Why most research manuscripts get rejected by leading medical journals? - Part 1

This topic has been discussed to death; yet, it lives on. Undoubtedly, the objective of every researcher is to publish the results of the research, given the tremendous labor that goes into every clinical research. Besides, it is through sharing of research outcomes that medicine has come this far and continues to scale newer milestones. The current COVID-19 pandemic has driven home this message loud and clear even among the general population. Nevertheless, a minuscule number of submitted manuscripts submitted to the leading indexed journals see the light of the day. That said, the number of articles that do make it to the publication stage every year is humongous. A September 2018 report by University World News stated, 'No one knows how many scientific journals there are, but several estimates point to around 30,000, with close to two million articles published each year.' That is indeed a HUGE number. What is it then that cuts the ice?

The list of reasons for rejection is long, and the common reasons that come to mind are in fact a small proportion of the causes for rejection. While there have been numerous articles over the years listing these causes for rejection, it might be helpful to look into each cause with some real-world examples. The following list is a very broad overview of why a research manuscript might be rejected. I have not listed some very obvious reasons like lack of ethics approval, lack of patient consent, and plagiarism because I think we have reached a stage where every researcher is aware of these laws and the precondition of every journal that the researchers abide by these laws. Furthermore, it is assumed here that the manuscript is not outside the scope of the target journal. These mandatory tenets being met, we will look at what are the reasons for rejections.

In a series of articles on the same topic over the next few weeks, we will look at each factor listed below in detail, with some real-world examples. These, I hope, will help researchers right from the planning stage, to ensure that their research has a good chance of getting published. The following issues will be discussed in the subsequent articles.

1. Objectives of the study.
2. Type of study.
3. Study design.
4. Sample size and statistical analysis.
5. Strength of the data
6. Novelty of the research.
7. Quality of the manuscript.
8. Non-adherence to the journal guidelines.
9. Miscellaneous causes like formatting issues, inconsistent data and others.

Why most research manuscripts get rejected by leading medical journals?- Part 2: Objectives and study design

In the first part above, we saw a general overview of the reasons why most articles on medical research submitted to indexed journals with a high impact factor get rejected. We will now get into the details of each with some examples.

Perhaps, the commonest yet most overlooked causes for rejection are the study objectives and design. Several factors need to be considered before the objective(s) and design/type of the study are decided. Objectives and study type and design go hand-in-hand and cannot be discussed exclusively from each other. What often happens is that researchers either have data that they wish to get published as a retrospective study (because I have good and large amount of data, why not publish it!), or they observe some trend in their clinical practice and think this can be conducted as a prospective study and published. However, in reality, these are endpoints and not starting points for a study. I will start with a simple example that all of us can relate to, and follow the same example to see how the objectives and study type/design need to be planned. Please note here that this is a hypothetical example and the data mentioned regarding the outcomes is not necessarily true. Further, in this part, we will focus on a retrospective study and will take up the prospective studies in part 3.

Let us say you wish to publish a study that shows metformin still plays an important role in diabetes management in the era of DPP4s. You might already have the data of several patients on metformin or a DPP4 alone or on a combination of both, or were switched from metformin to DPP4, or were on one of them and the other was added. Let us assume the data shows better glycemic control with a combination of DPP4 with Metformin rather than DPP4 alone. Since the data already exists, it is handed over to a statistician and after the analysis is done, the manuscript writing starts with an objective to submit it to the best journals with the highest impact factor. This is exactly where the plot goes wrong. A simple search might show that there are already numerous studies that have reached the same conclusion. Why would the journal want to publish your data? At the same time, since there are many studies with similar outcomes, it is obvious that despite similar outcomes numerous such studies were published. This is where the study design becomes important.

