Sanchez-Tamay[1,2], Lupita Ana Maria Valladolid-Sandoval[1,2], Jhosmer Ballena-Caicedo[1,2], Juan Carlos Bustamante-Rodríguez[1,2], Angie Chuquimbalqui Coronel[1,2], Christian Humberto Huaman-Vega[1,4] and Carmen Inés Gutierrez De Carrillo[1,2]

[1]*Instituto de Investigación de Enfermedades Tropicales, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Amazonas, Perú*

[2]*Facultad de Medicina (FAMED), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Amazonas, Perú*

[3]*Universidad Continental, Lima, Perú*

[4]*Instituto de Investigación de Salud Integral Intercultural, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Amazonas, Perú*

Abstract: This is a methodological review focused on raking, or iterative proportional fitting, as a tool for improving representativeness in studies with non-probability sampling. The paper synthesizes the theoretical foundations, practical considerations, and applications of raking in biomedical research. The method operates by iteratively adjusting sample weights so that the marginal distributions of selected variables match the known distributions of the target population. Its implementation requires reliable auxiliary information about the population of interest and careful selection of adjustment variables.

The review addresses critical aspects such as weight quality evaluation, management of extreme values, and computational considerations in raking implementation. The method's advantages are discussed, including its capacity to simultaneously adjust multiple variables and its applicability when only marginal information about the population is available. Its limitations are also examined, such as the potential generation of extreme weights and dependence on precise population data. Finally, practical examples are presented in various contexts, from hospital studies to research in university populations, demonstrating the method's versatility. The application of raking has proven particularly valuable in epidemiological and health services studies, where non-probability samples are common. This review provides a comprehensive methodological guide for researchers seeking to implement raking, emphasizing the importance of rigorous application and transparent documentation.

Keywords: Sampling Studies, Analytic Sample Preparation Methods, Biostatistics, Epidemiology.

INTRODUCTION

In recent decades, non-probability sampling has become increasingly prevalent in various disciplines, particularly health sciences and social research [1]. This shift from traditional probability sampling methods has been driven by multiple factors, including cost constraints, logistical challenges, and the need for rapid data collection in dynamic populations [2]. However, while non-probability sampling offers practical advantages, it introduces significant challenges regarding representativeness and potential selection bias, which can compromise the validity and generalization of research findings [3].

Researchers have developed various statistical approaches for creating sample weights that adjust for selection bias in non-probability samples to address these methodological challenges. This paper is a methodological review that aims to synthesize and explain the theoretical basis, implementation procedures, and practical considerations of one of the most widely used post-sampling adjustment techniques: raking. These methods seek to improve the representativeness of study samples by adjusting for known differences between the sample and the target population. Thus, a prominent approach has emerged as particularly valuable: raking (also known as iterative proportional fitting) [4,5].

The effectiveness of this weighting method depends critically on the availability and quality of auxiliary information about the target population and the selection process. This is because raking requires knowledge of the population marginal distributions for key variables. Recent methodological advances have expanded these approaches to incorporate multiple data sources and complex sampling designs,

*Address correspondence to this author at the Instituto de Investigación de Enfermedades Tropicales, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Amazonas, Perú;
Email: vicvepo@gmail.com

E-ISSN: 1929-6029/25

enhancing their utility in contemporary research settings [6].

Despite their potential benefits, applying sample weights in non-probability samples presents several challenges that researchers must carefully consider. These include handling extreme weights, selecting appropriate auxiliary variables, and evaluating weight [7]. Additionally, the impact of weighting adjustments on variance estimation and statistical inference requires careful consideration, as inappropriate application of weights can potentially increase rather than decrease total survey error [8].

Given the growing reliance on non-probability sampling in research and the complexity of available weighting methods, there is a clear need for a comprehensive understanding of these approaches and their appropriate application. This review aims to synthesize current knowledge about creating sample weights for non-probability samples, examine methodological developments, and provide practical guidelines for researchers. We focus on theoretical foundations, implementation challenges, and empirical evidence on the effectiveness of different weighting approaches in various research contexts.

TYPES OF NON-PROBABILITY SAMPLING

Understanding the fundamentals of non-probability sampling first requires establishing the different classifications of sampling methods in scientific research. According to Lohr [9], sampling methods can be categorized according to various criteria, with the most fundamental distinction being probability and non-probability sampling. In probability sampling, each unit has a known and non-zero probability of being selected, while in non-probability sampling, this probability is unknown or not controlled by the researcher.

The complexity of modern sampling designs is reflected in the different approaches that can be adopted. Moreover, sampling can be conducted in a single stage, where study units are directly selected, or in multiple stages, where selection occurs sequentially through different hierarchical levels. Additionally, the sampling frame can be based on complete lists of units or geographic areas, each approach with its methodological advantages and limitations [10].

The types of non-probability sampling have been extensively documented in the methodological literature, each with specific characteristics and applications. Convenience sampling, one of the most used in biomedical research, is characterized by selecting participants based on their accessibility and proximity to the researcher. While this method offers practical advantages regarding time and resources, there are inherent limitations regarding representativeness and potential selection bias [11].

Consecutive sampling, frequently employed in clinical studies, involves recruiting all individuals from the accessible population who meet the selection criteria during a specific period. Thus, this method can approximate probability sampling when extended over a sufficiently long period and under stable conditions, although it maintains the inherent limitations of non-probability methods [12].

Purposive or judgment sampling represents a more directed approach, where participants are deliberately selected for their specific characteristics. This method is particularly useful in qualitative research and specific case studies, where the objective is to obtain rich and deep information rather than achieve statistical representativeness. This approach requires substantial knowledge of the studied phenomenon and the target population [13].

