Aleatory and epistemic uncertainties can completely derail medical research results

Indrayan A

ABSTRACT

Aleatory uncertainties are generated by intrinsic factors such as studying a sample rather than the whole population and the source of epistemic uncertainties is extraneous such as limitations of knowledge. These uncertainties inflict all the findings in empirical medical research, but they are rarely appreciated. This article highlights these uncertainties and shows with the help of an example how apparently valid and reliable findings can completely derail due to these uncertainties. We conclude that aleatory and epistemic uncertainties should get due consideration while drawing conclusions and before the results are put into practice. Methods to reduce their impact on results are also presented.

KEYWORDS: Aleatory uncertainties, epistemic uncertainties, non-reproducible findings, wastage in research

Introduction

Despite enormous progress in medicine in recent times, the reports of non-reproducible findings[1] keep coming up. Thus, legitimate questions are raised on the actual worth of enormous resources invested in medical research around the world.

Most of the research in health and medicine is empirical where a set of data is collected on the patients or healthy subjects and analyzed with the help of statistical methods to come up to a result. The expectation is that such results are reliable and valid, and can be used to advance the science of medicine when accompanied by rational thinking. As many as 761,674 citations appeared in MedLine database with 2017 as the year of publication[2] but perhaps only a small fraction of them succeeded in advancing the practice of medicine. Chalmers[3] and Yordanov et al. [4] have highlighted this enormous waste in medical research. Among several reasons, such as choice of topic, design, sample size, confounding, and bias and errors, one important reason for this waste is the failure to take proper account of the aleatory and epistemic uncertainties. In medical research, these uncertainties are common and afflict the results in unpredictable ways. The conclusion, which is supposed to consider not just the result but also factors such as biological plausibility, previous knowledge, and corroborative evidence, can also go haywire. The net consequence is that many results lack reproducibility and are unusable. This communication first explains aleatory and epistemic uncertainties for the benefit of those who are not aware and then illustrates them with the help of an example. The example shows how apparently valid result loses credibility because of these uncertainties. In the end, we suggest some methods to reduce the impact of these uncertainties on the results.

Aleatory and Epistemic Uncertainties

Aleatory uncertainties are intrinsic to the study and arise mostly due to sampling fluctuations. Results from one sample of patients generally differ from another sample from the same target population due to individual variation. The factors causing these variations can be categorized as biological (such as age, gender, hereditary, immunity level, physiological functions, and biochemical parameters), environmental (life style, stress and anxiety, climate, pollution exposure, infection, and such other factors), and social support (family ties, interaction with...
friends, financial security, etc.). In addition, imperfect tools also contribute to aleatory uncertainties. When my blood pressure differs from yours, it is not merely due to differences in age and gender but also due to the variation in several other associated factors. At the study level, results differ from one study to another due to differences in design, method of collection of data, variables under consideration, confounding, method of analysis, interpretations, and reporting. Most aleatory uncertainties can be minimized by controlling the individual factors and by being more careful at each step of the study. However, the effect of these uncertainties cannot be completely eliminated. Two identical studies in the same population can still lead to different results due to uncontrollable sampling fluctuations.

Epistemic uncertainties are more intricate and require deeper explanation. These were first highlighted in the context of medical research by Indrayan in 2008[9] but still not fully appreciated. These uncertainties arise mainly due to the limitation of our knowledge. According to one paradigm, what we do not know is much more than what we know. All the studies are necessarily based on existing knowledge and that can be very incomplete in some situations. For example, only those risk factors of cancer of the prostate are studied that can be conjectured, and these conjectures depend on our present knowledge. Nobody tries to find how echographic findings modify this risk because that is not in our conjecture yet. As the knowledge expands, more and more factors are added and they tend to get more exactitude. In science, the unknown domain contributes to what we generally call chance.

