OPEN ACCESS

Received: 18/02/25
Accepted: 04/04/25
Online: 10/04/25

SPECIAL ARTICLE

Sample size and precision calculation for epidemiological studies: development and implementation of the *Calculadora Prevalencia*, an R package

Víctor Juan Vera-Ponce[1,a] | Fiorella E. Zuzunaga-Montoya[2,a] | Christian Humberto Huaman-Vega[1,b] | Nataly Mayely Sanchez-Tamay[1,c] | Carmen Inés Gutierrez de Carrillo[1,a]

[1] Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Amazonas, Perú.
[2] Universidad Científica del Sur, Lima, Perú.
[a] MD.
[b] Psychology student.
[c] Medical student.

Keywords:
sampling; sample size; stratified sampling; random sampling; systematic sampling (source: MeSH-NLM).

ABSTRACT

Determining sample size and assessing precision are essential components of epidemiological research. This article presents *Calculadora Prevalencia,* an R-based tool designed to facilitate these calculations by incorporating both methodological and logistical factors. The calculator supports various scenarios, including finite and infinite populations, stratified sampling, and adjustments for instrument sensitivity and specificity. Its utility is demonstrated through practical examples across diverse contexts: infinite urban populations, finite rural populations, stratified university sampling, and the analysis of existing datasets. The tool also addresses logistical considerations, estimating the number of subjects to contact based on rejection and eligibility rates, and projecting expected fieldwork duration. Its versatility enables both prospective planning and retrospective data evaluation, while the innovative inclusion of logistical components offers a realistic perspective on the resources needed to carry out successful epidemiological studies.

Cálculo de tamaño muestral y precisión para estudios epidemiológicos: desarrollo e implementación del paquete CalculadoraPrevalencia en R

Palabras clave:
muestreo; tamaño de la muestra, muestreo estratificado; muestreo aleatorio; muestreo sistemático (fuente: DeCs-BIREME).

RESUMEN

La determinación del tamaño muestral y la evaluación de precisión son elementos cruciales en investigación epidemiológica. Este artículo presenta CalculadoraPrevalencia, una herramienta en R que facilita estos cálculos incorporando aspectos metodológicos y logísticos. La calculadora maneja diversos escenarios: poblaciones finitas e infinitas, estratificación y ajustes por sensibilidad/especificidad de instrumentos. Mediante ejemplos prácticos en diferentes contextos (población urbana infinita, población rural finita, muestreo universitario estratificado y análisis de datos existentes), se demuestra su aplicabilidad. La herramienta integra consideraciones logísticas, calculando sujetos a contactar según tasas de rechazo/elegibilidad y estimando tiempos de trabajo de campo. Su versatilidad facilita tanto la planificación prospectiva como la evaluación de datos existentes, mientras que la innovadora incorporación de aspectos logísticos proporciona una visión realista de los recursos necesarios para estudios epidemiológicos exitosos.

Cite as: Vera-Ponce VJ, Zuzunaga-Montoya FE, Huaman-Vega CH, Sanchez-Tamay NM, Gutierrez de Carrillo CI. Sample size and precision calculation for epidemiological studies: development and implementation of the *Calculadora Prevalencia*, an R package. Rev Peru Cienc Salud. 2025;7(2):156-63. doi: https://doi.org/10.37711/rpcs.2025.7.2.4

Correspondence:
Víctor Juan Vera-Ponce
vicvepo@gmail.com

© The author. This is an article under the Creative Commons license, CC-BY 4.0

INTRODUCTION

Sample size calculation is a decisive element in any scientific research, representing a critical decision that determines the capacity to obtain accurate and meaningful results, with a direct impact on the validity and feasibility of the project [1,2].

Its fundamental importance lies in controlling the inherent errors of research. An appropriate calculation keeps these errors at acceptable levels, providing the statistical power necessary to detect significant differences when they truly exist, an aspect crucial to the validity of the results [3]. Contrary to common intuition, a larger population size does not necessarily require a proportionally larger sample, since the sample size is determined primarily by the magnitude of the proportions to be detected and the desired level of precision [4].

In this review, we present the fundamental concepts of sample size calculation and introduce a calculator developed by our team that incorporates functions often overlooked by other tools, including a logistic component that facilitates the evaluation of time feasibility—thus ensuring both scientific rigor and practical viability in epidemiological studies.

Fundamental concepts for sample size calculation

Sample size estimation in prevalence studies requires an understanding of several key statistical concepts, described below.

