Ensconcing a biostatistics clinic in tertiary care research institute of India: A descriptive study

Kishore Kamal¹, Meenakshi Sharma², Mahajan Rahul³, Kapoor Rakesh⁴, Vipin Koushal⁵

¹Departments of Biostatistics, ²Community Medicine and SPH, ³Dermatology Veneereology Leprology, ⁴Radiotherapy and ⁵Hospital Administration, Postgraduate Institute of Medical Education and Research, Chandigarh, India

ABSTRACT

Context: Data collection and statistical analysis are integral components of research. The beauty of statistics lies in its ability to evaluate evidence in the face of uncertainty. However, lack of dedicated biostatistical consultation units, rote academic teaching and training lead to poor statistical analysis. Thus, we aim to explore and understand the challenges of establishing a Biostatistics Clinic (BC) in a tertiary care research institute. A secondary aim is to identify the stage of research at which participants approach biostatisticians. Material and Methods: The data for the current study came from a consultancy unit named as “Biostatistics Clinic” in the department of biostatistics from Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India. The residents and students who approached the department of biostatistics for consultancy regarding study design, sample size, statistical analysis and other analytical work were the sampling units. Results: A total of 208 residents and students made 404 visits to biostatistics clinic. The male and female visits were 118 (56.7%) and 90 (43.3%), respectively. Majority of visitors approached for data analysis (171; 75%) followed by study design and sample size calculation for protocol preparation (43; 18.9%). Leading reference to approach the biostatistics clinic was through a faculty (91; 43.8%) followed by self (54; 26%). Conclusion: Despite the thrust and apparent advantages of contacting statistician at the beginning of the study, majority approached only at the data analysis stage. Therefore, repeated and improved efforts are required to spread the message of approaching statistician early.

Keywords: Analytical support clinic, biostatistical consulting, biostatistics clinic, statistical consultancy

Introduction

Data collection and statistical analysis are integral components of research. The individual, family, and community practitioner must decide between competing interventions based on evidence. The decision to provide the best patient outcome requires a robust data-based decision to segregate uncertainty from certainty. The beauty of statistics lies in its ability to evaluate evidence in the face of uncertainty. However, its improper usage to obtain P value and confidence intervals clouds the reliability and validity of the studies. The unscientific research and data from the same hamper patient care. Additionally, biostatistics’ incomplete and incorrect usage has emerged as a significant reason for rejecting articles. The reasons for this could be the lack of rigorous statistical training, statistical collaborations, and the use of software without technical understanding among researchers. Another reason can be consulting a statistician after the data collection instead of the planning stage. The literature recommends approaching a statistician at the beginning of the research. The replication crisis often occur due to not involving a statistician at the conceptualisation stage that resulted in

Address for correspondence: Dr. Kishore Kamal, Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh - 160 012, India. E-mail: kkishore.pgi@gmail.com

Received: 17-08-2021 Accepted: 16-12-2021 Published: 14-05-2022

Access this article online

Quick Response Code:

Website: www.jfmpc.com

DOI: 10.4103/jfmpc.jfmpc_1653_21

How to cite this article: Kamal K, Sharma M, Rahul M, Rakesh K, Koushal V. Ensconcing a biostatistics clinic in tertiary care research institute of India: A descriptive study. J Family Med Prim Care 2022;11:2051-5.
Numerous studies emphasised the need and setting up of biostatistical consultations.\textsuperscript{17–19} Costing studies in the literature have suggested the worth of such consultation units.\textsuperscript{20–22} Many tertiary care institutes in developed countries have separate biostatistical consultation units. However, the biostatistics departments in India were established primarily for teaching purposes, and there are no particular consultation units with dedicated time for consultations. Hence Biostatistics Clinic (BC), a first of its kind, was established at a tertiary care institution in northern India in 2018. The primary objective of the current study is to explore and understand the challenges of establishing a BC at a tertiary care institution in India. A secondary aim is to identify the stage of research at which participants approach biostatisticians.

### Methodology

The current study was conducted in the Post Graduate Institute of Medical Education and Research (PGIMER), a 2,200 bedded multi-speciality tertiary care research institute and “Institute of National Importance” situated in Chandigarh, India. The communication to initiate the biostatistics clinic was circulated offline (official letters to each department, oral communication) and online (social media, email, and message) modules to the resident doctors, faculty, and students for broader reach. Subsequently, emails, messages, and WhatsApp were used as reminders to utilise the biostatistics clinic service. The consultancy time was kept from 2:00 to 5:00 pm every working day to facilitate the clinicians. The timings of consultations were flexible and adjusted on individual requests. The beneficiaries were requested to officially email or call to fix an appointment to avoid inconvenience and time waste. Although participants were asked to visit departments, few consultations were given through WhatsApp, Skype, and emails. In general, we decided to keep half an hour for each consultation-the same was extended or reduced depending on the problems and requests for consultations. The participants’ visit record, such as the purpose of the visit, date, time, speciality, designation, mobile number, and email, was maintained.