In the context of medicine, at least two types of limitations of knowledge can be identified. One is the global ignorance as mentioned in the previous paragraph and the other is incomplete knowledge at the personal level. There may be aspects about which a particular researcher does not know although that knowledge is available with others through literature, experience, discussion with colleagues, and such other sources. Sometimes the researcher knows but fails to consider while planning a study, either due to negligence and lack of resources, or lack of knowledge about how to take care of various factors. Global ignorance and personal inadequacy both result in unknown inaccuracies but the former affects at the macro level (such as for all studies on peritonitis) and second only a particular study. In research set up in an institution, these two may be confounded and can hardly be segregated once the study is completed.

Another source of epistemic uncertainty is incomplete information on the patients under research. If it is on patients, some of whom come in coma, nothing much can be elicited and the patients’ management has to start based on whatever can be observed. Many patients, who are in full senses, are not able to provide complete history as they forget and do not have records. Some even suppress the information as for an injury or for drug abuse to avoid hassles. Such instances also put in a question mark on the results of some research.

All these uncertainties have the potential to adversely affect the credibility of results. We illustrate this with the help of a fictitious but simple example.

Example: Estimating the Incidence of Sexual Adverse Effects of Finasteride in a Population

Finasteride is commonly prescribed for male-pattern hair loss but is known to cause side effects, particularly adverse effect on sexual functions such as decreased libido and erectile dysfunction.[6] The incidence of such side effects is low and varies from population to population because of its possible dependence on factors such as age distribution of the patients, their pre-existing physical health condition, and the prevalence of anxiety-stress syndrome. Such risk factors seem to have not been studied for side effects of finasteride, but we consider them for this made-up example for illustration. An estimate of the incidence of such side effects is important for discussing this with patients while prescribing the drug.

Suppose a study was carried out on a group of 1,000 consecutive male patients who were prescribed finasteride orally 1 mg daily for 12 months. They reported that they never took finasteride earlier and agreed to participate in the study. They answered questions on their background information such as age, physical health, and anxiety-stress syndrome at the time of recruitment. No question was asked about sexual functions at this time as that could have alerted them about such side effects. At the end of the 12-month period, 800 of them could be assessed for their sexual functions – others dropped out. Only questions regarding libido and erection relative to their initial status were asked for a focused response. These are the only side effects under this study. In total, 5.1% reported such adverse effects. This is the point estimate. No other information was recorded. What kind of uncertainties does this express for the estimate of the incidence of sexual adverse effects of finasteride?

The most obvious source of aleatory uncertainty is the sampling fluctuations. Another group of 800 patients from the same population may reveal sexual adverse effects in 4.1% or 5.2% or any other. If the sample is simple random from a specified target population, a statistical confidence interval (CI) can be easily built around the point estimate. The convention is 95% CI. In this example, this would be 3.6%–6.6% as per the established procedure. Note that this is quite wide despite a large n and shows the high vulnerability of the estimate. Another problem is that the CI is based on simple random sample, which is not the case in this example. However, consecutive cases in this example can be considered to simulate simple random sampling and no adjustment is needed. But the requirement of consent has the potential to make it biased as the consent is affected by factors such as knowledge, anxiety, and satisfaction with the consent process.[7] This can have a major impact but let us assume for this example that it is minor and the plausible range is marginally higher to 3.5%–6.7% instead of the CI of 3.6%–6.6%. Specific values of the wider range are only guesses just to illustrate the point about the effect of unaccounted uncertainties.

The major problem, however, is nonresponse by 20% of the subjects in this example. This is substantial and can severely affect the estimate depending on whether none or many of them had the side effects under consideration. The nonresponders are
generally those who belong to low socioeconomic status, high illness burden, etc.\(^8\) and adds to the uncertainty to the estimate of the incidence of side effects in this case. The reasons for nonresponse are generally not known, hard to elicit, and come mostly under the epistemic domain. Let the new plausible range incorporating this uncertainty be 3.3%–6.9%.