Quota sampling attempts to replicate the proportions of specific population characteristics in the sample. This method ensures that certain population subgroups are represented in the desired proportions, although selection within each quota remains non-probabilistic. This method is common in market research and social studies, where maintaining certain demographic proportions is sought [14].

Snowball or chain-referral sampling is especially valuable for accessing hidden or hard-to-reach populations. This method uses the social networks of initial participants to recruit new subjects and is particularly useful in studies on marginalized or stigmatized populations. A more sophisticated variant is respondent-driven sampling, which incorporates mathematical procedures to adjust for selection biases inherent to the snowball method [15].

Finally, saturation sampling, primarily used in qualitative research, continues the data collection process until no substantial new information is obtained. This method requires continuous data analysis during the collection process to determine the point of theoretical saturation. Although it does not seek statistical representativeness, this method is valuable for ensuring a deep understanding of the studied phenomenon [16].

FOUNDATIONS OF REPRESENTATIVENESS: RANDOMNESS VERSUS RANDOM ASSIGNMENT

A fundamental aspect of modern scientific research is the distinction between random sampling and random assignment, concepts that, although related, fulfill different functions in the research process [17]. Random sampling refers to the process by which we select units from a population to form our study sample. In this process, each unit in the population has a known and non-zero probability of being selected [9]. For example, in a study of university students, random sampling would imply that each student at the university has the same probability of being selected to participate in the study. The importance of random sampling lies in its ability to ensure statistical representativeness: the characteristics of the sample tend to reflect the characteristics of the population from which it was drawn. This allows valid inferences about the population based on sample data, what we know as external validity or generalizability [18].

On the other hand, random assignment, also known as randomization, is a distinct process after the sample is selected. It refers to how already selected participants are distributed among different study groups, typically treatment and control groups [19]. For example, in a clinical trial, random assignment would randomly determine who receives the experimental treatment and the placebo once we have our participants. The main purpose of random assignment is to create groups comparable in all relevant characteristics, measured and unmeasured. This comparability is crucial for establishing causal relationships, as it ensures that the only systematic difference between groups is the intervention under study.

The distinction between these two processes is crucial because each addresses different types of validity in research. Random sampling primarily relates to external validity: can we generalize our findings to the broader population? In contrast, random assignment relates to internal validity: can we establish valid causal relationships between our variables of interest? [20].

Combining these two types of randomization generates different scenarios with distinct inferential capabilities. When we have both random sampling and random assignment, we can make valid causal inferences and generalize them to the population of interest. This is the ideal scenario, although it is not always practical. A study with random sampling without random assignment can generalize its results to the population, but only in terms of associations, not causality. On the other hand, a study with random assignment without random sampling can establish valid causal relationships. Still, these strictly apply only to the population from which the sample was drawn [21].

In modern research practice, the absence of complete randomization has led to the development of sophisticated methods to "emulate" experimental conditions. Indeed, methods for adjusting for confounders, including propensity score matching and multivariable regression models, can help approximate the conditions of a controlled experiment. These methods and inverse probability weighting techniques have become essential tools for strengthening the validity of inferences in observational studies and non-probability sampling [22].

Thus, evaluating representativeness in non-probability sampling requires careful consideration of multiple factors. This includes comparison with population data when available, conducting sensitivity analyses to assess the robustness of conclusions, and external validation of results. These procedures are fundamental for establishing the validity and scope of findings derived from non-probability samples.

Table 1: Differences between Random Sampling and Random Assignment

	Random Assignment	Non-Random Assignment
Random Sampling	Valid causal inference Generalization to the population Maximum internal and external validity	It does not allow direct causal inference Allows generalization of associations Good external validity
Non-Random Sampling	Valid causal inference Generalization is limited to the sample Good internal validity	Limited causal inference Limited generalization Requires additional assumptions

THE NEED FOR POST-SAMPLING ADJUSTMENT

Post-sampling adjustments have emerged as a crucial methodological necessity in contemporary research, particularly in contexts where probability sampling is not feasible or when there are deviations from ideal sampling designs. This need has intensified in recent decades due to various factors, including increasing survey non-response rates and logistical challenges in implementing traditional probability sampling designs [1].

The practical limitations of modern research frequently lead to samples that do not adequately reflect the structure of the target population. Nevertheless, even in well-designed studies, representativeness can be compromised by differential non-response bias, incomplete coverage of the sampling frame, and selective participation. These problems are particularly relevant in biomedical and epidemiological research, where participant characteristics may systematically differ from the general population's [23].

The need for post-sampling adjustments becomes more evident when considering the implications of lack of representativeness in estimates. Furthermore, an unadjusted forecast can lead to biased conclusions, especially when important differences between the sample and the population in key variables exist. This bias can be particularly problematic in studies that seek to inform public policies or clinical decisions, where external validity is crucial [8].

Advances in statistical methodology have provided sophisticated tools to address these challenges. These post-sampling adjustment methods can help "recalibrate" the sample to reflect the target population better. These adjustments are especially important in scenarios where auxiliary information about the population is available, but the obtained sample does not adequately reflect known population distributions [24].

Indeed, developing methods such as raking responds directly to this need. These methods allow the incorporation of external information to improve sample representativeness, provided that reliable data on relevant population characteristics are available. However, it is crucial to recognize that these adjustments are not a panacea; their effectiveness depends on the quality of available auxiliary information and the validity of underlying assumptions [4,5].