Expected proportion or prevalence

This represents the anticipated frequency of the event of interest in the study population, expressed as a value between 0 and 1. For example, an expected prevalence of 0.20 indicates that we expect to find the event in 20% of the population under study. Here arises an apparent paradox: we must anticipate what we expect to find before conducting the study. Although this situation is challenging, it can be addressed through different methodological strategies [5].

There are three main approaches to resolving this apparent paradox. The first consists of consulting existing literature, using data from previous similar studies as a reference. The second involves asking experts in the field to estimate the proportion that might reasonably be expected in the study. The third approach, particularly valuable, is conducting a pilot study [6].

Ultimately, the pilot study emerges as an invaluable methodological tool that goes beyond merely estimating sample size. Through the pilot, we can assess the effectiveness of the recruitment process, verify the quality of the proposed measurements, and anticipate potential challenges in participant follow-up, including the estimation of possible losses during the study. Although there is no clear consensus, existing literature suggests that a minimum pilot sample of 30 observations is recommended.

However, in situations where no reliable prior estimate exists, it is common to use a prevalence of 0.5, which yields the largest possible sample size, thus ensuring sufficient representation regardless of the true prevalence.

Confidence level

The confidence level is another essential concept that represents the probability that the calculated confidence interval contains the true population value. It is typically set at 0.95 (95%) or 0.99 (99%). This value reflects our statistical certainty: with a 95% confidence level, we can assert that if the study were repeated multiple times, in 95% of the cases, the calculated interval would contain the true population parameter. As with precision, a higher confidence level requires a larger sample size [7].

Precision

Precision and confidence level are fundamental concepts that determine the quality of estimates in prevalence studies. While the confidence level (typically 95%) indicates the probability that the population parameter lies within the calculated interval, precision determines the width of that interval. For instance, with a prevalence of 20%, a precision of 5% yields an interval of 15-25%, whereas a 2% precision narrows it to 18-22%. Greater precision provides more accurate estimates for decision-making and allows the detection of small yet relevant differences, though it requires considerably larger samples [1].

The selection of precision must balance scientific needs and available resources, considering that both parameters multiply their effect on sample size. In exploratory studies, a precision of 5-7% may be acceptable, whereas definitive studies aimed at informing health policies may require 2-3%. This decision must also take into account the magnitude of the parameter: a 5% precision might be excessive for a 1% prevalence but insufficient for a 50% one. Balancing precision, reliability, and feasibility is thus

one of the most crucial aspects in the design of epidemiological research.

Finite or infinite population

The distinction between finite and infinite populations in sample size calculation goes beyond the mere population size. This methodological decision depends on multiple factors during study planning. Traditionally, the infinite population approach is recommended when the sample represents less than 5% of the total population. Nevertheless, this criterion should be evaluated in light of the inferential goals: if the aim is to generalize beyond the immediate population, the infinite population approach is more appropriate, even for relatively small populations [8].

The finite population correction factor (FPC) plays a key role in this decision, adjusting the sample size when dealing with finite populations. The FPC formula (√[(N-n)/(N-1)]) reflects how the sampled proportion affects the precision of estimates. When the sample represents a significant portion of the total population, the FPC reduces the required sample size, recognizing that each sampled individual provides more information about the entire population [9].

When working with very large populations or when broader inferences are desired, the infinite population approach allows for more conservative and potentially generalizable estimates. It is preferable when there is uncertainty about the exact population size or when it may vary during the study. The choice between both approaches has practical implications for precision and efficiency: using the finite population approach when appropriate can optimize resources, whereas the infinite population approach favors generalization—thus requiring a balance depending on the study's objectives.

Sensitivity and specificity of the test

Sensitivity and specificity are fundamental parameters that allow for the adjustment of sample size calculations according to the accuracy of the measurement method used to identify the event of interest. These parameters are particularly relevant when the gold standard is not employed. Sensitivity represents the ability of the method to correctly identify true positive cases, while specificity reflects its ability to correctly identify true negative cases. Both parameters are expressed as proportions ranging from 0 to 1 [10].

For example, if a validated questionnaire is used to detect depression with a sensitivity of 0.90 and a specificity of 0.95 (since it is not a gold-standard test), this means that the instrument will correctly detect 90% of true depression cases and correctly identify 95% of non-depression cases. These values affect the estimation of true prevalence and, consequently, influence the required sample size to achieve accurate estimates [10].