Now consider the fact that these 800 patients were those who were prescribed finasteride, but their actual intake is not known. Some may have missed intermittently for a short period each time and some may have discontinued for a long period after observing side effects. Suppose this was asked at the end of the 12-month period and all reported regular intake, where regularity is defined as at least 90% intake. We assume for our example that there is no misreporting. This may not be so in other setups.

The antecedents under this study are age, physical health, and anxiety-stress syndrome. Other antecedents such as hand preference and sexual orientation\(^9\) are excluded. We will come to their measurement in a while but realize for now that the incidence of side effects should be different in different age-groups, and similarly for different levels of physical health and anxiety-stress levels. The point estimate of 5.1% obtained in this study is the average over the variation in these antecedents in the study group. The same is true for our plausible range of 3.3%–6.9% stated above. These should have been calculated separately for younger subjects and older subjects, and similarly for different physical health groups and anxiety-stress groups. If a patient of age 29 years in good physical health and no anxiety comes for a consultation, he cannot be told that the incidence of side effects is between 3.3% and 6.9% for his kind of patients. It is likely to be much lower for such patients and much higher for patients of age, say, 50 years and in debilitating condition. Considering this variation in various subgroups, it would be wise to consider a much wider plausible range such as 3.0%–7.2% in place of 3.3%–6.9% arrived earlier. There is no way to calculate this exactly but the plausible range illustrates the point that the CI widens. Note how the original CI for average incidence loses reliability due to such aleatory uncertainties. We have not considered one more aspect of epistemic uncertainties as stated in the preceding paragraphs. These uncertainties will ultimately reflect on the plausible interval of the estimate of the incidence of side effects. Suppose the new interval is 2.8%–7.4% against 3.0–7.2 reached earlier. These are conservative estimates – if we actually calculate, the interval may be wider.

On the outcome side, this example restricts to decreased libido and erectile dysfunction. There is an epistemic gap regarding the method to measure them exactly as no widely acceptable scale is available. The impact of this uncertainty on the incidence of side effects is also largely unpredictable but it is easy to imagine that this would add to the uncertainty in the estimate. For illustration, let us say that this uncertainty increases the plausible interval from 2.8%–7.4% to 2.6%–7.6%. Note also that the side effects in this study were assessed by interview method where the response would depend on the perception of the responder. This will not contribute to uncertainty if the study aims to assess this perception and is explicitly stated.

The bigger problem, however, is that the side effects studied in this study are only decreased libido and erectile dysfunction. Even if other side effects such as depression, rash and gingival hypertrophy\(^1\) are excluded, other sexual adverse effects such as lower sperm count and trouble having orgasm can also occur. There might be others in the epistemic domain about which we do not know yet. These will increase the incidence and not decrease it if all sexual adverse effects are to be considered.

Suppose the new range to account for this, raises the upper limit from 7.6% to 8.0%. The lower limit remains at 2.6% as guessed earlier. This adjustment is required only if the patient is to be informed of regarding incidence of all the sexual adverse effects and not restricted to the two considered in this study.

In this example, the actual 95% confidence interval (CI) was 3.6%–6.6%. This is what is generally reported. But the plausible range as just explained could be 2.6%–8.0%. This is the conservative range considering that we have ignored the effect of certain uncertainties as stated in the preceding paragraphs. It may be alleged that we have inflated the range too much but our experience suggests that we are conservative for the effect of the specified uncertainties. The CI is based on widely accepted statistical methods but further widening is our conjecture with hardly any scientific basis. Perhaps, there is no way to delineate this exactly and that is not important either. The explanation of the aleatory and epistemic uncertainties in this setup should leave no doubt that the actual plausible limits are much wider than the CI. Thus, the point estimate of 5.1% and 95% CI of 3.6%–6.7% can be hardly believed despite the study following all the steps for an adequate prevalence study in this example.

