Review Article

Non-response problem in health surveys

Shishir Kumar [1*], Chhaya Gangwal [2], Avni Punia [3]

[1] Department of Community Medicine, IGIMS, Patna, India.
[2] Department of Mathematics, Patna University, Patna, India.
[3] Avni Punia**,** All India Institute of Medical Sciences, Madurai, India.

Corresponding author and reprints: Dr. Shishir Kumar, Additional Professor (Biostatistics), Department of Community Medicine, Indira Gandhi Institute of Medical Sciences, Patna, India.
Email: shishirpunia@gmail.com
Received: 05 May 2024 **Accepted:** 10 Jun 2024 **Published:** 11 Jul 2024

Abstract

The impact of non-participation in surveys depends on how many people don't respond and the extent to which those who don't differ from those who do on key questions. In health research, non-participation can skew results, affecting our understanding of disease and healthcare needs. Understanding why people don't respond to surveys is crucial for accurately gauging health issues. Despite the popularity of surveys in health research, the problem of non-response bias persists due to low response rates. This paper explores various methods, their strengths, and weaknesses, to address non-response issues in health surveys.

Keywords: Bias; Health Surveys; Research Design.

Cite this article as: Kumar S, Gangwal C, Punia A. Non-response problem in health surveys. *Soc Determinants Health*. 2024;10(1):1-5. DOI: http://dx.doi.org/10.22037/sdh.v10i1.45236

Introduction

Non-response poses a growing concern in survey research. It happens when people either can't or won't answer questions asked by interviewers, occurring in both sample surveys and censuses. The impact of non-response varies widely across different types of surveys, affecting survey quality in two main ways. Firstly, it reduces the amount of available data, making population estimates less precise. Secondly, if there's a link between the variable being studied and response behavior, conclusions drawn from responses may not be valid for the entire population. Keeping non-response to a minimum is crucial. However, if substantial non-response persists despite efforts, corrective measures are necessary to prevent inaccurate population assessments, often involving a mix of adjustment techniques and standard estimation methods.

2. The Events of Non-response.

2.1. Technical application

The objective of any survey is to determine specific population characteristics. However, due to various errors, the true value is often elusive. Non-response stands out as a significant factor contributing to inaccurate population estimates. It refers to the failure to gather measurements or observations from certain elements chosen for inclusion in a sample. Non-response affects nearly all surveys and, if widespread, can seriously undermine the validity and reliability of survey results[1,2]. While the classification of non-response errors depends on survey contexts, the following categories shed light on its classification:

(1) Not-at home: This group includes respondents who are not home when the enumerator visits. This is common in surveys where respondents are unaware of

the enumeration or are temporarily away. Recalls can help reduce this category, labeling it as "temporarily unavailable" to indicate a delay rather than denial of the interview.

(2) Refusal: These respondents decline to provide information for various reasons or don't respond to enumerators/questionnaires, or are away from home during the survey period. Refusals can be temporary or permanent, with repeated attempts unlikely to succeed. "Unobtainable" is a better term for this category, denoting a denial rather than a delay of observation.

(3) Incapacity or Inability: This non-response type may be due to mental or physical illness or language barriers preventing response during the survey period. While it could fit under "unobtainable," distinguishing between unwilling and willing but incapable respondents may be useful in certain situations.

(4) Not Found: This category includes respondents who are not identified or followed, often because of cost constraints or inaccessibility. Cases of not attempted interviews fall into this category, possibly due to dangerous surroundings or inaccessibility.

(5) Lost Information: Information may be lost after field attempts due to poor quality questionnaires, loss or destruction during transit, or incomplete forms due to cheating. While the described typology applies to most survey situations, caution is needed in complex sampling designs.

2.2. The Extent of Non-response:

The more non-response there is, the more concerned we should be about its impact on survey estimates. Bias tends to increase as the rate of non-response rises. While it's challenging to measure bias objectively, quantifying the extent of non-response is relatively straightforward. However, comparing non-response rates across different surveys is difficult due to various factors such as the survey's purpose, type of sampling, design, fieldwork efficiency, interviewer performance, measures to reduce non-response, survey timing, target population, questionnaire length, and question wording. To enable proper survey comparisons, it's essential to establish a framework that controls the factors influencing non-response rates, allowing for informed judgments on the extent of non-response.

2.3. Models for Non-response[1]

The first step in developing theories to address non-response is to create a mathematical model that explains how non-response occurs. Two commonly used models for treating non-response are the "random response model" and the "fixed response model."

In the random response model, each individual in the population has an unknown probability of responding, which may vary from person to person. When an interviewer contacts someone for questioning, a probability mechanism determines whether that person responds.