Application of strata

Stratified sampling enhances representativeness and efficiency in prevalence studies by dividing the population into mutually exclusive subgroups. This strategy is particularly valuable when there are distinct subgroups with possible variations in the prevalence of the event under study, such as in university settings where academic years constitute natural strata [11].

The design effect is a crucial parameter that adjusts the sample size to account for the complexity introduced by stratification. A value greater than 1 indicates the need for a larger sample than that required in simple random sampling to maintain the same level of precision, due to the additional variability introduced [12].

This methodology allows the generation of specific estimates by subgroup, facilitates the identification of differences between strata, and optimizes efficiency when the strata are internally homogeneous but heterogeneous between groups. It can be implemented proportionally (respecting population proportions) or disproportionately (oversampling specific strata), depending on the objectives and practical considerations of each study.

Sample size versus precision calculation

In prevalence studies, there are two fundamental approaches: sample size calculation and precision calculation, each addressing different research needs [13,14].

Sample size calculation—the traditional and prospective approach—is used during the planning phase. The researcher starts with predefined parameters (expected prevalence, desired precision, and confidence level) to determine how many subjects need to be studied. This approach is useful when recruitment flexibility and sufficient resources are available.

Conversely, precision calculation represents a retrospective approach, valuable when working with a predetermined sample size due to budgetary, time,

or accessibility constraints, or when using existing databases.

Logistical considerations after sample size calculation

Once the theoretical sample size required for a prevalence study has been determined, it is essential to consider logistical factors that will influence its practical implementation. The calculated sample size represents the final number of participants needed for analysis; however, fieldwork realities involve various factors that affect the total number of subjects that must initially be contacted [15,16].

The rejection rate and eligibility rate are two critical factors in logistical planning. The former represents the proportion of individuals who, when contacted, choose not to participate; the latter indicates the percentage of contacted individuals who meet the inclusion and exclusion criteria. For instance, if 300 final participants are needed, with an expected rejection rate of 20% and an eligibility rate of 80%, approximately 470 individuals must initially be contacted. These estimates should be based on prior experience or pilot studies.

The operational capacity of the research team determines how many subjects can be processed per day. This factor depends on available personnel, time required for instrument administration, availability of equipment or facilities, and scheduling constraints for both the research team and the target population. The total time required to complete recruitment is calculated considering all these elements; for example, if 470 people must be contacted and 220 participants can be processed per month, the recruitment period will extend over approximately two and a half months.

Use and implementation of the calculator

The *CalculadoraPrevalencia* has been designed as a versatile tool for sample size and precision calculation in prevalence studies, implemented within the R programming environment. This tool integrates multiple methodological aspects that traditionally had to be considered separately, allowing for a more comprehensive analysis tailored to the specific needs of each study. Researchers wishing to use the calculator need only install the package in R using the command: devtools::install_github("VicVePo/CalculadoraPrevalencia").

Once installed, the package can be loaded into an R session with the command: library ("Calculadora Prevalencia"). This simple installation and loading process allows researchers to immediately access all the functionalities for sample size calculation and logistical planning included in the package, thus facilitating the design and organization of epidemiological studies.

The structure of the *Calculadora Prevalencia* is based on a main function that accepts various parameters depending on the specific characteristics of the study: expected prevalence, desired precision, and confidence level for sample size calculation, with the flexibility to include the population size when known (N) or to specify *NA* for infinite populations. Additionally, it allows adjustments for measurement quality through sensitivity and specificity parameters, which can be modified according to the method used. For studies employing stratified sampling, the calculator also enables the specification of both the proportions of each stratum (*prop_estratos*) and the design effect (*efecto_diseno*), thus offering a comprehensive tool for diverse epidemiological research scenarios (see Figure 1).

```
resultado <- CalculadoraPrevalencia(
  prevalencia = NA,                # proporción esperada
  precision = NA,                  # Precisión
  nivel_confianza = NA,            # nivel de confianza
  N = NA,                          # población infinita o finita
  sensibilidad = NA,               # sensibilidad
  especificidad = NA,              # especificidad
  prop_estratos = proporciones,    # proporciones calculadas
  efecto_diseno = NA,              # efecto de diseño
  calcular = "NA")                 # cálculo de "muestra" o "precision"

print(resultado)
```