We have not considered one more aspect of epistemic uncertainty and not many seem to be alive to this problem. Almost any result in medical research is derived from the existing or past cases but is generally used on future cases. No future case can be in any sample – thus, even the most immaculately drawn sample remains imperfect representation. It is presumed that future cases, at least in immediate future, would be similar.
to those studied. The experience suggests that this works well
in most situations but may not hold in some situations such as
when some new development occurs, say, for the assessment
of sexual adverse effects. The other aspect is using the result for
one population on the other without making any adjustment.
For example, the normal body temperature 98.6°F was obtained
for the German population. This is accepted almost all
over the world but no large scale study has been carried out
in developing countries to confirm that this is valid for them.
The same can be stated also for a large number of laboratory
based parameters.

Our example illustrates the sizeable impact of aleatory and
epistemic uncertainties in the simple case of estimating the
incidence of side effects. The objective in most medical studies
is more complex such as finding the relative importance of
various risk factors. In such studies, the role of covariates and
confounding factors can be prominent, and the aleatory and
epistemic uncertainties in the assessment of these factors can
make a substantial dent on the validity and reliability of
the results. At present, these uncertainties are mostly ignored and
the results become unreproducible.

Minimizing the Impact of Uncertainties

The realization of the presence of aleatory and epistemic
uncertainties in an empirical result is an achievement by itself
because these are rarely discussed in the context of medical
research. The steps to control these uncertainties can be taken
only after such a realization.

Aleatory and epistemic uncertainties remain but their impact
on results can be controlled by taking certain steps. Some
of these are already well known and many studies routinely
adopt them, although without proper contextualization. For
example, the proper choice of design can control most aleatory
uncertainties caused by variation in antecedent factors, their
interaction, and the confounders. Strict inclusion and exclusion
criteria help to remove many sources of aleatory
uncertainties and to focus on a specific type of subjects but the
generalizability suffers in this process. The results should have
explicit mention of this limitation and no tall claims should be
made for the general class of patients. Random selection and
randomization in clinical trials is done to ameliorate the effect
of unknown factors in the epistemic domain in the hope that
these will average out. These methods are effective in their
mandate in the case of large samples but no validated scoring system is available for many
other factors. Thus, the first step for reproducible research in
such cases should be to develop a scoring system and validate
it for the population under study. The procedure for this has
been described by several workers for different conditions.

We have earlier mentioned epistemic uncertainties in the
context of the exact measurement of the antecedents and
outcome in our example. There are a large number of medical
factors that defy direct measurement. The usual practice
for such factors is to devise a scoring system and use it after
proper validation. Many scoring systems such as quality of life
scores and APACHE scores have been validated in different
populations but no validated scoring system is available for many
other factors. Thus, the first step for reproducible research in
such cases should be to develop a scoring system and validate
it for the population under study. The procedure for this has
been described by several workers for different conditions.

For the epistemic uncertainties generated by global ignorance,
perhaps the only alternative is to expand the horizon and
conduct more focused research on possible conjectures.
Development and use of improved medical methods are always
helpful in furthering research. This is a long term process and
possibly no short-cuts are available. Till that time, we have to live
with it but we need to be cautious in our conclusions and accept
uncertainty. However, for incomplete medical knowledge of
individual workers, help can be taken from tools such as expert
systems. These are available for some conditions and more can
be developed. This is easily said than done because of problems
such as congregating a set of real experts for each topic. Such
experts are rare, and they are extremely busy in their profession
and generally not available for this kind of exercise. The other
such tool is the etiology diagram of the type presented by
Indrayan and Malhotra for myocardial infarction. This
diagram helps in not missing out factors of importance in
its causation. More such diagrams can be developed to facilitate
general practitioners and researchers.