The implementation of post-sampling adjustments also requires careful consideration of the trade-offs involved. While these adjustments can reduce bias in estimates, they may also increase variance, especially when the resulting weights are highly variable. Therefore, the decision to apply post-sampling adjustments must balance bias reduction with the potential increase in the variability of estimates [7].

It is important to emphasize that the need for post-sampling adjustments should not be interpreted as a justification for poor study designs. Rather, quite the contrary, as these methods should be seen as complements, not substitutes, for rigorous research design [6]. Careful planning of the sampling design and implementation of strategies to minimize non-response and other biases remain fundamental to research quality.

METHODS OF CREATING SAMPLE WEIGHTS: RAKING (ITERATIVE PROPORTIONAL FITTING)

Raking, also known as iterative proportional fitting, represents one of the most robust methods for creating sample weights in contemporary research [25]. This method has evolved significantly with modern computational capabilities. The procedure iteratively adjusts sample weights so that the marginal distributions of selected variables match the known distributions of the target population [4,5].

The theoretical foundation of raking is based on the premise that it functions as a "balance adjustment" process, where we seek to make our sample more similar to the population we want to study. Imagine we are studying university students and know that the university has 60% men and 40% women, but in our sample, we have 70% men and 30% women. What raking does is "give more weight" to the responses of women and "less weight" to those of men so that when we analyze the data, the proportions more closely resemble the reality of the university.

It is as if we were giving different "importance" to the responses: if we have fewer women in our sample than in the actual population, their responses "will count more" to compensate for this underrepresentation. The process is repeated with several characteristics simultaneously (for example, age, faculty, and year of study) until our weighted sample better reflects the characteristics of the entire university population.

The key is having reliable information about what the population is like (for example, data from university records) to make these adjustments correctly. It is like having a "photograph" of how our sample should look and adjusting our data to make it more similar to that photograph.

SELECTION OF VARIABLES

The selection of variables for raking must be carefully aligned with the study objectives and the planned analytical approach. In studies with a descriptive focus, where the main aim is to characterize disparities or inequities of a phenomenon in the population, the variables for raking should include fundamental sociodemographic characteristics that can influence the sample's representativeness. For example, in population studies of disease prevalence, age, sex, socioeconomic level, and geographic location are essential for adjustment, as these characteristics typically influence participation patterns and can affect prevalence estimates [26].

In studies with a predictive focus, where the objective is to develop or validate prediction models, the variables selected for raking should include those known to be associated with the outcome of interest, which could be imbalanced in the sample. It is suggested that, in these cases, in addition to basic sociodemographic variables, known risk factors and variables that influence the sample selection process should be considered. However, it is crucial to avoid including the raking variables that will be used as main predictors in the model, as this could introduce artificial biases in the model's predictive capacity [27].

In studies with an explanatory or causal focus, the selection of variables for raking should be guided by the underlying causal structure of the problem under study. The priority variables for adjustment act as potential confounders or effect modifiers in the relationship of interest. These variables must be associated with the probability of selection in the sample and the exposure or outcome of interest. However, avoiding including the raking variables in the causal pathway between the exposure and the outcome is fundamental, as this could introduce biases in estimating the causal effect. Additionally, the possible interaction between these variables should be considered and ensured that the raking adjustment does not distort the causal relationships of interest [18].

ADVANTAGES AND LIMITATIONS

The raking method presents significant advantages that contribute to its widespread adoption in modern research. One of its main strengths lies in its ability to simultaneously adjust multiple variables while preserving the relationships between variables in the original sample. The method is particularly valuable when only marginal information about the population is available, a common situation in many research contexts [23].

However, the method also presents important limitations that must be carefully considered. A significant challenge is the possible generation of extreme weights, especially when substantial discrepancies exist between the sample and the population in some categories. Additionally, it should be noted that the method does not guarantee the representativeness of interactions between variables, which can be problematic when these interactions are of substantive interest to the study [23].

PRACTICAL CONSIDERATIONS

The successful implementation of raking requires careful attention to various practical aspects that can significantly affect the quality and usefulness of the generated weights. The first fundamental aspect is the evaluation of weight quality, which is essential to ensure the validity of subsequent analyses. This evaluation should include a detailed examination of the weight distribution, paying special attention to their variability and outliers. Researchers should verify that the weighted marginal distributions effectively match the target population distributions within specified tolerances and that the resulting weights are stable and consistent with the study design [28].

The management of extreme values in the weights represents another significant practical challenge. It has been suggested that large or small weights can arise when important discrepancies exist between the sample and the population in specific categories. To address this problem, various weight truncation and smoothing strategies have been developed. A common approach is establishing upper and lower limits for the weights, typically based on percentiles of their distribution or multiples of their mean. Therefore, it is recommended that any adjustment of extreme values be carefully documented, and its implications for the final estimates should be evaluated through sensitivity analysis [28,29].

In addition to truncation, other strategies such as weight trimming, which involves setting bounds to remove or reduce the influence of outliers, and post-stratification, which adjusts weights based on full crossclassifications rather than marginal distributions, can also be considered. These approaches may offer improved control over weight variability and enhance the robustness of statistical estimates, particularly in complex or highly unbalanced samples.

The availability of appropriate statistical software is crucial for the effective implementation of raking. Modern statistical programs such as R offer specific functions to perform raking adjustments. For example, both the 'anesrake' package and the 'survey' package in R provide robust tools for implementing raking and analyzing data with complex weights. While several statistical programs have this function included, the choice of software should be based on the specific needs of the study, the research team's familiarity with the different tools, and the software's ability to handle the size and complexity of the data [30].