### Sample size

We did not formally calculate sample size as the current study is hypothesis-generating in nature rather than hypothesis confirmation.

### Data analysis

The resident doctors and students approached the BC for consultancy regarding study design, sample size calculation, statistical analysis, and other analytical work were the units of data collection and analysis. The variables of interest were statistical consultation types (statistical analysis, sample size, study design), refer by whom (faculty, self, friends), stage of consultation (protocol conceptualisation, after data collection or all the stages), and time (before lunch, after lunch or after duty hours). Subsequently, a spreadsheet was used to enter data from the hard copy specifying the purpose of the visits. We used descriptive measures such as frequency (%), bar chart, and pie-chart to report study findings using MS excel\textsuperscript{\textregistered} (Microsoft, Washington (US)).

### Results

A total of 208 students made 404 visits to the biostatistics clinic [Table 1]. The male and females visits were 118 (56.7%) and 90 (43.5%), respectively. Leading reference to the biostatistics clinic was through a faculty (43.8%) followed by self (26%). More than 2/3\textsuperscript{rd} (143) of the visitors approached once, followed by 2-4 times (24%) and ≥ five times (7.2%). An overwhelming majority of visits came for data analysis (75%) followed by study design and sample size calculation for protocol preparation (18.9%). The most frequent visitors were the junior residents (42.7%), followed by senior residents and doctorate students (18.8%). Majority (69.1%) of visitors approached post-lunch (2:00 – 5:00 pm) followed (13.6%) by before lunch (9:00 – 1:00 pm) and after (9.9%) office hours (5:00 – 8:00 pm) respectively. Most (90%) of the visitors called and fixed an appointment before visiting.

We found that only 65 (31.3%) students out of 208 came for repeated consultations. However, only five (2.4%) students visited the clinic, from conceptualisation to data analysis.

### Table 1: Characteristics of Visitors to Biostatistics Clinic in a Tertiary Care Medical Research Institute

| Variables              | Categories | Count (%) |
|------------------------|------------|-----------|
| Gender                 | Male       | 118 (56.7)|
|                        | Female     | 90 (43.3) |
| Reference              | Faculty    | 91 (43.8) |
|                        | Self       | 54 (26.0) |
|                        | Friend     | 36 (17.3) |
|                        | Not Specified | 23 (11.1) |
|                        | Notice     | 4 (1.8)   |
| Consultation frequency | 1 Time     | 143 (68.8)|
|                        | 2-4 Times  | 50 (24.0) |
|                        | ≥ 5 Times  | 15 (7.2)  |
| Purpose\textsuperscript{*} (n=228) | Data Analysis | 171 (75.0) |
|                        | Protocol Preparation | 43 (18.9) |
|                        | Interpretation | 13 (5.7)  |
|                        | Software & Data Entry | 1 (0.4)   |
| Designation\textsuperscript{*} (n=213) | Junior Residents | 91 (42.7)  |
|                        | PhD        | 40 (18.8) |
|                        | Senior Residents | 40 (18.8) |
|                        | Research Staff | 24 (11.3) |
|                        | Postgraduate | 18 (8.4)  |
| Appointment\textsuperscript{*} (n=404) | 09:00-01:00 pm | 55 (13.6) |
|                        | 01:00-02:00 pm | 30 (7.4)  |
|                        | 02:00-05:00 pm | 279 (69.1) |
|                        | 05:00-08:00 pm | 40 (9.9)  |

\textsuperscript{*}Few residents and students came for more than one purpose. "Few Students designation changes during the study period. "A total of 404 appointments was given to 208 residents and students.
The offer of authorship at data analysis stages often leads to multiple testing for hunting statistically significant P. The unidirectional discussion without a holistic overview of study characteristics may lead to Garbage (low powered studies, flawed study design, ill-conceived master sheet) In, Garbage (non-replicable studies, inappropriate statistical analysis, Incorrect results) Out (GIGO).

The partial or incomplete dataset leads to repetitive preparation and coding of the same dataset, multiple testing, repetition of similar analysis on different occasions. It can blur the distinction between independent and dependent variables. The emergency approach in analysis often leads to ignorance of vital intricacies of the study design, assumptions of the datasets. Further, it may lead to incorrect statistical analysis and interpretations of the data. The offer of authorship at data analysis stages often leads to data torturing to hunt for a P<0.05 or other statistically significant criteria. It eventually leads to inflation of Type-I error and is a substantial contributor to the replicability crisis.