Figure 1. Matrix of the Sample Size/Precision Calculator in R

```
resultado_logistico <- logistica_estudio_transversal(
  n_final = muestra$tamano_muestral,
  tasa_rechazo = NA,              # % de rechazo
  tasa_elegibilidad = NA,         # % de elegibilidad
  sujetos_por_dia = NA,           # Número observaciones por día
  dias_laborables_mes = NA)       # Número de días laborables por mes

print(resultado_logistico)
```

Figure 2. Matrix of the Logistic Section after Sample Size Calculation in R

The versatility of the *CalculadoraPrevalencia* is demonstrated by its ability to operate in two distinct modes through the "calcular" parameter. The "sample" mode determines the required sample size given a desired level of precision, while the "precision" mode calculates the expected precision for a predetermined sample size. This dual functionality enables its application both in prospective planning and in the evaluation of study designs with sampling constraints. Additionally, the tool is complemented by a function for logistic analysis (*logistica_estudio_transversal*), which translates the theoretical sample size into practical requirements. It takes into account rejection rates, eligibility criteria, and operational capacity to provide realistic estimates of the time and resources needed (see Figure 2).

The results generated by the calculator are presented in a structured format that includes not only the calculated sample size or precision but also all the parameters used in the computation. This facilitates the documentation of the planning process and allows for transparent communication of methodological decisions in scientific publications.

Application of Examples

Example 1: Prevalence study of depressive symptoms in the city of Chachapoyas

To estimate the prevalence of depression among adults in Chachapoyas (population: 80,000 inhabitants), the following parameters were considered: expected prevalence of 20%, precision of 5%, 95% confidence level, and the use of the PHQ-9 instrument (sensitivity 88% and specificity 92%). This can be visualized in Figure 3.

The calculation indicates a required sample of 438 individuals. Considering a 25% rejection rate, 80% eligibility, and a capacity to assess 8 participants per day, approximately 730 individuals need to be contacted, requiring an estimated 92 days of fieldwork. This calculation can be found in the Appendix section (*see Appendix 1*).

Example 2: Prevalence study of arterial hypertension in the district of Luya

The district of Luya (4,200 inhabitants) illustrates the sample size calculation for a finite population. For this study, the following parameters were considered: expected prevalence of 25%, precision

```
> muestra <- CalculadoraPrevalencia(
+   prevalencia = 0.20,           # prevalencia esperada
+   precision = 0.05,             # Nivel de precisión
+   nivel_confianza = 0.95,       # nivel de confianza
+   N = NA,                       # población infinita o finita
+   sensibilidad = 0.88,          # sensibilidad de la prueba
+   especificidad = 0.92,         # especificidad de la prueba
+   calcular = "muestra")
>
> print(muestra)
$tamano_muestral
[1] 438
```

Figure 3. Initial Calculation Using the *CalculadoraPrevalencia* in R

```
> muestra <- CalculadoraPrevalencia(
+   prevalencia = 0.25,           # prevalencia esperada
+   precision = 0.05,             # Nivel de precisión
+   nivel_confianza = 0.95,       # nivel de confianza
+   N = 4200,                     # población infinita o finita
+   sensibilidad = 0.95,          # sensibilidad de la prueba
+   especificidad = 0.92,         # especificidad de la prueba
+   calcular = "muestra")
>
> print(muestra)
$tamano_muestral
[1] 386
```

Figure 4. Calculation Results for Finite Population in Luya

of 5%, 95% confidence level, and the use of digital sphygmomanometers (sensitivity 95%, specificity 92%). The sample size calculation is shown in Figure 4.

The calculation indicates a sample of 245 individuals, smaller than the approximately 425 that would result from assuming an infinite population, due to the finite population correction factor. For logistical planning, considering a 20% rejection rate, 70% eligibility, and a capacity to assess 6 participants per day, it will be necessary to contact 690 individuals over approximately four months. This calculation can be found in Appendix 2.

The logistical analysis shows that 690 individuals must be contacted initially to achieve the calculated sample size. Implementing the study in Luya will require around four months of fieldwork, taking into account the geographical and cultural characteristics of the area. It is essential to establish a sampling strategy that considers the spatial distribution of the population across the district — including both the main town and surrounding villages — even if non-probabilistic sampling is used. Coordination with local authorities and community health agents will be crucial to facilitate access to the population and maximize participation.