EXAMPLES OF RAKING USE IN NON-PROBABILITY SAMPLING

Example 1: Factors Related to Anxiety during COVID-19

The practical application of raking can be illustrated with various cases documented in recent literature. The study on anxiety during the COVID-19 pandemic in Peru provides an excellent example of how to apply raking in a real investigation. The researchers faced a common challenge in studies with online surveys: their sample was not representative of the Peruvian population. This problem became evident when comparing the characteristics of survey respondents with data from the 2017 National Census.

In this particular case, the researchers identified four key variables available in both their survey and the census: age (categorized into groups: 18–24, 25–34, 35–44, 45–54, 55+ years), sex (male, female), educational level (incomplete secondary or less, complete secondary, undergraduate or more), and region (Lima, Rest of Coast, Andes, Jungle). It is important to note that these variables may vary according to the study and context; they must be available both in your sample and in a reliable source of population data, such as a census or a representative national survey.

The implementation of raking in R began with creating an initial unit weight and establishing the survey design using the svydesign() function. Subsequently, the known population distributions for

```r
# First we load the database (in this case it's called "data")
# We create the weights
data$initial_weight <- 1
design <- svydesign(ids = ~1, weights = ~initial_weight, data = data)

# We need to specify the census distributions (or the database where we have the data)
# For age
pop.age <- data.frame(
  agecat = c("0", "1", "2", "3", "4"),
  Freq = c(0.1725, 0.2301, 0.2020, 0.1591, 0.2362))

# For sex
pop.sex <- data.frame(
  sex = c("0", "1"),
  Freq = c(0.4847, 0.5153))

# For region
pop.region <- data.frame(
  region = c("0", "1", "2", "3"),
  Freq = c(0.3472, 0.2448, 0.3260, 0.0820))

# Now the raking
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~region),
                        population.margins = list(pop.age, pop.sex, pop.region))

# To extract the weights after raking
raking_weights <- weights(adjusted_design)

# Add the weights to your original database
data$raking_weight <- raking_weights

# Now the raking with upper limit
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~region),
                        population.margins = list(pop.age, pop.sex, pop.region),
                        control = list(maxit = 50,       # maximum iterations
                                       epsilon = 1e-7,    # tolerance
                                       bound = 3500))     # upper limit for weights
```

Figure 1: Implementation of the raking method in R, for example, 1: specification of marginal distributions and iterative adjustment process.

the adjustment variables (age, sex, and region) were specified using data frames containing the categories and their respective marginal frequencies. For age, five groups were established with population proportions ranging from 0.1725 to 0.2362; for sex, a binary distribution (0.4847 and 0.5153); and four categories with proportions from 0.0820 to 0.3472 for the region. The raking process was executed using the rake() function, specifying both the sample variables (sample.margins) and the population distributions (population.margins).

Two versions were implemented: a basic one and another with additional control parameters, including an upper limit of 3500 for the weights, a maximum of 50 iterations, and a convergence tolerance of 1e-7. These specific parameters were selected to ensure computational stability and prevent the generation of excessively large weights, which could disproportionately influence the results and inflate variance. The iteration limit and convergence threshold were chosen to guarantee marginal alignment without risking non-convergence or unnecessary computational burden. The resulting weights were extracted using the weights() function and incorporated into the original database for use in subsequent analyses.

Once these weights were calculated, they were used in all subsequent analyses. In this case, two Poisson regressions with robust variance were applied, one unweighted and one weighted. The application of raking revealed important differences in the studied associations. In the unweighted analysis, the male sex showed a significantly protective association (aPR=0.61; 95%CI:0.46-0.80; p<0.001), while after raking, this association was attenuated and lost statistical significance (aPR=0.65; 95%CI:0.42-1.00; p=0.051). Regarding age, both analyses showed positive associations for the 30-39 age group, but the effect in the 40-49 age group became stronger and

Table 2: Differences between Unweighted and Weighted Robust Variance Poisson Regression Analysis of Determinants of Anxiety During COVID-19

Characteristics	Unweighted			Weighted		
	aPR	95% CI	p-value	aPR	95% CI	p-value
Sex						
Female	—	—		—	—	
Male	0.61	0.46, 0.80	<0.001	0.65	0.42, 1.00	0.051
Age groups						
18 - 29 years	—	—		—	—	
30 - 39 years	1.8	1.29, 2.53	<0.001	1.73	1.11, 2.71	0.016
40 - 49 years	1.73	0.90, 3.37	0.103	2.72	1.25, 5.92	0.012
50 - 59 years	2.95	1.40, 6.69	0.006	1.74	0.70, 4.35	0.234
60 years and older	1.46	1.01, 2.14	0.047	1.42	0.87, 2.32	0.156
Region*						
Metropolitan Lima	—	—		—	—	
Rest of Coast	1.08	0.70, 1.69	0.722	1.48	0.79, 2.77	0.225
Jungle	1.72	0.95, 3.23	0.079	1.83	0.72, 4.63	0.201
Highlands	0.59	0.42, 0.82	0.002	0.68	0.40, 1.14	0.139
Working?*						
No	—	—		—	—	
Yes	1.19	0.88, 1.61	0.248	1.91	1.23, 2.96	0.004
Living alone?*						
No	—	—		—	—	
Yes	0.76	0.47, 1.24	0.271	0.7	0.33, 1.47	0.344