A good way to start is to take a look at the studies with identical outcomes (better glycemic control with a combination of DPP4 and Metformin rather than DPP4 alone) published in the leading journals with high impact factors say in the last 10 years, especially those that have been cited frequently, and to go through the methodologies and limitations of these studies. This not to replicate the study designs that have been followed, but to know what parameters make each of these studies different from each other, although their conclusions **appear similar.** The outcomes might have been better/poor in some patient profiles than in others or there might be a correlation of the outcomes with some factors. Some examples could be:

1. Age group
2. Male/female sex
3. A particular DPP4 rather than DPP4 as an entire class
4. Region/race/ethnicity of the subjects
5. Number of years since the diagnosis of diabetes
6. Duration of followup
7. Associated secondary outcomes like renal function, cardiovascular events, weight loss/gain, lipid profiles, etc.
8. Association with coexisting factors like weight, physical activity, etc at baseline.
9. Pre-existing factors that could have affected the outcomes e.g. hypertension, history of myocardial infarction, etc.

Now with a fair idea of what data is already available and where the gaps are, it might be good to go back to your data and think about what can be the highlight of your study based on the data you have and what is already published. As an example, maybe most studies have included all the DPP4s as an entire class of drugs. Whereas in your data maybe 50-60% of the cohort was on a particular DPP4. That can be the strength rather than the

weakness of the study. It might make sense to include only this cohort and leave out the rest so that the strength of the data is robust. Another example - maybe it has been already published that the combination of metformin with DPP4 shows better outcomes in patients who were obese at baseline. Can you make it more specific by analyzing the outcomes by different BMI levels rather than obesity in general? Again, it would depend on your data- what parameters have been monitored and documented that can be used to create a study design that makes a meaningful contribution to research, which the journal editors think their readers will be interested in. The flipside here could be that if you filter down the data to such levels, your sample size might become very small, which will discuss in the next paragraph. However, the positive is that you now have a clear focus for your study, which means a higher chance for publication, and if the sample size is small, it is better to wait until you have data from more patients before starting work on the analysis. Once there is clarity about the focus area, it is important to have a research question (study hypothesis) in mind, as that will decide the design that would be best. Some examples could be:

1. Metformin + DPP4 has better glycemic control in those with BMI 25-30 kg/m^2 than in those with BMI > 30 kg/m^2
2. Metformin + DPP4 has better glycemic control when started at the time of diagnosis than when the 2nd drug is added later.
3. Metformin + DPP4 has better glycemic control in those with diastolic blood pressure <80 mm Hg than in those with diastolic BP> 80 mm Hg **or** Metformin + DPP4 abcgliptin has better glycemic control in those with diastolic blood pressure <80 mm Hg while Metformin + DPP4 xyzgliptin has better glycemic control in those with diastolic BP> 80 mm Hg
4. Among patients who did not show response to Metformin or DPP4 montherpay, more patients below 40 years of age were prescribed Metformin + DPP4 abc gliptin while those above 40 years were prescribed Metformin + DPP xyz gliptin.
5. Metformin + DPP4 showed improved glycemic control within 3 months in those with normal serum creatinine while the same results were seen in 6 months in those with elevated serum creatinine.
6. Metformin + DPP4 showed improved glycemic control in patients with no hyperlipidemia but not in those with high lipid levels.

Among the examples of study hypotheses above, you can see that for the qts 5, at least a 6 months' data is required (longitudinal study). In example 4, data a single timepoint would suffice (cross-sectional study). For examples 1-3, the researcher would need to decide the cutoff time period that will be considered for the study (cohort studies). Example 6 compares the outcomes in those with the presence of a particular risk factor vs those who do not have that risk factor. (case controlled study). Next, the sample size and statistician comes in. You need to provide him/her details of the focus and objective of your study as also the study hypothesis, and share the available literature so that the statistician can suggest an appropriate sample size with a rationale for the same. Not all journals ask for the rationale for sample calculation for a retrospective study, but many do. If you wish to target specific journals for your manuscript, it is good to go through the journal guidelines at this stage to know if sample size calculation is required. At the end of all the above work, you now clearly know what is the data and information required if you wish to submit the planned study for publication in an indexed journal. Thus, there is clarity on what is already available with you, what is useful, what should be discarded, what is still missing.

