

ISSN 0970-0218



Apr-Jun 2021 / Vol 46 / Issue 2

Indian Journal *of* Community Medicine

Official Publication of Indian Association of Preventive and Social Medicine

www.ijcm.org.in



Medknow

Medical Biostatistics as a Science of Managing Medical Uncertainties

Abhaya Indrayan

Department of Clinical Research, Max Healthcare Institute, New Delhi, India

Abstract

Biostatistics is generally understood as the branch of statistics that deals with data relating to biological processes. While this remains the core of biostatistics activities, medical biostatistics has additional features. It is rarely realized that the ultimate function of medical biostatistics is to manage medical uncertainties, particularly those that are data based. We propose to define medical biostatistics as the science of managing empirical uncertainties in health and medicine. This definition describes the subject more appropriately and has the potential to put it on a pedestal it deserves because of its focus on medical aspects. This note provides the rationale for this proposal.

Keywords: Definition of medical biostatistics, management, medical uncertainties, variation

INTRODUCTION

From the epitome of crunching numbers, statistical science has traveled a long distance. It is time that it is realized as a management science. This is especially true for medical biostatistics.

Conventionally, biostatistics deals with biological data including agriculture, veterinary science, and fisheries. However, most of us understand biostatistics as a science dealing with the data on life and health of human beings. When restricted to humans, it seems better to qualify this as “medical” biostatistics. This qualifier also makes it more medical than statistical with “medical” + “bio” component exceeding “statistics” component. Another rarely realized feature of this subject is its seminal role in managing data-based medical uncertainties. Accordingly, we propose to define medical biostatistics as the science of managing empirical uncertainties pertaining to human health. This definition provides an entirely new orientation to the subject, integrates it fully well with medical disciplines, and removes its alienation from medical professionals. It can also raise the bar and bring in a new mandate for this subject. This note describes the rationale for this proposal.

As of now, medical biostatistics is rarely recognized as a distinct subject, and even the commonly used name biostatistics has no widely accepted definition. A selected sample of the description of biostatistics at some “reputed” websites is as follows:

1. Using the tools of statistics, biostatisticians help answer pressing research questions in medicine, biology, and public health (Department of Biostatistics, University of Washington, <https://www.biostat.washington.edu/about/biostatistics>)
2. Biostatistics is the application of statistical principles to questions and problems in medicine, public health, or biology (Boston University School of Public Health, http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_BiostatisticsBasics/BS704_BiostatisticsBasics_print.html)
3. Biostatistics is an innovative field that involves the design, analysis, and interpretation of data for studies in public health and medicine (Graduate School of Public Health, University of Pittsburgh, <https://www.publichealth.pitt.edu/biostatistics>)
4. Biostatistics is the development and application of statistical methods to a wide range of topics in biology. It encompasses the design of biological experiments, the

Address for correspondence: Prof. Abhaya Indrayan,
Department of Clinical Research, Max Healthcare Institute, Saket,
New Delhi - 110 017, India.
E-mail: a.indrayan@gmail.com

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Indrayan A. Medical biostatistics as a science of managing medical uncertainties. *Indian J Community Med* 2021;46:182-5.

Received: 04-09-20, **Accepted:** 24-03-21; **Published:** 29-05-21

Access this article online

Quick Response Code:



Website:
www.ijcm.org.in

DOI:
10.4103/ijcm.IJCM_763_20

collection, and analysis of data from those experiments and the interpretation of the results (Wikipedia, <https://en.wikipedia.org/wiki/Biostatistics>)

5. Biostatistics is statistical processes and methods applied to the collection, analysis, and interpretation of biological data and especially data relating to human biology, health, and medicine (Merriam-Webster Dictionary, <https://www.merriam-webster.com/dictionary/biostatistics>).

The common theme in these descriptions of the subject is the processing of biological data – their design, analysis, and interpretation. There is no mention of management or medical uncertainties anywhere. This conventional perception of biostatistics does not do justice to its functions and to the enormous contribution it makes to improve the quality of biological decisions. The perception can be changed by qualifying it by the term “medical” and restricting it to the issues having a direct impact on human health. With this change, the ultimate objective of medical biostatistics too would be to contribute to the efforts to improve people’s health like any other medical science. It already makes this contribution through the efficient management of data-based medical uncertainties, and this must reflect in the definition of the subject. The following details give the rationale of how and why medical biostatistics is the science of managing empirical medical uncertainties.

Medical Uncertainties

Medical uncertainties are well known, but they are easy to appreciate when their presence is realized at two levels. At the individual patient level, it is the potential fallibility of decisions regarding diagnosis, treatment, and prognosis of health conditions. At the group or community level, medical uncertainty comprises a lack of assurance regarding the role of primordial and proximal risk factors of various conditions of ill-health and regarding the exact effect of various promotive, preventive, and treatment interventions. In both these setups, a prominent component is the uncertainty regarding the present state and the future course with or without intervention.