*Each variable was adjusted for age and sex.
aPR: adjusted Prevalence Ratio.
95% CI: 95% confidence interval.
Source: authors' elaboration.

```r
# First we load the database (in this case it's called "data")
# We create the weights
data$initial_weight <- 1
design <- svydesign(ids = ~1, weights = ~initial_weight, data = data)

# We need to specify the census distributions (or the database where we have the data)
# For age
pop.age <- data.frame(
  agecat = c("0", "1", "2"),
  Freq = c(0.35, 0.45, 0.20))

# For sex
pop.sex <- data.frame(
  sex = c("0", "1"),
  Freq = c(0.45, 0.55))

# For faculty
pop.faculty <- data.frame(
  faculty = c("0", "1", "2", "3"),
  Freq = c(0.30, 0.20, 0.25, 0.25))

# For work
pop.works <- data.frame(
  works = c("0", "1"),
  Freq = c(0.30, 0.70))

# Now the raking
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~faculty, ~works),
                        population.margins = list(pop.age, pop.sex, pop.faculty, pop.works))

# To extract the weights after raking
raking_weights <- weights(adjusted_design)

# Add the weights to your original database
data$raking_weight <- raking_weights

# The rest of the code remains the same until the raking
# Now the raking with upper limit
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~faculty, ~works),
                        population.margins = list(pop.age, pop.sex, pop.faculty, pop.works),
                        control = list(maxit = 50,       # maximum iterations
                                       epsilon = 1e-7,    # tolerance
                                       bound = 3500))     # upper limit for weights
```

Figure 2: Implementation of the raking method in R, for example 2.

significant only after weighting (aPR=2.72; 95%CI:1.25-5.92; p=0.012). Notably, the significant association with residing in the Sierra region (aPR=0.59; 95%CI:0.42-0.82; p=0.002) in the unweighted analysis disappeared after raking (aPR=0.68; 95%CI:0.40-1.14; p=0.139). A particularly relevant finding was that employment status, which showed no significant association in the unweighted analysis (p=0.248), emerged as an important factor after raking, with workers showing a higher probability of the outcome (aPR=1.91; 95%CI:1.23-2.96; p=0.004).

These differences underscore the importance of raking adjustment to obtain more representative estimates of the general population.

Example 2: Anxiety and Irritable Bowel Syndrome

Now, let us consider a study designed to examine the association between anxiety and Irritable Bowel Syndrome (IBS) in university students. In this case, four key auxiliary variables were used for adjustment: age (categorized into three groups with population distributions of 35%, 45%, and 20%), sex (with a population distribution of 45% men and 55% women), faculty (divided into four categories with distributions of 30%, 20%, 25%, and 25%), and employment status (30% working, 70% not working). Additionally, information on the prevalence of anxiety (25%) was available from a previous mental health campaign conducted at the university.

It is important to mention that implementing raking in this context requires careful consideration of the auxiliary variables to include. While data on the prevalence of anxiety in the student population is available, this variable should not be included in the raking process, as it constitutes a primary outcome variable in the study. Including outcome variables in the adjustment process could introduce artificial biases in the associations of interest. Therefore, raking is implemented using only the sociodemographic variables: age, sex, and employment status.

The technical implementation of raking began with assigning initial unit weights to all observations, thus establishing a neutral starting point for the adjustment process. Population distributions were specified using data frames in R, providing a clear structure for each auxiliary variable. The adjustment process was performed using the rake() function, which iteratively adjusted the weights so that the marginal distributions of the sample matched the known population distributions while maintaining the natural relationship between anxiety and IBS.

An important technical aspect was the inclusion of control parameters in the raking process, establishing an upper limit of 3500 for the weights, a maximum of 50 iterations, and a tolerance of 1e-7. These values were selected to ensure both computational efficiency and statistical robustness. The upper weight limit was applied to avoid extreme weights that could distort the estimates or increase variance excessively. The number of iterations and convergence tolerance were chosen to strike a balance between achieving accurate alignment with the population margins and avoiding overfitting or convergence issues. The generated weights were subsequently incorporated into the regression analyses, allowing the examination of the association between anxiety and IBS to be more representative of the general student population.

In this manner, the comparative analysis between weighted and unweighted results revealed substantial differences in the magnitude of the association between anxiety and IBS. In the unweighted analysis, students with anxiety showed a 65% higher probability of presenting IBS compared to those without anxiety (aPR=1.65; 95%CI:1.35-2.02; p<0.001). However, after applying raking, this association strengthened considerably, showing that students with anxiety had a more than four times greater probability of presenting IBS (aPR=4.39; 95%CI:3.01-6.40; p<0.001).

This substantial difference in magnitudes suggests that unweighted sampling might significantly underestimate the association between these conditions in the general student population. Adjustment through raking, by correcting imbalances in the sociodemographic characteristics of the sample, revealed a stronger association that maintained its statistical significance.

Example 3: Diabetes Mellitus and Cardiovascular Disease

This substantial difference in magnitudes suggests that unweighted sampling might significantly underestimate the association between these conditions in the general student population. Adjustment through raking, by correcting imbalances in the sociodemographic characteristics of the sample, revealed a stronger association that maintained its statistical significance.

To illustrate the application of raking in the hospital context, let us consider a study to examine the association between type 2 diabetes mellitus and cardiovascular disease (CVD) in adult patients. This example leverages a significant advantage of hospital research: access to complete administrative and clinical records. Researchers conducting their survey or primary data collection may encounter difficulties obtaining responses from certain patient groups. For example, elderly patients or those from rural areas might be underrepresented in the final sample. Raking allows for adjusting these imbalances using the available administrative information.