We will look at prospective studies in the next part of this series.

Why most research manuscripts get rejected by leading medical journals?- Part 3: Prospective study design

In this part, I will focus on prospective research. For the ease of understanding and continuity, I will take the same example as that used in part 2 i.e., you wish to publish a study that aims to explore whether metformin still plays an important role in diabetes management in the era of DPP4s. In the previous article, we saw how a retrospective study design on this topic can be made to stand out from the existing studies, so as to have a good chance of an indexed journal being interested in publishing it. In this part, with the same example, we will see how a publication-worthy prospective study can be designed and more importantly, the pitfalls to avoid. First, some basics about the types of prospective studies:

Prospective studies can be interventional or non-interventional. As the terms suggest, in an interventional study, the treatment might be changed during the course of the study based on a pre-designed plan in the protocol, or it might be changed based on certain outcomes. In non-interventional studies, the same treatment would continue throughout the duration of the study. One more type of non-interventional study is a cohort study which recruits people with a similar condition (disease and/or treatment) and collects information about them for a number of years.

If the study has 2 or more arms (as most prospective studies do), another aspect that needs to be considered is patient allocation. In this regard, double-blind studies with random allocation are considered as the highest level of evidence and are more likely to have higher acceptability.

So, in a prospective study that aims to explore whether metformin still plays an important role in diabetes management in the era of DPP4s, the examples of the above types of prospective studies could be:

A study with subjects recruited and allocated to 2 arms: one group is administered only DPP4 and another administered DPP4+Metformin, and the outcomes followed over say 2 years. This would be a prospective, non-interventional study.

Subjects recruited and allocated to 2 arms: one group is administered only DPP4 and another administered DPP4+Metformin. However, after 12 weeks of treatment, if the HBA1c level is more than 6.5, another drug is added. The drug might be pre-specified in the protocol or might be at the discretion of the treating physician. This would be a prospective, interventional study.

A study where all subjects in one group are on DPP4+Metformin and the other group on DPP4 alone and are followed over a long term for e.g., 10 years or 20 years to see how many patients in each group develop adverse outcomes associated with diabetes, like myocardial infarction, stroke, renal failure, etc, would be a cohort study.

In each of the examples, the subjects might be randomly or non-randomly allocated to the 2 arms depending on certain baseline criteria defined in the protocol. Again, these can be single-blind or double-blind, wherein the researcher as well as the patients do not know what drug is being administered, while the person who is treating the patients is not aware about the study, so that there is no bias in which patients gets what treatment.

So with the same study (better glycemic control with a combination of DPP4 with Metformin rather than DPP4 alone) being conducted retrospectively, we saw in part 2 of the series, what could be the pitfalls that need to be avoided. When the same study is planned to be conducted prospectively, the do's and don'ts are somewhat different. I am listing below a sort of checklist that needs to be considered before planning a prospective study so that after all the effort it is not rejected by the journals due to a flaw in the design. Some factors in the list below remain the same as those listed in part 2 of the series.

What is your study hypothesis (research question) and what data is already published from previous studies? I have explained these with detailed examples in part 2 ; hence, will avoid a repetition here.

Why do you want to a prospective study? This is an important question because there are many ethical and legal factors involved apart from the time and cost. Assuming the time and cost factors are taken care of, what is the contribution this research will make to the existing literature for which it is worth performing the study despite the ethical and legal challenges? Examples of such concerns could be - obtaining informed consent from the patients, possible legal implications in case of any serious adverse events, insurance cover for the patients is required in some countries for prospective studies, justifying the intervention in case of interventional studies, justifying the random allocation to different treatment arms, and others. Is this study likely to bring out something new which could have future implications in the understanding/treatment of the disease and/or use of certain drugs? Hence, minute scrutiny of all research on the topic that is already available, what are the gaps in existing literature, and your research question becomes absolutely essential. At this stage it should also be decided as to what patients profiles to the minutest details for inclusion can you define that will make your study different from those in the existing literature.