Empirical uncertainties are data based. The handling of such uncertainties is easy when they are divided into aleatory and epistemic components.^[1] These terms may sound new to medicine but are commonly used in seismic science and economics. Aleatory uncertainty arises from endogenous factors such as inherent biological variation, environmental factors, sociocultural and psychological factors, and random variation due to observers, instruments, and laboratories. Epistemic uncertainty arises from a lack of knowledge, conceptual errors, nonavailability of valid tools, and biases of various types. The sources of epistemic uncertainty are exogenous.

Management

Management essentially is a value addition process that tries to optimize the output by properly organizing the inputs. It involves elements such as goal setting; identifying quality and quantity of inputs such as men, machine, methods, material, and money in a production line, and their adequate

and timely provision; minimizing risk opportunities and maximizing conducive environment for optimal functioning of the inputs; gauging performance; and taking rectifying and promoting steps – thus starting the cycle all over again.^[2] Management is a flexible process and does not adhere to consistency and conformity. It is an art of accomplishing an assignment by translating complexity, specialization, and talents into performance.^[2] In the following paragraphs, we examine the application of this management process to medical uncertainties and illustrate how medical biostatistics methods accomplish this.

In the case of management of empirical medical uncertainties, value addition is in terms of controlling these uncertainties so that their impact on decisions is minimal. The description and assessment of these uncertainties are an integral part of this process. Both these activities are done using medical biostatistics methods. The performance is assessed in terms of reaching valid and reliable results. This is the key output in this case also as is for management elsewhere.

Management of Medical Uncertainties

The basic inputs for the management of empirical medical uncertainties are the data. These are invariably inflicted with aleatory variations and epistemic bottlenecks and require expert handling. The study design is a tool that helps to organize these inputs. An immaculately executed perfect design would substantially minimize the risk of reaching an invalid or unreliable conclusion and maximize the power of the study for fixed inputs. Considerations such as the definition of the study units and the variables under consideration, sample size, method of selection, the role of confounders, potential sources of bias including reliability and validity of medical assessments, and the method of analysis of data, are the elements that provide definite help in enhancing the chance of reaching a valid and reliable result. With tools such as probability and its derivatives that include frequency distribution, sensitivity, specificity, relative risk, and odds ratio; estimation methods in terms of effect size, its confidence interval, and meta-analysis; the test of hypothesis for assessing the absence of medically significant effect; and trend analysis that sieves clear signals from noise; medical biostatistics serves the purpose of managing uncertainties quite admirably. Biostatistics models provide the road map to optimize the output in terms of improved results for given inputs. Consideration of various probabilities awards it flexibility instead of consistency and conformity and makes the process of management of uncertainties more efficient and realistic. Decision analysis that combines value judgments regarding the utility of various possible outcomes with the evidence-based risk assessments at the stage of diagnosis and treatment is also an important function of the methods of medical biostatistics. All these help to effectively manage data-based medical uncertainties.

Aleatory uncertainties are the basic ingredients of all statistical methods and can be adequately managed by these methods because these uncertainties arise from variations, and handling

variation is the core of these methods. The same cannot be stated about epistemic uncertainties. Sensitivity analysis^[3] can be effectively used to delineate the impact of some epistemic uncertainties, although not all. However, epistemic uncertainties can be rarely minimized because they belong to the unknown domain. There is no solution for some epistemic gaps except further research because most epistemic gaps are rooted in the lack of knowledge. When the underlying process of emergence and progression of a health condition is unclear, modeling can help understand this process in some cases, although that may have to be based on conjectures in this case. These models may or may not stand the test of the time, but they add to the knowledge base. No science is available that can adequately deal with the unknown except, to some extent, statistics that pools all the unknowns together under the “error term,” provides methods to examine them, and helps to draw a valid inference. Medical biostatistics does all this for human health. The following example illustrates the role of medical biostatistics.

Example: Convalescent plasma therapy for COVID-19 patients

Consider the presently (August 2020) raging coronavirus disease (COVID) epidemic. This is a new disease and so many things are not known about this disease. Let us examine how medical biostatistics can help in managing medical uncertainties regarding some aspects of this disease. The treatment of COVID is in epistemic domain as no treatment is known yet. No sensitivity analysis can be done in this case, but clinical trials on different possible modalities are being conducted around the world to address the uncertainty regarding treatment modality. Most prominent of these is convalescent plasma therapy which had given encouraging results earlier with other serious infectious diseases.^[4] We explain how medical biostatistics helps in managing medical uncertainties in this setup.