The implementation of raking was performed using three fundamental auxiliary variables available in the hospital's administrative records: age, sex, and place of residence. The age distribution was categorized into three groups, reflecting the demographic structure of the population served at the hospital (25% in the youngest group, 45% in the intermediate group, and 30% in the oldest group). The sex distribution was close to parity (48% men and 52% women), while the residence variable distinguished between urban and rural populations (70% and 30%, respectively), reflecting the hospital's coverage area.

Table 3: Differences between Unweighted and Weighted Robust Variance Poisson Regression on the Association between Anxiety and IBS

Characteristics	Unweighted			Weighted		
	aPR	95% CI	p-value	aPR	95% CI	p-value
Anxiety						
No	Ref.			Ref.		
Yes	1.65	1.35, 2.02	<0.001	4.39	3.01, 6.40	<0.001

Adjusted for sex, categorized age, faculty, alcohol consumption, and smoking activity.
cPR: crude Prevalence Ratio. aPR: adjusted Prevalence Ratio.
95% CI: 95% confidence interval.
Source: authors' elaboration.

```
# After loading the database (called "data")
# We create the weights
data$initial_weight <- 1
design <- svydesign(ids = ~1, weights = ~initial_weight, data = data)

# We need to specify the census distributions (or the database where we have the data
# For age
pop.age <- data.frame(
  agecat = c("0", "1", "2"),
  Freq = c(0.25, 0.45, 0.30))

# For sex
pop.sex <- data.frame(
  sex = c("0", "1"),
  Freq = c(0.48, 0.52))

# For residence
pop.residence <- data.frame(
  residence = c("0", "1"),
  Freq = c(0.70, 0.30))

# Now the raking
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~residence),
                        population.margins = list(pop.age, pop.sex, pop.residence))

# To extract the weights after raking
raking_weights <- weights(adjusted_design)

# Add the weights to your original database
data$raking_weight <- raking_weights

# Now the raking with upper limit
adjusted_design <- rake(design = design,
                        sample.margins = list(~agecat, ~sex, ~residence),
                        population.margins = list(pop.age, pop.sex, pop.residence),
                        control = list(maxit = 50,       # maximum iterations
                                       epsilon = 1e-7,     # tolerance
                                       bound = 3500))      # upper limit for weights
```

Figure 3: Implementation of the raking method in R, for example 3.

It is important to note that implementing raking in this context requires careful selection of auxiliary variables. It is suggested that the selected variables should be related to both the probability of response and the outcome variables of interest. In this case, demographic variables (age, sex) and residence (insurance type, origin) are appropriate candidates for adjustment.

Furthermore, although the hospital may have records on the prevalence of diabetes or cardiovascular disease in its population, these variables should not be included in the raking process, as they are the primary variables of interest in the study. Kalton and Flores-Cervantes (2003) point out that including outcome variables in the adjustment could introduce artificial biases in the studied associations.

Specific controls were implemented in the raking process, including an upper limit of 3500 for the weights, a maximum of 50 iterations, and a tolerance of 1e-7. These parameters are particularly important in the hospital context, where populations may present very heterogeneous characteristics and where the precision of estimates is crucial for clinical decision-making. The upper limit on weights was used to reduce the influence of extreme values that could disproportionately affect the estimates and inflate the variance. The iteration cap and the convergence tolerance were selected to ensure the stability of the algorithm while achieving accurate marginal alignment without unnecessary computational burden or risk of non-convergence.

The weights generated through this process allow subsequent analyses to be more representative of the total population served at the hospital, adjusting for possible selection biases in the initial sample.

The analysis of the association between diabetes mellitus and the outcome of interest showed notable differences between unweighted and weighted results after applying raking. In the unweighted analysis, although a positive trend was observed, the association did not reach statistical significance, showing that patients with diabetes had a 30% higher probability of presenting the outcome compared to those without diabetes (aPR=1.3; 95%CI:0.74-2.17; p=0.34). However, after applying raking to adjust for the

Table 4: Differences between Unweighted and Weighted robust Variance Poisson Regression on the Association between Diabetes and CVD

Characteristics	Unweighted			Weighted		
	aPR	95% CI	p-value	aPR	95% CI	p-value
Diabetes Mellitus						
No	Ref.			Ref.		
Sí	1.3	0.74, 2.17	0.34	3.13	1.61, 6.07	<0.001

Adjusted for sex, categorized age, residence, alcohol consumption, and smoking activity
cPR: crude Prevalence Ratio. aPR: adjusted Prevalence Ratio
95% CI: 95% confidence interval
Source: authors' elaboration

sociodemographic characteristics of the hospital population (age, sex, and place of residence), the association strengthened substantially and reached statistical significance, revealing that patients with diabetes had a more than three times greater probability of presenting the outcome (aPR=3.13; 95%CI:1.61-6.07; $p<0.001$). This substantial difference suggests that unweighted analysis might significantly underestimate the association between diabetes and the outcome in the general hospital population, highlighting the importance of adjustment through raking to obtain more representative estimates.

This hospital example illustrates several advantages of raking in biomedical research: the ability to leverage existing administrative data, the possibility of adjusting for multiple characteristics simultaneously, and the improvement in the representativeness of final estimates. It also demonstrates how the availability of complete hospital records can strengthen the implementation of post-sampling adjustment methods.