The steps suggested in this communication may make it difficult
to reach a conclusion but that is the price we should be willing
to pay for valid and reliable research results.
Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References

1. Ioannidis JP. Why most published research findings are false? PLoS Med 2005;2:e124.
2. MedLine. MEDLINE® Citation Counts by Year of Publication (as of January 2019). Available from: https://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html.
3. Chalmers I. Avoidable waste in the production and reporting of research evidence. Lancet 374:86-9.
4. Yordanov Y, Dechartres A, Atal I, Tran VT, Boutron I, Crequit P, et al. Avoidable waste of research related to outcome planning and reporting in clinical trials. BMC Med 2018;16:87.
5. Indrayan A. Medical Biostatistics. 2nd ed. Chapman & Hall/CRC Press; 2008. p. 8.
6. Mayo Clinic. Finasteride (Oral Route). Available from: https://www.mayoclinic.org/drugs-supplements/finasteride-oral-route/side-effects/drg-20063819/p=1. [Last accessed on 2020 Jan 14].
7. Kinnersley P, Phillips K, Savage K, Kelly MJ, Farrell E, Morgan B, et al. Interventions to promote informed consent for patients undergoing surgical and other invasive healthcare procedures. Cochrane Database Sys Rev 2013;6:CD009445.
8. Dad T, Tighiouart H, Fenton JJ, Lacson E Jr, Meyer KB, Miskulin DC, et al. Evaluation of non-response to the in-center hemodialysis consumer assessment of healthcare providers and systems (ICH CAHPS) survey. BMC Health Ser Res 2018;18:790. Article number 790.
9. Rezende HD, Dias MF, Trueb RM. A comment on the post-finasteride syndrome. Int J Trichol 2018;10:255-61.
10. Toombs M, Nasir B, Kisely S, Rammuthugala G, Gill NS, Beccaria G, et al. Cultural validation of the structured clinical interview for diagnostic and statistical manual of mental disorders in Indigenous Australians. Australas Psychiatry 2019;27:362-5.
11. Motofei IG, Rowland DL, Georgescu S, Tampa M, Paunica S, Constantin VD, et al. Post-finasteride adverse effects in male androgenic alopecia: A case report of vitiligo. Skin Pharmacol Physiol 2017;30:42-5.
12. Wikipedia. Human body temperature. Available from: https://en.wikipedia.org/wiki/Human_body_temperature. [Last accessed on 2019 18 Sep].
13. Indrayan A, Malhotra RK. Medical Biostatistics. 4th ed. Chapman and Hall/CRC Press; 2018. p. 38, 93, 122, 297.
14. WHO. The Use of the WHO-UMC System for Standardised Case Causality Assessment. Available from: https://www.who.int/medicines/areas/quality_safety/safety_efficacy/WHOcausality_assessment.pdf. [Last accessed on 2020 Jan 14].
15. Wand H, Iversen J, Wilson D, Topp L, Maher L. Developing and validating a scoring tool for identifying people who inject drugs at increased risk of hepatitis C virus infection. BMJ Open 2012;2:e000387.
16. Tao S, Hoffmeister M, Brenner H. Development and validation of a scoring system to identify individuals at high risk for advanced colorectal neoplasms who should undergo colonoscopy screening. Clin Gastroenterol Hepatol 2014;12:478-85.
17. Lee YH, Bang H, Park YM, Bae JC, Lee BW, Kang ES, et al. Non-laboratory-based self-assessment screening score for non-alcoholic fatty liver disease: Development, validation and comparison with other scores. PLoS One 2014;9:e107584.
18. Zhang Z, Zhang H, Khanal MK. Development of scoring system for risk stratification in clinical medicine: A step-by-step tutorial. Ann Transl Med 2017;5:436.
19. Wasserstein RL, Schirm A, Lazar NA. Moving to a world beyond “P<0.05”. Am Stat 2019;73(Suppl 1):1-19.
20. Wen LY, Howard SK. Value of expert systems, quick reference guides and other cognitive aids. Curr Opin Anaesthesiol 2014;27:643-8.