GENERATING WORKING HYPOTHESES FOR ORIGINAL RESEARCH STUDIES

Received: Sept 3, 2019

Accepted: Sept 5, 2019

Durga Prasanna Misra^{1*} <http://orcid.org/0000-0002-5035-7396>

Vikas Agarwal¹ <http://orcid.org/0000-0002-4508-1233>

¹Department of Clinical Immunology and Rheumatology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India

*Corresponding author:

Durga Prasanna Misra, MD, Assistant Professor, Department of Clinical Immunology and Rheumatology, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow-226014, India;

Twitter handle: @DurgaPrasannaM1; **E-mail:** durgapmisra@gmail.com

Abstract

A hypothesis is a statement of the expected outcome of a research study, generally based on analysis of prior published knowledge, or with reference to the previous work of the investigators. The hypothesis forms the foundation of a research proposal. A study based, and planned, on a sound hypothesis may have a greater likelihood of meaningfully contributing to science. After the generation of a hypothesis, it is equally important to appropriately design and adequately power a study (by ensuring a sufficient sample size) in order to test the hypothesis. Adhering to principles discussed forthwith shall help young researchers to generate and test their own hypotheses, and these are best learnt with experience.

Keywords: Hypothesis, Research design, Statistics as topic, Research ethics

How to cite: Misra DP., Agarwal V. Generating working hypotheses for original research studies. Cent Asian J Med Hypotheses Ethics 2020;1(1):14-19. <https://doi.org/10.47316/cajmhe.2020.1.1.02>

SCIENCE – THE SEARCH FOR THE TRUTH

Scientific knowledge is built upon by engaging in research about natural phenomena[1]. Such research could simply observe and report natural phenomena (observational research), or could attempt to modulate natural phenomena based on knowledge derived from these observations (interventional research) [2]. A hypothesis forms the

cornerstone of the research roadmap for a particular research study. In this article, we shall discuss, about the relevance and testing of a hypothesis, from the viewpoint of medical research, with examples.

HYPOTHESIS-FREE AND HYPOTHESIS-BASED RESEARCH

Hypothesis refers to a statement, from the investigators, regarding the likely outcome of a research study [3]. Certain types of studies may be free from pre-stated hypotheses. Such exploratory studies help gather background information regarding a particular disease or patient population, when such information is not otherwise available. For example, population-based surveys to detect the incidence or prevalence of diseases in the community need not have a hypothesis. Similarly, studies exploring the patient characteristics of new diseases, or established diseases in populations where no prior data is available regarding this, may also be hypothesis-free. Such studies aim to collect background information to help understand disease characteristics. Yet another type of hypothesis-free research involves the use of novel techniques such as genomics, transcriptomics, proteomics, and metabolomics analysis [4], or attempts to derive novel information from big data in the real world using artificial intelligence [5].

A majority of studies conducted in the present-day world have a prior stated hypothesis. With respect to such studies, the investigators provide a statement regarding what they feel may be the likely outcome of the study. For example, in an observational study on the patient characteristics of rheumatoid arthritis from a state in India, the authors may propose that they perceive that rheumatoid arthritis (RA) in this region may affect younger individuals, and be more severe with respect to disease activity, presence of deformities, and extra-articular manifestations compared to other cohorts of RA from the Western world. Considering an example of an interventional study, where drug X is being compared against established drug Y for the management of RA, the investigators may propose a hypothesis that drug X is likely to be superior to drug Y in terms of attainment of remission of disease activity by 6 months. They may expect 60% patients on drug X to be in remission at 6 months, compared to 40% patients attaining remission at 6 months on drug Y. It is important to state a hypothesis a priori whenever possible, since the study should be designed thereafter to attempt to

prove, or negate, the hypothesis. Herein arises the concept of the null hypothesis, and the alternative hypothesis. When a difference is being proposed between the experimental and the control groups, the null hypothesis states that, in reality, there is no difference between the experimental and control groups. The alternative hypothesis opposes the null hypothesis, and instead states that there does exist a difference, in the real world, between the experimental and the control groups [6]. Generally, the results of the study shall prove either the null hypothesis or the alternative hypothesis to be true.