On the other hand, the fixed response model assumes two strata within the population: potential respondents and potential non-respondents. The size and composition of each stratum are unknown beforehand and are determined by survey specifications. After disregarding these strata, a sample is chosen from the population, making the number of respondents a random variable in both models.

If complete enumeration were to occur instead of sampling, in the random response model, the determination of respondents would still be a random process, while in the fixed response model, it would be fixed. Despite their differences, both models share some similarities. They both involve two stochastic mechanisms: the sampling mechanism and the response mechanism. The main distinction lies in the order in which these mechanisms are applied.

In the fixed response model, the response mechanism is activated first, determining the two strata, followed by sample selection. In contrast, the random response model selects the sample first, and then the response mechanism is activated for each selected individual.

The random response model allows for the estimation of response probabilities, which can be used in adjustment procedures or linked to personal characteristics. The fixed response model typically yields simpler formulas. However, the accuracy of estimates under this model depends on the realized response and non-response strata, making it difficult to determine the accuracy of the estimation method. As a result, more attention is often given to the random response model.

3. Selection of Auxiliary Variables[3]

Discovering a possible connection between the variable being studied and response behavior is crucial. However, using sample data alone, it's impossible to establish such a relationship because the values of the variable for non-respondents are unknown. To glean insights about non-respondents, we need information about them. Auxiliary variables serve as one such source of information. These variables are measurable for both respondents and non-respondents and can be categorized into two types:

3.1 Information collected by interviewers without face-to-face interviews, such as town type, housing type, approximate year of non communicable disease, and neighborhood social status in housing surveys.

3.2 Information obtained from administrative records.

Analyzing the relationship between auxiliary variables and response behavior helps understand the characteristics of non-respondents and may provide additional insights into the relationship between the variable under study and response behavior. Auxiliary variables that exhibit a clear connection with response behavior are vital for adjusting non-response. It's assumed that auxiliary variables are nominal, meaning their different values are only used to distinguish between groups, and arithmetic operations on these values are not allowed. This nominal assumption isn't restrictive in practice, as many variables are nominal, and other variable types can be easily expressed in terms of nominal variables.

4. Reducing Bias from Non-Response through Subgroup Weighting[2,3]

4.1 Identifying the Relationship: When we suspect or find a relationship between what we want to know (Y) and whether people respond (R), we need to tackle the bias caused by non-response.

4.2 Splitting the Population into Subgroups: If we can split the whole population into smaller groups where the differences in Y and R don't matter or not significant, we can find a better average Y for each group more accurately.

4.3 Finding Group Averages: In each smaller group, you find the average income ($\bar{y}n^*$) of the people who responded (the ones who told you their income).

4.4 Combining Subgroup Estimates: We can then combine these average estimates from each subgroup ($\bar{y}1^*$, $\bar{y}2^*$, etc.) to get a better estimate of the average Y for the entire population ($\bar{y}^*$).

4.5 Weights for Accuracy: The accuracy of your estimates depends on how well you know the sizes of these smaller groups. If you know the sizes, you can use them directly as weights (wh). If you don't know them, you can estimate them using the sizes of the sample you got from each group (n_h).

4.6 Reducing Bias and Variance: By using this method, you can reduce bias (errors in your estimates) caused by non-response and get a better picture of the whole population's average income. The

more you know about the sizes of the smaller groups, the more accurate your estimates will be.

Suppose, we have a variable X that helps us split the population into H subgroups, each with sizes $N1, N2, \ldots, NH$. When we use subgroup weighting:

4.7 Estimating Subgroup Mean: First, we compute an estimator $\bar{y}n^*$ for the mean of each subgroup 'h'. This is calculated by taking the average of the Y values (income, for example) of the responding elements in each subgroup.

Formula: (Equation 1)

$$\bar{y}_n^* = \frac{1}{m_h}\sum_{i=1}^{m_h} y_{hi}$$

Here, m_h is the number of respondents in subgroup 'h', and $y_{h1}^*, y_{h2}^*, \ldots\ldots, y_{hmh}^*$ are their corresponding income values.

4.8 Combining Subgroup Estimates: Next, we combine these subgroup estimators to get an overall population estimator, denoted as $\bar{y}^*$.

Formula: (Equation 2)

$$\bar{y}^* = \sum_{h=1}^{H} w_h . \bar{y}_h^*$$

Here, w_h represents the weight assigned to subgroup 'h'. The type of estimator depends on the available information about these weights.