Residents and students flow to the clinic were regular in each month. However, there were spikes during the thesis and protocol submission months [Figure 1]. The thesis and protocol submission months in the institute are June and December. The participation by the clinical (39%) and non-clinical (34%) departments were almost the same [Figure 2]. Despite the availability of consultants and repeated reminders, most researchers (97.6%) do not follow the standard protocol of approaching at the planning stage. Moreover, numerous investigators (>90%) come for data analysis without a reference proposal. The majority (43.8%) of the participants seeking biostatistical consultations were referred by faculty. Therefore, it is crucial to periodically disseminate the information about the biostatistics clinic among the faculty colleagues, at least in the beginning, till the system becomes self-sustained. The majority of the studies emphasised developing collaborations between researchers and biostatistics consultants. However, ours is perhaps the first study to empirically demonstrate the importance of collaborations. Self-reference (26%) indicated that participants were aware of the need for an expert opinion on biostatistics. It could be possible due to increasing emphasis by methodologists, prior exposure, felt need, or lack of expertise in statistical analysis. However, there was a spur in seeking consultations during protocol and thesis submission months. Such exigencies were managed with flexible timings and a checklist for discussion in our clinic.

There were many challenges faced at the various stages of the biostatistics clinic. The timing and flexibility play a significant role in establishing the biostatistics clinic. Despite an emphasis on prior appointments to save time, many dropped in both morning and evening. Similarly, many of them were not punctual about prior appointments to save time, many dropped in both morning and evening. Similarly, many of them were not punctual about prior appointments to save time.

The current study showed that, despite the emphasis on consultation during the planning stage, most visitors (75%) approached at analysis stage after data collection. This situation is known as “Data repairer.” A significant challenge was balancing study design, objectives and variables, sample size, and selecting appropriate statistical techniques. During the analysis stage, approaching students were the most fixated for specific analysis, sample size justification despite apparent flaws. Most people coming at this stage were restless to hunt for P < 0.05, alternatively known as “p-value hacking.” Many methodologists have raised the alarm against the perils of P value hacking. This stage frequently leads to a precarious situation of slicing and dicing the data by researchers without a significant P value (p < 0.05) from the data. It eventually inflates type-I error. The inflation of error leads to non-replication of study findings and has emerged as a significant threat off late.

The inflation of error leads to non-replication of study findings and has emerged as a significant threat off late. The inflation of error leads to non-replication of study findings and has emerged as a significant threat off late.
P value paradigm, and they will not achieve nirvana (salvation) if the P value does not turn out to be significant. They required the gentlest touch, persuasion, teaching the difference between clinical and statistical significance, the importance of negative trials, and the need to approach the statistician at the conceptual stage of the study.

In the initial stages, many residents came with a pre mindset that statisticians enter the numbers in the computer and then computer do the analysis. Therefore, it is going to be a 5-minute job. This misapprehension guided them in contacting biostatistical consultancy 10-15 days before submitting protocols and theses. However, the single-visit approach to solving all the problems is flawed because it is difficult to understand the subtle intricacies of the studies and apply appropriate statistics in one go, despite, availability of computers. Kreft has succinctly summarised, “no piece of hardware can solve the challenging statistical issues underlying decisions about model specification”. Many resident doctors and students reported P value without confidence intervals (CI). Moreover, the mean and standard deviation were standard reporting measures irrespective of the skewed nature of the data. In our clinic, this was an opportunity to educate researchers to carefully plan and sync “research question, study design, sample size, data collection and entry, analysis and display of results with appropriate descriptive and inferential measures” at the conceptual stage.

Despite recommendations, individual, family, and community practitioners do not involve biostatisticians during study conceptualization. The studies without biostatisticians lead to many methodological flaws in data collection, analysis and interpretations that may adversely affect patient care.

**Conclusion**

Despite the thrust and apparent advantages of contacting statisticians at the beginning of the study, the majority approached only at the data analysis stage. Therefore, repeated and improved efforts are required to spread the message of consulting a statistician at the planning stage through BCs. Further, consultation units may be established at other institutions to strengthen and consolidate the research in the country. The strengthening of BCs and collaboration with biostatisticians will also enhance the best possible care for patients.

**Circular**

The lead author of this article officially started the biostatistics clinic on 10th May 2018 vide letter no. PGI/BIOSTAT/2018/87-143 after pilot testing.

**Financial support and sponsorship**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflicts of interest**

There are no conflicts of interest.