IMPLICATIONS FOR PUBLIC HEALTH AND MEDICAL DECISION-MAKING

The implementation of raking techniques in biomedical studies has important implications beyond methodological rigor. In contexts where probability sampling is not feasible—due to urgency, ethical constraints, or limited resources—raking offers a statistically sound alternative to improve representativeness. By aligning key sample characteristics with known population distributions, raking strengthens the validity of prevalence estimates, association measures, and risk profiles derived from non-probability data sources.

For public health surveillance, this means that data obtained through online surveys, rapid assessments, or convenience-based recruitment can still inform policy decisions, provided appropriate adjustments like raking are applied. This is particularly relevant during health emergencies (such as pandemics) or in hard-to-reach populations, where conventional sampling designs may be impractical or impossible.

In clinical epidemiology, raking can help ensure that associations between exposures and outcomes reflect the broader population and not just the study sample. This improves the external validity of research findings and supports the development of guidelines that are equitable and generalizable. For instance, adjusting for underrepresented groups (e.g., older adults, rural populations) ensures that their health risks are adequately captured in the data used for planning interventions.

Ultimately, the appropriate use of raking contributes to more accurate decision-making in both public health and clinical settings. It promotes the responsible use of non-probability data by correcting biases that could otherwise lead to misleading conclusions. As data sources become increasingly diverse and complex, methods like raking will continue to play a key role in bridging the gap between statistical robustness and real-world utility.

STRENGTHS AND LIMITATIONS

The raking method offers several notable strengths, making it particularly valuable in health research using non-probability samples. Its greatest advantage lies in the ability to simultaneously adjust multiple variables while relying only on marginal population distributions, often available from censuses or administrative records. This makes raking highly applicable in epidemiological studies, rapid health assessments, and online health surveys, where full cross-classified data or probability sampling designs may be unavailable.

Raking is more flexible than post-stratification because it does not require population totals for every possible combination of adjustment variables. It also surpasses methods such as propensity score adjustment in contexts where detailed population-level individual data are not accessible. Unlike model-based techniques, raking remains a nonparametric approach, reducing dependency on the correctness of the underlying model and allowing for more transparent and reproducible adjustments.

However, raking is not without limitations. One common issue is the potential generation of extreme weights, particularly when large discrepancies exist between the sample and the population in certain subgroups. These extreme values can inflate the variance of estimates and reduce statistical efficiency. To address this, weight trimming or truncation strategies are often applied, though these require careful justification and sensitivity analysis to avoid bias.

Additionally, while raking aligns marginal distributions, it does not guarantee representativeness in joint distributions or complex interactions between variables. In studies where such interactions are central to the research question, other methods like calibration weighting or multilevel modeling may be more appropriate. Despite these caveats, when applied rigorously and transparently, raking remains one of the most practical and powerful tools available for enhancing representativeness in non-probability health research.

CONCLUSIONS AND RECOMMENDATIONS

The raking method is valuable for improving representativeness in biomedical studies that rely on non-probability samples. Its application is especially relevant in contexts where only marginal information about the target population is available, and where probability sampling is not feasible due to practical constraints. Through illustrative examples in hospital and university settings, this review has demonstrated how raking can contribute to more accurate and generalizable findings in health research.

For the effective implementation of raking in biomedical research, it is recommended to: 1) identify and use reliable sources of population data for marginal distributions, preferably from official records or recent censuses; 2) carefully select auxiliary variables for adjustment, considering both their availability in the sample and their theoretical relevance to the study; 3) implement controls on extreme weights, establishing appropriate upper limits to avoid instability in estimates; 4) perform sensitivity analyses comparing weighted and unweighted results to evaluate the impact of the adjustment; and 5) document in detail the implementation process of raking, including population data sources, convergence criteria used, and any adjustments made to the weights. These methodological considerations are crucial to ensure the validity and reproducibility of studies that employ this post-sampling adjustment method.

ACKNOWLEDGEMENTS

A special thanks to the members of Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Amazonas, Peru, for their support and contributions throughout the completion of this research.

FINANCIAL DISCLOSURE

This study was financed by Vicerectorado de Investigación de la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

INFORMED CONSENT

This study is a review; therefore, informed consent is not required.

DATA AVAILABILITY

Data are available upon request to the corresponding author.

AUTHORS' CONTRIBUTION

Víctor Juan Vera-Ponce: Conceptualization, Investigation, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing

Fiorella E. Zuzunaga-Montoya: Methodology, Software, Data Curation, Formal analysis, Writing - Review & Editing

Nataly Mayely Sanchez-Tamay: Validation, Investigation, Project administration, Writing - Original Draft, Writing - Review & Editing

Lupita Ana Maria Valladolid-Sandoval: Visualization, Investigation, Project administration, Writing - Original Draft, Writing - Review & Editing

Jhosmer Ballena-Caicedo: Investigation, Project administration, Writing - Original Draft, Writing - Review & Editing

Juan Carlos Bustamante-Rodríguez: Investigation, Project administration, Methodology, Writing - Original Draft, Writing - Review & Editing

Angie Chuquimbalqui Coronel: Investigation, Methodology, Writing - Original Draft, Writing - Review & Editing

Christian Humberto Huaman-Vega: Investigation, Methodology, Writing - Original Draft, Writing - Review & Editing

Carmen Inés Gutierrez De Carrillo: Methodology, Resources, Supervision, Funding acquisition, Writing - Review & Editing

REFERENCES

[1] Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, *et al*. Summary Report of the AAPOR Task Force on Non-probability Sampling. J Surv Stat Methodol 2013; 1(2): 90-143. https://doi.org/10.1093/jssam/smt008