Example 3: Prevalence study of anxiety among students at Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas using stratified sampling

In this university study, seven faculties were considered as natural strata, each with its respective proportion. The student population is distributed as follows: Faculty of Health Sciences (18.40%), Faculty of Civil Engineering (17.92%), Faculty of Social Sciences (14.62%), Faculty of Systems Engineering (16.51%), Faculty of Administration (11.32%), Faculty of Education (9.43%), and Faculty of Agroindustrial Engineering (11.80%). An expected anxiety prevalence of 30% was considered, using the GAD-7 instrument (sensitivity 89%, specificity 82%), with a 5% precision and 95% confidence level (see Figure 5).

The calculation indicates that a total sample of 402 students is needed. The stratification by faculty yields the following distribution: 1) Health Sciences: 134 students; 2) Civil Engineering: 131 students; 3) Social Sciences: 107 students; 4) Systems Engineering: 121 students; 5) Administration: 83 students; 6) Education: 69 students; and 7) Agroindustrial Engineering: 86 students.

```
> # Creo las proporciones según la información obtenida
> proporciones <- c(0.1840, 0.1792, 0.1462, 0.1651, 0.1132, 0.0943, 0.1180)
>
> # Cálculo del tamaño muestral estratificado
> resultado_estratificado <- CalculadoraPrevalencia(
+   prevalencia = 0.30,              # prevalencia esperada del 20%
+   precision = 0.05,                # precisión del 5%
+   nivel_confianza = 0.95,          # nivel de confianza del 95%
+   N = NA,                          # tamaño de población finita
+   sensibilidad = 0.89,             # sensibilidad del 90%
+   especificidad = 0.82,            # especificidad del 95%
+   prop_estratos = proporciones,    # proporciones calculadas
+   efecto_diseno = 1,               # efecto de diseño
+   calcular = "muestra")            # calcular muestra
>
> print(resultado_estratificado)
$tamano_muestral
[1] 731

$tamanos_estratos
[1] 134 131 107 121  83  69  86
```

Figure 5. Stratified Sample Size Calculation Using *CalculadoraPrevalencia* in R

For the logistical analysis, considering a 40% rejection rate, 80% eligibility, and a capacity to assess 15 students per day, it will be necessary to contact 1,523 students for fieldwork (*see Appendix 3*).

Conclusions

Rigorous methodological planning is essential for the success of epidemiological studies. The *CalculadoraPrevalencia* emerges as a comprehensive tool that facilitates not only the calculation of sample size and precision but also the consideration of crucial practical aspects. Its versatility is evident in its ability to handle various scenarios — finite populations, stratified samples, and secondary data analyses.

The integration of the logistical component represents a significant advancement in study design. The automatic calculation of the total number of subjects to be contacted (considering rejection and eligibility rates) and the estimation of the time required for fieldwork allow for more realistic planning. The inclusion of parameters such as sensitivity and specificity of instruments enables more reliable estimates tailored to each context.

Recommendations for the effective use of the calculator

During planning, it is essential to carefully evaluate the initial parameters. The expected prevalence should be based on prior literature or pilot studies. The desired precision must balance scientific needs with available resources. Stratification should be considered when subgroups with potential differences in event prevalence exist, even if this increases logistical complexity.

For studies involving diagnostic tests, it is crucial to incorporate realistic data on sensitivity and specificity, conducting sensitivity analyses when these parameters are uncertain. Logistical planning should be conservative in challenging settings, slightly overestimating required resources rather than facing significant delays.

In secondary analyses, the evaluation of precision should be performed before conducting substantive analyses. Finally, it is important to remember that the calculator is a support tool, not a substitute for the researcher's critical judgment. Results must always be interpreted within the specific context of the study, considering practical, ethical, and scientific aspects that may not be fully reflected in numerical calculations.