4.9 Determining Weight Types: If we know the sizes $(N1, N2, \ldots, NH)$ of the subgroups, we can use fixed weights.

Formula: $w_h = \frac{N_h}{N}$ (Equation 3)

Here, N_h is the size of subgroup 'h' and N is the total population size.

4.10 Estimating Weights: If we don't know the subgroup sizes, we can estimate them using the sizes of the sample elements $(n1, n2, \ldots, nH)$ in each subgroup.

Formula: $w_h = \frac{n_h}{n}$ (Equation 4)

Here, n_h is the number of sample elements in subgroup 'h', and n is the total sample size.

These formulas help us estimate population parameters while accounting for differences in subgroup sizes and response rates.

5. Other Adjustment Methods

Several methods have been developed to address non-response, and we'll briefly discuss some of them here.

5.1. No Adjustment

In certain cases, no adjustment is needed. If there's no apparent relationship between the variable being studied and response behavior, the response can be viewed as a random sample from the population. Also, if statements are confined to the population of potential respondents, no correction is required. In all other cases, 'no adjustment' is only justified if the category "non-response" is included in all tables in publications.

5.2. Imputation

Imputation procedures address missing observations due to non-response by replacing values in the records of non-respondents. In "hot deck" imputation, data come from respondents of the current survey, while in "cold deck" imputation, data are from a previous survey. If the response structure of previous and current surveys is similar, results from cold deck and hot deck imputation will roughly align. Imputation can be done in several ways:

1) Imputation of a random respondent.
2) Imputation of the mean respondent.
3) Imputation of a random respondent within the same subgroup.
4) Imputation of the mean respondent within the same subgroup.
5) Imputation of a value obtained by fitting a model.

Procedures (1) and (2) don't reduce bias. Procedures (3) and (4) resemble subgroup weighting. The impact of procedure (5)

relies heavily on the model's fit and the reasonableness of model assumptions.

5.3. Adjustment for Not-at-Homes

The Politz and Simmons (1949)[4] method attempts to correct for not-at-home bias by estimating the probability of finding a person at home. Respondents are asked, for example, how often they were home during the previous days. The constructed at-home probability serves as a stratification variable. It's also worthwhile to seek a model explaining the relationship between the variable being studied and the at-home probability. Extending this model to the group of not-at-homes may yield more insights about this group.

5.4. Adjustment for Refusers

Measuring people's willingness to cooperate in the survey allows for adjustment similar to not-at-home adjustment. Additionally, willingness to cooperate serves as a gauge for the survey climate. Constructing a scale for this information might be somewhat more challenging than for not-at-home adjustment.

5.5. Double Sampling

Harsen & Hurvitz (1946)[5] proposed selecting a sample from respondents to gather more information about non-respondents. Specially trained interviewers take a subsample and make more intensive and accurate measurements on selected units to obtain missing information. However, time and budget constraints often hinder the use of double sampling.

5.6. The Principal Question

If the Hansen & Hurwitz (1946)[5] method is too costly, the principal question procedure could serve as an alternative. Many surveys revolve around one key question. If completing the entire questionnaire poses challenges during fieldwork, interviewers may focus on obtaining answers to the principal question. This effort may also extend to follow-up by letter or telephone.

6. Conclusion

Given the increasing rates of non-response in recent years, it's crucial to conduct thorough research on how non-response affects survey quality and to develop new methods to reduce non-response bias. The significant variations in survey objectives, designs, and implementations make it difficult to accurately interpret differences in non-response rates. Hence, there's a need to establish a theoretical framework that enables proper comparisons. If your study experiences high levels of non-response, consider how the characteristics of non-respondents might differ from those who participated in your survey.

References

1. Julian Baim, Martin Frankel, Joe Agresti, Jim Collins, Risa Becker and Van Podberesky. “Alternative Statistical Techniques to Address Non-Response Bias.”. Available from: https://www.pdrf.net/wp-content/uploads /2015/10/21_Syn13 BaimAlternative-StatisticalSess3 _ final.pdf
2. Daniel K. “Measurement error in household surveys: sources and measurement”. Available from: https:// unstats. un.org /unsd / hhsurveys /pdf/Household_ surveys.pdf
3. C.E. Sarndal and S. Lundstrom. “*Estimation in Surveys with Nonresponse”.* England: John Wiley & Sons Ltd, 2005.
4. Politz, Alfred and Willard Simmons. “An Attempt to Get the Not-At -Homes into the Sample Without Callbacks.” Journal of the American Statistical Association (1949) 44:9-32.
5. Hansen, M. H. and Hurwitz, W. N., "The Problem of Non-response in Sample Surveys," Journal of the American Statistical Association, Vol. 41, (1946), pp. 517-529.