GENERATING HYPOTHESES

A hypothesis could be generated based on a thorough analysis of the published literature on the subject [3]. In this context, before embarking on a research study, it cannot be overemphasized regarding the need to conduct an extensive literature search through multiple bibliographic databases regarding the subject of the study [7]. Scientists planning to conduct a research study should preferably search through both multidisciplinary (MEDLINE, Scopus, Web of Science) as well as specialist bibliographic databases (those catering to specific groups of scientists, such as chemical sciences) [8]. This shall help identify the breadth of the existing knowledge on the subject, as well as to understand which areas require further exploration. Such gaps in the existing knowledge could be addressed by the planned research study. This approach shall also help avoid unnecessary duplication of research efforts, by attempting to address hypotheses which have already been definitively proven or disproven by previous studies. It must be noted that the literature search may yet identify prior hypotheses which require further exploration. For example, it may be still relevant to further study the role of ex vivo transfer of regulatory T lymphocytes in patients with refractory RA, even if there are prior studies in this regard, if the results of such previous studies are equivocal. For example, four studies might have shown a beneficial effect, whereas, three others failed to demonstrate any beneficial effect. In such a situation, it may be relevant to analyze the methodology of these studies to identify potential lacunae which might have resulted in the positive or

negative results, and address such lacunae while planning one's one research study to test the hypothesis regarding the efficacy of ex-vivo transfer of regulatory T lymphocytes in refractory RA.

Another approach to generate a hypothesis could be based on one's own research. The prior work of the researchers might have led on to certain observations, which could form the basis for future research studies. We cite an example from our own work to demonstrate this. Our group conducted a randomized, placebo-controlled, cross-over trial to evaluate the clinical efficacy of the phosphodiesterase 5 inhibitor tadalafil in patients with refractory Raynaud's phenomenon (RP - a condition resulting in excessive vascular spasm and compromise of circulation in the extremities of the body). The hypothesis behind this study was a result of our analysis of the literature, which suggested the role of the short-acting phosphodiesterase 5 inhibitor sildenafil in patients with RP. We proposed that tadalafil, which has a longer duration of action (hence, the advantage of alternate day dosing, instead of the thrice daily dosing of sildenafil), with a similar mechanism of action, should be similarly effective in ameliorating RP. This trial revealed that tadalafil was effective in reducing the frequency and severity of RP, proving the initial study hypothesis [9]. Interestingly, we also observed that in patients with scleroderma (a disease associated with excessive, unregulated fibrosis in skin and other internal organs), which formed most of the study population, there was an improvement in clinically apparent skin fibrosis (although this was not the target of our study). Based on this observation, we hypothesized that tadalafil may have effects beyond just vasodilatation (leading on to improvement of RP), and may also affect the primary pathological process of fibrosis which results in scleroderma. We further designed two studies to test this hypothesis that tadalafil was an antifibrotic agent. We assessed skin fibrosis scores in patients with scleroderma before and after tadalafil, compared to a group of patients not treated with tadalafil, as well as effects

on the transcription of genes related to fibrotic pathways as demonstrated in skin biopsies from these patients before and after tadalafil. Our study demonstrated a beneficial effect of tadalafil on skin fibrosis, clinically as well as based on analysis of transcription of pro- and anti-fibrotic genes in the skin [10]. Further, we designed another study to culture skin fibroblasts from patients with scleroderma *in vitro*, and could demonstrate favorable modulation of the pro-fibrotic potential of these fibroblasts with tadalafil and other phosphodiesterase 5 inhibitors *in vitro* [11].

A third way by which a hypothesis is generated could be by means of random thought processes. Such hypotheses may not be based on biological plausibility, or on a thorough literature search. While it must be understood that hypotheses are equally likely to be proven or not by subsequent experiments, generally, those studies based on random hypotheses without a solid background are neither encouraged, nor acceptable, by the scientific community. Young researchers should refrain from generating and testing such random hypotheses.

ETHICAL CONSIDERATIONS

The most important ethical consideration while generating a hypothesis is that the hypothesis should be the original idea of the authors, not taken from elsewhere without permission of the individual (s) who first generated it, thereby maintaining the necessary intellectual property rights [12]. In the latter scenario, the authors feel that it would be ethically appropriate to include the original person who conceived the idea in the planning and the conduct of the study to answer the research hypothesis. From an editorial viewpoint, researchers and editors must never use the hypotheses in their own studies taken from papers that they handle or review, until such studies have been published in the public domain [13].

Another significant ethical concern relates to studies that lack a prior hypothesis, or those studies with prior hypotheses which collect data regarding a

number of variables. There might be a temptation to seek “significant” p values, also referred to as data dredging or p hacking. Such an approach may reveal “significant” associations, often with little biological meaning, and should be avoided [14].