REFERENCES

1. Mascha EJ, Vetter TR. Significance, Errors, Power, and Sample Size: The Blocking and Tackling of Statistics. Anesth Analg. [Internet]. 2018 [cited 2025 Mar 12];126(2):691-8. doi:10.1213/ANE.0000000000002741
2. Andrade C. Sample Size and its Importance in Research. Indian J Psychol Med. [Internet]. 2020 [cited 2025 Mar 12];42(1):102-3. doi:10.4103/IJPSYM.IJPSYM_504_19
3. Chow S-C, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research [Internet]. 3rd edition. New York: Chapman and Hall/CRC; 2017 [cited 2025 Mar 12]. doi:10.1201/9781315183084
4. Rao UK. Concepts in sample size determination. Indian J Dent Res [Internet]. 2012 [cited 2025 Mar 12];23(5):660-4. doi:10.4103/0970-9290.107385
5. Dattalo P. Determining Sample Size [Internet]. Oxford University Press; 2008 [cited 2024 Nov 3]. doi:10.1093/acprof:oso/9780195315493.001.0001
6. Wang X, Ji X. Sample Size Estimation in Clinical Research: From Randomized Controlled Trials to Observational Studies. Chest. [Internet]. 2020 [cited 2024 Nov 3];158(1S):S12-20. doi:10.1016/j.chest.2020.03.010
7. Case LD, Ambrosius WT. Power and sample size. Methods Mol Biol. [Internet]. 2007 [cited 2024 Nov 3];404:377-408. doi:10.1007/978-1-59745-530-5_19
8. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med. [Internet]. 2013 [cited 2024 Nov 3];35(2):121-6. doi:10.4103/0253-7176.116232
9. Population Correction Factor - an overview | ScienceDirect Topics [Internet]. [cited 2024 Nov 3]. Available from: https://www.sciencedirect.com/topics/mathematics/population-correction-factor
10. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. [Internet]. 2014 [cited 2024 Nov 3];48:193-204. doi:10.1016/j.jbi.2014.02.013
11. Lohr SL. Sampling: design and analysis. 2nd ed. Boston, MA: Brooks/Cole; 2010.
12. Kelley K. Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. Behav Res Methods [Internet]. 2007 [cited 2024 Nov 3];39(4):755-66. doi:10.3758/BF03192966
13. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J. [Internet]. 2003 [cited 2024 Nov 3];20(5):453-8. doi:10.1136/emj.20.5.453
14. Das S, Mitra K, Mandal M. Sample size calculation: Basic principles. Indian J Anaesth. [Internet]. 2016 [cited 2024 Nov 3];60(9):652-6. doi:10.4103/0019-5049.190621
15. Anastasi JK, Capili B, Norton M, McMahon DJ, Marder K. Recruitment and retention of clinical trial participants: understanding motivations of patients with chronic pain and other populations. Frontiers in Pain Research [Internet]. 2024 [cited 2024 Nov 3];4:1330937. doi:10.3389/fpain.2023.1330937
16. Smith PG, Morrow RH, Ross DA. Field Trials of Health Interventions [Internet]. 3rd ed. Oxford: OUP Oxford; 2015 [cited 2024 Nov 3]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK305515/

Funding sources

This research was conducted using the author's own resources.

Conflict of interest statement

The authors declare no conflicts of interest.

Contribución del autor

Conceptualización, metodología, análisis formal, investigación, recursos, redacción – borrador original, redacción - revisión y edición y visualización.

APPENDICES

Appendix 1. Logistic analysis of Example 1

```
> resultado_logistico <- logistica_estudio_transversal(
+   n_final = muestra$tamano_muestral,
+   tasa_rechazo = 0.25,          # 25% de rechazo
+   tasa_elegibilidad = 0.8,      # 80% de elegibilidad
+   sujetos_por_dia = 8,          # 10 sujetos por día
+   dias_laborables_mes = 22)     # 22 días laborables por mes
>
> print(resultado_logistico)
$muestra_final
[1] 438

$muestra_evaluar
[1] 584

$muestra_invitar
[1] 730

$dias_reclutamiento
[1] 92
```

Appendix 2. Logistic analysis of Example 2

```
> resultado_logistico <- logistica_estudio_transversal(
+   n_final = muestra$tamano_muestral,
+   tasa_rechazo = 0.20,          # 25% de rechazo
+   tasa_elegibilidad = 0.7,      # 80% de elegibilidad
+   sujetos_por_dia = 8,          # 10 sujetos por día
+   dias_laborables_mes = 20)     # 22 días laborables por mes
>
> print(resultado_logistico)
$muestra_final
[1] 386

$muestra_evaluar
[1] 483

$muestra_invitar
[1] 690

$dias_reclutamiento
[1] 87
```

Appendix 3. Logistic analysis of Example 3

```
> resultado_logistico <- logistica_estudio_transversal(
+   n_final = resultado_estratificado$tamano_muestral,
+   tasa_rechazo = 0.40,          # 25% de rechazo
+   tasa_elegibilidad = 0.8,      # 80% de elegibilidad
+   sujetos_por_dia = 15,         # 10 sujetos por día
+   dias_laborables_mes = 20)     # 22 días laborables por mes
>
> print(resultado_logistico)
$muestra_final
[1] 731

$muestra_evaluar
[1] 1219

$muestra_invitar
[1] 1523

$dias_reclutamiento
[1] 102
```