Sample size calculation - This is an absolute must in prospective studies and most studies will ask for the details of how it was calculated, what was considered, what was not considered and why.

The biggest pitfall that gets overlooked is the analytical plan. It is often assumed that the analysis will be done by the statisticians once the data is available. In fact the analytical plan is most important even before writing the protocol or allocating funds to the study. The analytical plan will determine the inclusion and exclusion criteria as also the robustness of the data. Let me give a couple of examples. I know of a prospective study which was quite novel and multicentric, with random allocation but was refused publication by all leading journals because it captured too much data. Despite a sample size of 1500, the protocol did not define the treatment regimen and left it to the discretion of the treating physicians, due to which at the stage of analysis it was realized that there are humungous number of subgroups, effectively having each subgroup with small samples, which made arriving at any meaningful conclusion of the primary objective impossible. Another prospective study about the effectiveness of a procedure was returned by several journals because the inclusion criteria were not robust, as a result of which the baseline characteristics of the included subjects was very diverse in terms of the primary disease. Thus, each subgroup in itself had a very small sample size. The analytical plan should also discuss how the primary and secondary objectives and subgroups will be analyzed. After the plan is discussed, you might often find that you need to redefine the patient profiles (which was done at step 2 above) for inclusion in the study.

Only when in-depth answers to all the above questions ae ready and have been discussed and debated among the investigators multiple times, should even the protocol writing begin. Most importantly, the statistician should a part of all meetings and discussions between the investigators. The statistician should also be allowed to have a look at the protocol before it is sent to the ethics committee or other relevant authorities for approval.

So with part 2 and this part 3 of the series, we have adequately covered the study types and designs for researchers to know what to consider before even planning a study, which is eventually intended for a publication.

Why most research manuscripts get rejected by leading medical journals?- Part 4: Writing the manuscript and Journal guidelines

In this final part, I will focus on writing the manuscript. Take a look at the image above. Even a Nobel laureate's manuscript could be rejected!! It might not necessarily mean the research was poor. It is about how well you can communicate about the research and its outcomes. I will divide this chapter into 2 parts:

1. The content of the manuscript itself.
2. The journal guidelines.

Content of the manuscript

Assuming you have followed everything mentioned in chapters 2 and 3, by now you would have a robust study design, the objectives well defined, a statistical analysis plan before data collection, robust and complete data, and the statistical analysis completed. The manuscript should be able to cover all of it comprehensively, without deviating from what was planned in the protocol. Of course, the journal is unlikely to ask for a copy of the protocol, but non-adherence to the protocol could affect the quality of data and the research itself, and the reviewers could have a lot of questions that need to be convincingly answered. The standard structure for a research article more or less is the same: Abstract, Introduction, Methods, Results, Discussion, Conclusion, References. If you have followed all the steps and guidelines mentioned in part 2 and 3 of this series, writing the manuscript is unlikely to be challenging. However, the challenge lies in the Journal guidelines, which I cover below.

Journal guidelines:

Most indexed journals have very strict formatting guidelines and absolute adherence is a must. Moreover, the list of guidelines is long and even a single punctuation error is unacceptable. I am listing below examples of most guidelines I have come across, though all journals might not necessarily have all of them. However, the big ones like Lancet will usually have all of these guidelines.

In case it is difficult for the author to format the manuscript according to the guidelines, it might be better to seek professional help.

English: The journal might specify which English is to be used- American or British. It makes a big difference because some spellings and conventions differ between the two. For e.g., 'analyzed' vs. 'analysed.' In British English, dates are presented in the order of date, month, and year (e.g., 21 June 2010), while in American English, the month precedes the date (June 21, 2010).

Style to be followed - e.g., AMA 10th guidelines/11th guidelines, APA style, Chicago style. These differ a lot in terms of formatting.