TESTING HYPOTHESES

The next step after generating a hypothesis is to test this out, by conducting a research study. Here, certain considerations have to be kept in mind. First, any hypothesis can only be tested in a particular population (the target population), and that too in a limited sample drawn from the target population (the study population) [15]. It remains possible that observations made on the study population to prove or disprove the study hypothesis may, or may not, represent the truth with respect to the study population.

Herein arises the concept of the power of a study, which is the ability to detect a true difference between groups, if such a difference truly exists. The greater the sample size of the study population, the more likely the results of the study are to approximate the real truth with respect to the target population. It is also important to understand the concept of significance level, generally taken by convention to be 5% (although this could vary depending on the type of study). This means that we are willing to accept a 5% probability that the null hypothesis for a particular study has been rejected by chance. Considerations of feasibility prevent the entire target population to be included in the study population, therefore, it is necessary to calculate a sample size, based on the hypothesized difference. This requires estimates of the mean values of a particular parameter in the study and control groups, and their standard deviations or standard errors, along with proposed limits of statistical significance and power. These are generally derived from prior published literature, or from pilot studies conducted by the investigators. Unless a study has a sufficient sample size, it is likely that it may not be able to appropriately test a hypothesis to approximate the true results in the target population [16].

Sometimes, studies may test more than one hypothesis. In this case, it is advisable to calculate sample size based on each of these hypotheses, and take into account the largest sample size arrived at from these calculations. It is considered good practice not to test too many hypotheses together in a single study. Also, while calculating sample size, one must take into consideration likely proportions of individuals who may not agree to participate in the study (non-response rates), or who

may drop out during the follow-up period. The final sample size calculated must be inflated to account for such expected non-response or losses to follow-up [17]. Figure 1 summarizes the key considerations in generating a hypothesis.

Figure 1: Generating and testing hypotheses

STUDY HYPOTHESES AND FUNDING

Conducting most types of studies requires funding. Researchers have to prepare grant applications with full study proposals, which are submitted to funding agencies for their consideration [18]. Speaking from personal experience, since funding is limited most of the time, research studies with a sound hypothesis and appropriate study design are more likely to be funded than those which lack these attributes. Furthermore, with regards to medical research, studies whose hypotheses have a potential to translate into benefits for patients or healthcare are more likely to get funded when compared to hypotheses that lack direct translational relevance.

CONCLUSION

The key attributes of sound hypotheses are two-fold: thorough literature search and preliminary work conducted by the investigators. It is equally important to appropriately design a study in order to adequately address the study hypothesis. Ethical considerations in the generation of a hypothesis are originality, with due attention to intellectual property rights, and the avoidance of data dredging. Generating and testing hypotheses are not easy, and best learnt with experience. These considerations might be relevant to keep in mind when early career researchers attempt to generate and test hypotheses.

FUNDING

None

AUTHOR CONTRIBUTIONS

Substantial contributions to the conception or design of the work; and the acquisition, analysis, or interpretation of data for the work – DPM, VA. Drafting the work – DPM. Revising it critically for important intellectual content – VA. Final approval of the version to be published - DPM, VA. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved – DPM, VA.

CONFLICTS OF INTEREST

The authors have no potential conflicts of interest to disclose.

DISCLAIMER

No part of the manuscript has been submitted simultaneously or published elsewhere.

REFERENCES

1. Maqbool F, Bahadar H, Abdollahi M. Science for the benefits of all: The way from idea to product. J Med Hypotheses Ideas 2014;8(2):74–77.
2. Thiese MS. Observational and interventional study design types; an overview. Biochem Med 2014;24(2):199–210.
3. Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. Can J Hosp Pharm 2014;67(1):31–34.
4. Leon-Mimila P, Wang J, Huertas-Vazquez A. Relevance of Multi-Omics Studies in Cardiovascular Diseases. Front Cardiovasc Med 2019;6:91.
5. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019;380(14):1347–1358.
6. Leppink J, O'Sullivan P, Winston K. Evidence against vs. in favour of a null hypothesis. Perspect Med Educ 2017;6(2):115–118.
7. Gasparyan AY, Ayvazyan L, Blackmore H, Kitas GD. Writing a narrative biomedical review: Considerations for authors, peer reviewers, and editors. Rheumatol Int 2011;31(11):1409–1417.
8. Gasparyan AY, Yessirkepov M, Voronov AA, et al. Specialist bibliographic databases. J Korean Med Sci 2016;31(5):660–673.
9. Shenoy PD, Kumar S, Jha LK, et al. Efficacy of tadalafil in secondary Raynaud's phenomenon resistant to vasodilator therapy: a double-blind randomized cross-over trial. Rheumatology (Oxford) 2010;49(12):2420–2428.