---

**References**

1. Romano R, Gambale E. Statistics and medicine: The indispensable know-how of the researcher. Transl Med UniSa 2013;5:28-31.
2. Association AS. Statistics Improve Health Care. Statistical Significance. 2017. p. 2017. Available from: https://www.amstat.org/asa/files/pdfs/StatSig/StatSigHealthCare.pdf. [Last accessed on 2019 Jul 03].
3. Grindstaff TL, Saliba SA. Avoiding manuscript mistakes. Int J Sports Phys Ther 2012;7:518-24.
4. Ali J. Manuscript rejection: Causes and remedies. J Young Pharm 2010;2:3-6.
5. Sullivan GM. What to do when your paper is rejected. J Grad Med Educ 2015;7:1-3.
6. Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. J Prosthodont Res 2011;55:1-6.
7. Ezeala CC, Nweke IN, Ezeala MO. Common errors in manuscripts submitted to medical science journals. Ann Med Health Sci Res 2013;3:376-9.
8. Zapf A, Rauch G, Kieser M. Why do you need a biostatistician? BMC Med Res Methodol 2020;20:23.
9. Esterhuizen TM, Li G, Young T, Zeng J, Machekano R, Thabane L. Advancing collaborations in health research and clinical trials in Sub-Saharan Africa: Development and implementation of a biostatistical collaboration module in the Masters in Biostatistics Program at Stellenbosch University, Trials 2021;22:478.
10. Kass RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N. Ten simple rules for effective statistical practice. PLoS Comput Biol 2016;12:e1004961.
11. Adams-Huet B, Ahn C. Bridging clinical investigators and statisticians: Writing the statistical methodology for a research proposal. J Investig Med 2009;57:818-24.
12. Carley S, Lecky F. Statistical consideration for research. Emerg Med J 2003;20:258-62.
13. Hensel WM. Double trouble? The communication dimension of the reproducibility crisis in experimental psychology and neuroscience. Eur J Philos Sci 2020;10:44.
14. Lewandowsky S, Oberauer K. Low replicability can support robust and efficient science. Nat Commun 2020;11:358.
15. Anvari F, Lakens D. The replicability crisis and public trust in psychological science. Compr Results Soc Psychol 2018;3:266-86.
16. Kerr NL. HARKing: Hypothesizing after the results are known. Pers Soc Psychol Rev 1998;2:196-217.
17. Perkins SM, Bacchetti P, Davey CS, Lindsell CJ, Mazumdar M, Oster RA, et al. Best practices for biostatistical consultation and collaboration in academic health centers. Am Stat 2016;70:187-94.
18. Welty LJ, Carter RE, Finkelstein D, Harrell FE Jr, Lindsell CJ, Macaluso M, et al. Perspective: Strategies for developing biostatistics resources in an academic health center. Acad Med J Assoc Am Med Coll 2013;88:454-60.
19. Arndt S, Woolson RF. Establishing a biostatistical core unit in a clinical research center. Am Stat 1991;45:22-7.
20. Rahbar MH, Dickerson AS, Ahn C, Carter RE, Hessabi M, Lindsell CJ, et al. Characteristics of biostatistics, epidemiology, and research design programs in institutions with clinical and translational science awards. Acad Med J Assoc Am Med Coll 2017;92:229-36.
21. Rubio DM, del Junco DJ, Bhore R, Lindsell CJ, Oster RA, Wittkowski KM, et al. Evaluation metrics for biostatistical and epidemiological collaborations. Stat Med 2011;30:2767-77.

22. Zhang G, Chen JJ. Biostatistics faculty and NIH awards at US medical schools. Am Stat 2015;69:34-40.

23. Adda J, Decker C, Ottaviani M. P-hacking in clinical trials and how incentives shape the distribution of results across phases. Proc Natl Acad Sci 2020;117:13386-92.

24. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol 2015;13:e1002106.

25. Szucs D. A tutorial on hunting statistical significance by Chasing N. Front Psychol 2016;7:1444.

26. Miłkowski M, Hensel WM, Hohol M. Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. J Comput Neurosci 2018;45:163-72.

27. Baker M. Is there a reproducibility crisis? A nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help. Nature 2016;533:452-5.

28. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One 2009;4:e5738.

29. Hengartner MP. Raising awareness for the replication crisis in clinical psychology by focusing on inconsistencies in psychotherapy research: How much can we rely on published findings from efficacy trials? Front Psychol 2018;9:256.

30. Tobi H, Kuik DJ, Bezemer PD, Ket P. Towards a curriculum for the consultant biostatistician: Identification of central disciplines. Stat Med 2001;20:3921-9.

31. Kreft IGG. Special issue: Hierarchical linear models: Problems and prospects. J Educ Behav Stat 1995;20:109-13.