CME

Understanding of Scales of Measurements and its Consideration in Clinical Research

Sada N. Dwivedi

In the present digital world, regardless of scale of measurements, often collected data are likely to be inappropriately considered as numeric. It might be mainly because even data on qualitative cum categorical variables are collected using certain numeric codes which look like quantitative instead of qualitative variables. Such an inappropriate understanding about scales of measurements often leads to consideration of inappropriate analytical methods providing inaccurate results. Hence, such practice results in inaccurate interpretations and sometimes disastrous implications. Therefore, understanding measurement scales goes a long way to appropriately plan & execute study; collect accurate data, analyze appropriately, interpret results accurately and deriving policy-driven implications.

Access this article online

Website:
www.cijmr.com

Keywords:
Coronary heart disease, Scale of measurements, Quantitative vs. Qualitative variables

10.58999/cijmr.v2i01.102

Introduction

The understanding of scale of measurement of every variable in a study, including exposure and outcome(s), goes a long way to help in planning a study, collecting data, managing data, analyzing data and interpreting the analytical results accurately and appropriately.[1-7] The variables are broadly categorized as qualitative and quantitative. The qualitative variables are non-numeric, whereas the quantitative variables are numeric. To elaborate further, the qualitative variables are further grouped as those on the nominal scale or ordinal scale. Each of them is non-numeric. Numbers serve merely as tags or labels to identify the individuals/ items. Accordingly, a nominal scale variable consists of two or more categories helping in counting only (e.g., gender as male and female). On the contrary, an ordinal scale variable consisting of two or more categories also ranks those categories (e.g., degree of symptomatic morbidity as normal, mild, moderate and severe; Likert type item). In other terms, in case of an ordinal scale, numbers also describe the magnitude of a variable. However, in both scales of qualitative variables, numbers are non-numeric and the difference between numbers does not carry any meaning.

To elaborate quantitative variables further, they are also grouped as those on interval or ratio scales. The interval scale helps in assigning numeric values in any range to arbitrary assessments such as feelings about pain and Likert scale (i.e., sum of scores on various Likert-type items). In contrary to qualitative variables, the interval scale can quantify the difference between the values. In this case, mean and median may be calculated. Also, the difference between two such variables is meaningful. In other words, the variables on the interval scale are measured in an exact manner. They are not measured in a relative way where the presence of zero is arbitrary. On the other hand, the ratio scale is a measured variable (e.g., height; weight) where zero is absolute. Therefore, due to zero features, it does not have negative numbers. Hence, it allows for comparing differences or intervals. As such, for variables on ratio scale, all measures of location, including mean, median and mode, and all measures of dispersion, including standard deviation and coefficient

International Centre for Health Research, R D Gardi Medical College, Ujjain, Madhya Pradesh, India

Correspondence to: Sada N. Dwivedi, International Centre for Health Research, R D Gardi Medical College, Ujjain, Madhya Pradesh, India. E-mail: dwivedi7@hotmail.com

Submitted: 06/03/2023 Revision: 15/04/2023
Accepted: 04/05/2023 Published: 20/05/2023

How to cite this article: Dwivedi SN. Understanding of Scales of Measurements and its Consideration in Clinical Research. Central India Journal of Medical Research. 2023;2(1):13-16.

of variation, may be calculated. Also, they allow unit conversions like kilogram and gram.

Quantitative vs. Qualitative Variables

Once data are recorded using structured / semi-structured questionnaire, while analysing, sometimes every variable is considered as quantitative regardless of quantitative or qualitative forms just because every variable is recorded using numbers. To utter shock, while reviewing articles submitted to reputed journals like Indian Journal of Medical Research, sometimes one can come across inappropriately reported mean and standard deviation of qualitative variables like gender (1: Male; 2 : Female) and occupation (1 to 8 categories). It might happen knowingly or unknowingly just because of misunderstanding that every variable is recorded in numerical forms, including qualitative variables being recorded using code numbers. As obvious, frequency and percentage need to be reported for categorical variables like this. Analysis needs to deal with due consideration regarding considered scale of measurement of each variable under the collected data including exposure variable (e.g., smoking, drinking) and outcome (e.g., hypertension, coronary heart disease), while assessing association between smoking and coronary heart disease (CHD). One needs to be aware about appropriate descriptive analytical approach in view of specific scale of measurement of each considered variable.