10. Ahmed S, Rai MK, Misra DP, Agarwal V. Tadalafil reduces skin fibrosis and profibrotic genes expression in patients with systemic sclerosis (abstract). *Arthritis Rheumatol* 2017;69 (suppl 10).
11. Agarwal V, Rai MK, Agarwal V, Singh H, Chaturvedi S. Phosphodiesterase-5 Inhibitors Attenuate Fibrotic Phenotype and Restore Anti-Fibrotic Responses of Cutaneous Fibroblasts in Patients with Scleroderma [abstract]. *Arthritis Rheumatol* 2016;68 (suppl 10).
12. Jolly M, Fletcher AC, Bourne PE. Ten Simple Rules to Protect Your Intellectual Property. *PLOS Comput Biol* 2012;8(11):e1002766.
13. ICMJE recommendations. Available from: <http://www.icmje.org/icmje-recommendations.pdf> [Accessed Aug 11, 2019].
14. Marín-Franch I. Publication bias and the chase for statistical significance. *J Optim* 2018;11(2):67–68.
15. Banerjee A, Chaudhury S. Statistics without tears: Populations and samples. *Ind Psychiatry J* 2010;19(1):60–65.
16. Akobeng AK. Understanding type I and type II errors, statistical power and sample size. *Acta Paediatr* 2016;105(6):605–609.
17. Whitley E, Ball J. Statistics review 4: sample size calculations. *Crit Care* 2002;6(4):335–341.
18. Kanji S. Turning Your Research Idea into a Proposal Worth Funding. *Can J Hosp Pharm* 2015;68(6):458–464.

Шынайы зерттеулер үшін жұмыс гипотезаларын қалыптастыру

Түйіндеме

Гипотеза - бұл бұрын алынған білімді талдау немесе зерттеушілердің алдыңғы жұмысына негізделген ғылыми зерттеудің күтілетін нәтижесінің тұжырымы. Гипотеза зерттеу ұсынысының негізінде жатады. Дұрыс гипотезаға негізделген және жоспарланған зерттеу жұмысы ғылымға қосқан үлесінің жоғары ықтималдылығымен сипатталады. Гипотезаны ойластырғаннан кейін, зерттеуді дұрыс жобалау және жеткілікті түрде жүргізу (таңдаудың жеткілікті мөлшерін қамтамасыз ету) бірдей маңызды болып табылады. Бұл гипотезаны тексеру үшін қажет. Осы жұмыста қарастырылған қағидаларды ұстану жас зерттеушілерге өздерінің гипотезаларын құруға және тексеруге көмектеседі.

Түйін сөздер: Гипотеза, Оқу дизайны, Тақырып ретінде статистика, Зерттеу этикасы

Дәйексөз үшін: Мисра Д. П., Агарвал В. Шынайы зерттеулер үшін жұмыс гипотезаларын қалыптастыру. Медициналық гипотеза мен этиканың Орта Азиялық журналы. – 2020. – №1 (1). – Б. 14-19. <https://doi.org/10.47316/cajmhe.2020.1.1.02>

Разработка рабочих гипотез для оригинальных научных исследований

Резюме

Гипотеза – это сформулированный ожидаемый результат научного исследования, базирующийся на анализе имеющихся знаний или на ранее проведенных исследованиях. На основании гипотезы разрабатывается план исследования. Чем правильнее составлена рабочая гипотеза, тем выше ценность исследования. После разработки гипотезы, правильное планирование исследования так же важно, как его корректное выполнение (с обеспечением достаточного размера выборки) для проверки и тестирования гипотезы. Принципы, которые изложены в статье, помогут молодым исследователям правильно составлять и проверять свои гипотезы.

Ключевые слова: Гипотеза, Дизайн исследования, Статистика как тема, Этика исследования

Для цитирования: Мисра Д. П., Агарвал В. Разработка рабочих гипотез для оригинальных исследований. Центральноеазиатский журнал медицинских гипотез и этики. – 2020. – №1(1). – С. 14 -19. <https://doi.org/10.47316/cajmhe.2020.1.1.02>