Page size, margins, spacing, font: What page size to use (e.g., A4 or letter), how much margin is required on all 4 sides, whether the manuscript should be single-spaced, 1.5-spaced, or double-spaced. What font and what font size is to be used.

Indentation, hyphens, text alignment: Whether a new paragraph should be indented, whether hyphens are allowed, whether the text is to left-aligned or justified.

Spacing: spaces before and after symbols e.g., ' = ' or '='. Even if this is not specified, consistency is expected. Any one style is to be consistently followed for all symbols throughout the manuscript.

Page and line numbers: Whether to number pages, if yes, where? Are line numbers required? If yes, should they be continuous or restart on each page.

Writing numbers: Whether all numbers are to be written as numerals or those up to ten are to be spelled out. Also, how to write a range of numbers (whether to use a hyphen or en dash).

Which abbreviations are allowed/not allowed?

Word count limits- How many words are allowed in the title, abstract, main text, no. of references, tables, figures allowed.

Title page: How many words/characters are allowed in the title, are abbreviations allowed/not allowed? Is the first letter of every word to be capitalized or only the first word? Is a running title required- how many words and characters are allowed. How to write the author names- first name+ surname, initials+surname, surname+first name, surname+initials. Are the affiliations to be

indicated by superscript letters or numerals. What additional information is required on the title page -e.g., Word count, Conflict of Interest Disclosures etc.

Abstract and keywords: Whether it should be structured/unstructured. If structured, what should be titles of the subsections. What is the word count allowed. What details are mandatory. Keywords are usually to be chosen from the MeSH list.

Main text: What should be the titles of the sections? Should they be numbered? Should they be left-aligned or center-aligned? What font and formatting is required? Details of ethics approvals are always required in the Methods section. Whether subsections are allowed/required in the Methods and Results section and if yes, how many and should they be numbered? Whether a conclusion section is required/allowed. Is there a word limit for any of the sections?

Citations: There are very very specific guidelines for these - e.g., superscript, brackets, parentheses, first author name with the year, first two authors' names with the year. It also needs to be checked whether the punctuation marks (periods, commas, colons, semi-colons) are to be placed after the citation or before. Whether there should be a space before the citation.

Writing 'p' values: Some journals might have very specific guidelines for writing 'p' values. e.g. exact values might be required, a '0' before the decimal might not be allowed for values below a certain number, significant 'p' values might need to be written in bold or italics.

Formatting references- This is again very important and not even a single spacing error is acceptable. There are specific guidelines about how to write author names (first name+ surname, initials+surname, surname.first name, surname.initials.), how many authors names are allowed before et al., how to write the journal name (abbreviated, italics, bold, followed by period/comma or not), year (to be written after the authors' names or after the journal name, to be followed by comma or colon/semi-colon, followed by space or not), volume number (bold, plain, followed by comma or semi-colon, followed by space or not), page numbers (full form or shortened form, separated by hyphen or en dash), doi required or not.

Formatting tables: Where to place the tables, the design/layout of the table, how to indicate footnotes to the table (there is often a risk of symbols to be followed in the given order to indicate the footnotes).

Formatting figures: Where to place the figures, what format is allowed, what resolution is required. In fact, there are very detailed guidelines for figures which vary from journal to journal, for each type of figure e.g., charts, colored images, histological images etc.

Disclosures: This section is usually required and includes conflict of interest, acknowledgments, funding, etc.

Additional sections: Some journals require additional sections -e.g., what is already known, what does this study add to the existing knowledge etc. which have limits as low as 85 characters including spaces.

So, I have covered most types of journal guidelines. In case I come across some more, I will add them here. Note, however, not all journals have all the above guidelines, but most leading journals do. If these seem intimidating, seeking professional help for writing the manuscript can be considered.

Before I sign off, I hope this series will be of help in planning your research and achieving a successful publication. Please do not hesitate to write to me in case you have any queries.

Best wishes for your publication!!!