Change in Scale of Measurement

Often a change in scale of measurement is considered. For example, a quantitative variable is often considered as categorical variable. For example, as a clinical outcome, a systolic blood pressure above 130 mmHg may be considered as hypertensive and otherwise non-hypertensive. After this change in scale, we miss the information, like under this example, one cannot know about individual's systolic blood pressure. This is also true in case of changes in scale of measurements of exposure variables like smoking, breastfeeding, physical activity and so on. Also, it remains true in case of changes in confounders like age recorded in completed years. Therefore, it is always advisable to collect data in raw forms so that there is opportunity to explore and decide optimal scales of measurements.[4]

If outcome variable is quantitative (e.g., systolic blood pressure), to derive its associated factors, one needs to use multiple linear regression analysis using a set of independent variables (quantitative and/ or categorical variables) including exposure variable. Once it is converted as a categorical variable (hypertensive/ non-hypertensive), appropriately a logistic regression analysis needs to be carried out using same set of independent variables (quantitative and/ or categorical variables). In other words, even analytical method may change as a result of change in scale of measurement of an outcome variable. Further, if outcome remains same, a change in scale of measurement of even one of the considered independent variables may result into different regression model[4]. Therefore, one should not adopt casual approach while finalising scales of measurements of available variables under a study. The analytical results solely rely on structure of data set like how many variables are being considered, and what are their scales of measurements.

Ideally, after change in scales of measurements regardless of outcome/ exposure/ other independent variables, originally existing relationship should not get distorted. As such, if required, changes in scales need to be guided by existing relationship. In other words, changes in scales need to be guided more objectively along with common sense. In case of nominal categories, merger of only those categories might be acceptable who cannot be retained separately due to a small number of people, and theoretically in order. For example, while merging some of the occupation groups, skilled and non-skilled sub-categories need not to be merged together. Likewise, in case of ordinal scale, mildly diseased people need not to be merged with healthy people. Further, Likert type items need not to be presumed to be quantitative until unless they are on the scale of 7 and more points. However, summation of scores related to multiple Likert type items provides likert scale data for every individual which can be analysed quantitatively[6]. While dealing with any data set, these basic issues go in a long way to ensure accuracy in data analysis, its results, interpretations and implications.

Decision Regarding Scale of Measurements of Available Variables

To begin with, one needs to examine theoretically the possibility of retaining some of the variables in their original scales of measurements regardless of nominal, ordinal, interval and ratio scales. For remaining variables, after categorising them under each of these four scales of measurements, one has to understand available data and explore, assess and decide the changed scales of measurements more objectively. Strictly speaking, not a single variable in a data set needs to be considered in the form of questionable scale of measurement. Further, a particular variable need not to have same scale of measurement regardless of areas of research. For an instance, in case of three studies, that is, on children, adults and geriatrics groups, there will be three different scales of measurements of age in case it is being changed from completed years to categorical forms.

Prompting Points

There are various points due to which an objective finalisation of scales of measurements of each considered variable including exposure variable and outcome variable alone may not guarantee truly existing relationship between exposure variable and outcome variable. Some of them are briefly described below:

Clinical Relevance

As a first point, out of a series of available variables, only those variables need to be considered for further analyses which are relevant from public health program point of view, regardless of statistical significance. No need to consider nuisance variables just because they are available. As a rule of thumb, all relevant variables found statistically significant at 25% level of significance need to be included in further analyses[7].

Confounders

As a second point, missing potential variables (i.e., confounders) may distort the relationship between exposure and outcome variable. To be more specific, every other variable which can influence the relationship between exposure and outcome is known as confounder. For example, along with known confounders like age and gender, physical activity is a well known confounder for prevailing relationship between smoking and coronary heart disease (CHD). Hence, if no data is collected on drinking, without its adjustment true relationship between smoking and CHD may not be obtained. In other words, data on each of the potential confounders should be collected under any particular study.