There is a considerable uncertainty regarding the actual efficacy of plasma therapy relative to the standard care in serious and critical patients. If it is effective, how much and what kind of patients are benefitted – young or old, with comorbidities or without comorbidities, serious or critical, patients going to intensive care unit (ICU) or those not going to ICU – and how the hospital length of stay is affected, what kind of side effects does it express, for how long does it protect, and so on so forth. To keep the discussion manageable, let us restrict to mortality as the outcome of interest. The effect, thus, would be the difference in mortality this treatment makes in serious and critical patients relative to the standard care. To begin with, this requires that a medically significant effect is specified. This is a clinical problem but determining a sample size to be able to detect that kind of effect when present with a specified power requires statistical calculations. This addresses the uncertainty regarding how big a trial should be.

A randomized control trial provides highest level of evidence for all such new regimens but, in this case, getting informed consent could be difficult and other ethical issues may arise^[5]

because critical patients may resent if plasma is not given to them. A large percentage of patients perceive this therapy as a solution to their ailment. Thus, the applicability of the results would remain suspicious. Aleatory uncertainties regarding the effect of moderator (preexisting) variables on the outcome can be managed by ensuring baseline equivalence of the cases and controls by the statistical strategy of randomization if the sample size is large or by matching if the size is small. As just mentioned, randomization can cause ethical problems in allocation of a patient to the control group, whereas matching is feasible in a record-based study. The moderator variables in this case include age, gender, and number of comorbidities, which need to be matched. If they do not match, help may have to be taken from tools such as standardized mortality ratio or logistic regression to find the adjusted effect. These statistical tools help to considerably alleviate the uncertainties regarding the actual effect and give a credible result.

A medical biostatistician would know about the disruptive effect of mediator variables. In this case, these are the use of ventilator and admission to ICU that can substantially alter the outcome. Management of uncertainty regarding the effect of these mediators requires either poststratification or calculation of adjusted rates. Both are fundamentally statistical solutions. The management of the effect of mediator variables becomes even more intricate if the trial is multicentric because each hospital may have its own set of guidelines despite agreeing to a common protocol.

Once the uncertainty regarding the effect of the moderator and mediator variables is managed through steps outlined above, much of the difference in mortality between the two groups can be safely attributed to the treatment – plasma therapy in this case. However, some unexplained variation may remain due to unknown or unanticipated factors. Statisticians deal with them as random error. If this error is substantially less than factor variance, a valid result regarding efficacy can be obtained. Statistical tests of significance are used for this purpose. The significance should be tested for the presence of medically significant effect and not for null effect.

CONCLUSION

The explanation in the preceding paragraphs and the example may convince that it is appropriate to define medical biostatistics as the science of managing medical uncertainties, particularly those that arise from data in an empirical setup. No definition is perfect or acceptable to everyone, but it seems to describe the subject adequately. It highlights the relevance of the subject for medicine and raises the bar with a new mandate for the subject with a commitment to the human variation and biological processes. “Management” brings a new responsibility that is lacking so far and increases the accountability of the subject. This is important in view of severe criticism of some of the statistical methods.^[6] As stated earlier, this definition emphasizes that “medical” and “bio” are the dominant part of medical biostatistics and exemplifies the

complete fusion of statistics with medicine that Feinstein^[7] emphasized so much. Conventionally, biostatistics has come to be identified with medicine rather than other biological disciplines – thus restricting it to the medical uncertainties pertaining to human health looks appropriate. The ambiguity is removed by calling it medical biostatistics, instead of just biostatistics.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Indrayan A. Aleatory and epistemic uncertainties can completely derail medical research results. *J Postgrad Med* 2020;66:94-8.
2. Magretta J. What Management Is: How It Works and Why Its Everyone's Business. New York: The Free Press; 2002. p. 1-4. Available from: https://www.academia.edu/8327880/What_Management_Is_How_It_Works_and_Why_It_s_Everyone_s_Business. [Last accessed on 2020 Sep 04].
3. Saltelli A, Tarantola S, Campolongo F, Ratto M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Chichester: John Wiley; 2004. Available from: http://www.andreasaltelli.eu/file/repository/SALTELLI_2004_Sensitivity_Analysis_in_Practice.pdf. [Last accessed on 2020 Sep 04].
4. Sun M, Xu Y, He H, Zhang L, Wang X, Qiu Q, *et al*. A potentially effective treatment for COVID-19: A systematic review and meta-analysis of convalescent plasma therapy in treating severe infectious disease. *Int J Infect Dis* 2020;98:334-46.
5. Perotti C, Del Fante C, Baldanti F, Franchini M, Percivalle E, Vecchio Nepita E, *et al*. Plasma from donors recovered from the new Coronavirus 2019 as therapy for critical patients with COVID-19 (COVID-19 plasma study): A multicentre study protocol. *Intern Emerg Med* 2020;15:819-24.
6. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond “ $p < 0.05$ ”. *Am Stat* 2019;73 Suppl 1:1-9.
7. Feinstein AR. *Clinical Biostatistics*. Saint Louis: The CV Mosby Company; 1977. p. 4.