[2] Mercer AW, Kreuter F, Keeter S, Stuart EA. Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference. Public Opin Q 2017; 81(S1): 250-71. https://doi.org/10.1093/poq/nfw060

[3] Brick JM. The Future of Survey Sampling. Public Opin Q 2011; 75(5): 872-88. https://doi.org/10.1093/poq/nfr045

[4] Wang VH-C, Lei J, Shi T, Pagán JA. Weighting the United States All of Us Research Program data to known population estimates using raking. Prev Med Rep 2024; 43: 102795. https://doi.org/10.1016/j.pmedr.2024.102795

[5] Wang VH-C, Holm J, Pagán JA. Use calibration to improve the precision of estimates obtained from All of Us data. J Am Med Inform Assoc JAMIA 2024; ocae181. https://doi.org/10.1093/jamia/ocae181

[6] Kim J, Tam S-M. Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference. Int Stat Rev 2020; 89. https://doi.org/10.1111/insr.12434

[7] Bethlehem J. Selection Bias in Web Surveys. Int Stat Rev 2010; 78(2): 161-88. https://doi.org/10.1111/j.1751-5823.2010.00112.x

[8] Valliant R, Dever JA, Kreuter F. Practical Tools for Designing and Weighting Survey Samples [Internet]. Cham: Springer International Publishing; 2018 [citado el 26 de octubre de 2024]. (Statistics for Social and Behavioral Sciences). https://doi.org/10.1007/978-3-319-93632-1

[9] Lohr SL. Sampling: design and analysis. 2nd ed. Boston, Mass: Brooks/Cole; 2010. 596 p.

[10] Encyclopedia of Survey Research Methods [Internet]. SAGE Publications Inc 2024 [citado el 26 de octubre de 2024]. Disponible en: https://us.sagepub.com/en-us/nam/encyclopedia-of-survey-research-methods/book227714

[11] Etikan I, Musa SA, Alkassim RS. Comparison of Convenience Sampling and Purposive Sampling. Am J Theor Appl Stat 2015; 5(1): 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

[12] Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? An Bras Dermatol 2016; 91(3): 326. https://doi.org/10.1590/abd1806-4841.20165254

[13] Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015; 42(5): 533. https://doi.org/10.1007/s10488-013-0528-y

[14] Teddlie C, Yu F. Mixed Methods Sampling: A Typology With Examples. J Mix Methods Res 2007; 1(1): 77-100. https://doi.org/10.1177/1558689806292430

[15] Heckathorn DD. Snowball versus respondent-driven sampling. Sociol Methodol 2011; 41(1): 355. https://doi.org/10.1111/j.1467-9531.2011.01244.x

[16] Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLOS ONE 2020; 15(5): e0232076. https://doi.org/10.1371/journal.pone.0232076

[17] Imbens GW, Rubin DB. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press; 2015 [citado el 26 de octubre de 2024]. https://doi.org/10.1017/CBO9781139025751

[18] Hernan MA, Robins JM. Causal Inference: What If. CRC Press; 2024. 312 p.

[19] Pearl J. Causality: Models, Reasoning, and Inference. Cambridge University Press; 2000. 412 p.

[20] Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Belmont, CA; 2001. 656 p.

[21] VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med 2017; 167(4): 268-74. https://doi.org/10.7326/M16-2607

[22] Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci Rev J Inst Math Stat 2010; 25(1): 1. https://doi.org/10.1214/09-STS313

[23] Kalton G, Flores Cervantes I. Weighting Methods. Httplst-liepiiep-Unescoorgcgi-Binwwwi32exeinepidoc1int2000019622100 2003; 19

[24] Little RJA, Rubin DB. Statistical Analysis with Missing Data. John Wiley & Sons; 2019. 462 p.

[25] Deville J-C, Särndal C-E, Sautory O. Generalized Raking Procedures in Survey Sampling. J Am Stat Assoc. 1993; 88(423): 1013-20. https://doi.org/10.2307/2290793

[26] Groves RM, Peytcheva E. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opin Q 2008; 72(2): 167-89. https://doi.org/10.1093/poq/nfn011

[27] Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. J R Stat Soc Ser A Stat Soc 2001; 174(2): 369. https://doi.org/10.1111/j.1467-985X.2010.00673.x

[28] Battaglia MP, Hoaglin DC, Frankel MR. Practical Considerations in Raking Survey Data. Surv Pract 2009; 2(5). https://doi.org/10.29115/SP-2009-0019

[29] Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a.Raking) | Abt Global 2024. Disponible en: https://www.abtglobal.com/insights/publications/presentation/extreme-survey-weight-adjustment-as-component-of-sample

[30] giscus [Internet]. [citado el 26 de octubre de 2024]. Disponible en: https://giscus.app/en/widget?origin=https%3A%2F%2Fsdaza.com%2Fblog%2F2012%2Fraking%2F&session=&theme=null&reactionsEnabled=1&emitMetadata=0&inputPosition=bottom&repo=sdaza%2Fsdaza.github.io&repoId=MDEwOlJlcG9zaXRvcnk1MzY3MjE3MQ%3D%3D&category=Comments&categoryId=DIC_kwDOA5PmLc4CTBt6&strict=1&description=Demography%2C+Sociology%2C+Simulation%2C+Data+Science&backLink=https%3A%2F%2Fsdaza.com%2Fblog%2F2012%2Fraking%2F&term=Raking+weights+with+R+%7C+Sebastian+Daza

Received on 11-02-2025 Accepted on 16-03-2025 Published on 24-04-2025

https://doi.org/10.6000/1929-6029.2025.14.22