Co-Linearity

As a third point, consideration of two highly correlated variables in same model might not only distort the results for the targeted relationship but also their relationships with outcome. Hence, after finalisation of scales of measurements of all available variables, such co linearity problems need to be ruled out through exploratory analysis. After identifying such pairs, at a time only one of the two variables need to be considered in a model building. Because of this, in a study, sometimes multiple epidemiological models might be developed using varying sub-sets of variables identified as discussed here. If required, one can further assess and identify optimal model through available statistical methods.

Effect Modifiers

As a fourth point, like confounders, missing potential effect modifiers even among non-collinear confounders may also distort the relationship between exposure and outcome variable. To be more specific, if extent of relationship between exposure and outcome varies at different levels of another variable, this variable is known as effect modifier for relationship between exposure and outcome. For example, drinking is a well known effect modifier for prevailing relationship between smoking and CHD. Hence, even if data collection on drinking is done, without adjustment regarding its effect modification true relationship between smoking and CHD may not be obtained. In other words, not only data on each of the potential effect modifiers should be collected under any particular study but also adjustment regarding its effect modification needs to be carried out. To adjust regarding effect modification of an effect modifier, it needs to be multiplied with exposure variable. For example, if smoking is exposure variable (yes/ no) where as drinking is effect modifier (yes/ no), as a result of their multiplication, new variable (smoker as well as drinker; smoker but not drinker; non-smoker but drinker; neither smoker nor drinker) will emerge. This resulting new variable needs to be also considered along with identified non-collinear confounders and exposure variable in a model building.

Confounder vs. Effect Modifier

As such, a variable may be only confounder; only effect modifier; confounder as well as effect modifier both; and neither confounder nor effect modifier. In case of only confounder, it has to be simply included in subset of variables to develop a model. But, in case of only effect modifier, only new variable obtained through multiplication of this effect modifier with exposure variable needs to be included in subset of variables. However, in case of confounder and effect moodier both, this effect modifier as well as generated new variable will be included in the subset of variables for model building.

Need of Adjustment in Explored minimum Sample Size

A minimum sample size required for a study is often explored and targeted. For this, along with other inputs, often an optimal effect size in view of relationship between exposure and outcome is considered. Hence, as a fifth point, considered minimum sample size may sometimes restrict to get accurate analysis and analytical results. In other words, explored minimum sample size[5] may need to be further increased in view of expected lager number of variables to be dealt with while analysing the data especially for model building[7]. In this regard, it is often quoted that there should be a minimum of 10 observations for each variable.

Conclusion

Once a quality data is available under a given study, understanding and application of scales of measurements play a pivotal role in carrying out analysis appropriately leading to accurate results with tremendous strength of implications.

References

1. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996, 15(4): 361-387.
2. Armitage P, Berry G, Mathews JNS. Statistical Methods in Medical Research. 2002 (4th Edition), Blackwell Science Ltd., Oxford, UK.
3. Altman DG. Practical Statistics for Medical Research. 1990 (1st Edition), Chapman & Hall/ CRC, London, UK.
4. Dwivedi SN, Singh Rajvir. On assessing the child spacing effect of breast feeding using Cox-proportional hazards models. Demography India. 2003; 32(2): 160-170.
5. Rubaish A, Wosornu L, Dwivedi SN. Using Deductions from Assessment Studies towards Furtherance of the Academic Program: An Empirical Appraisal of Institutional Student Course Evaluation," iBusiness, 2011, 3(2):220-228.
6. Dwivedi SN. How to Deal with Sample Size Exploration and its reporting in Clinical Research? Central India Journal of Medical Research. 2022; 1(3):1-5.
7. Harrell FE. Regression Modeling Strategies: with applications to linear models, logistic regression, and survival analysis. Springer. 2001.