Quantitative Methods

Module 6: Practice, Ethics & Integrity

6.01 Practice, Ethics & Integrity: Documentation

In this video I'll discuss the relevance of research documentation. Good documentation is critical for two things: objective replication and checking and verification of results and conclusions. In other words documentation is very important to research integrity.

In order to replicate a study, the hypothesis, research design and predictions need to be clearly stated. Replication also requires a clear description of the procedures and instruments and the protocol for contacting and interacting with participants. Finally, the data, data manipulation and statistics that were performed and the researchers interpretation of these results into a final conclusion also need to be documented. If any of these steps are unclear or cannot be verified after the study is performed, then this can lead to confusion when contradictory results are found in replication studies.

The hypothesis, **research design**, specific predictions and interpretation of results are generally explicitly stated in a research publication or journal article. Such a document also contains information on the research procedure, instruments, data, data manipulations and statistics, but this information is always *summarized*.
Giving all the details simply takes up too much space. Most researchers just want to know a study's goal, setup and outcome. However, the detailed information should be documented somewhere, so it can be made available to those researchers who *are* interested in performing an exact replication.

So what details should be documented exactly? Well to start, all information on the **instruments** and **materials** need to be available. With instrument information I mean the instruments used to measure or manipulate the variables of interest and the procedures or instruments used to control or measure extraneous variables.
The materials include written or verbal instructions, texts used in email communications, consent forms, documents containing debriefing information, etcetera, etcetera. Basically all materials, information and documents that were used by the experimenter or participant need to be retrievable.

Let's focus on the instruments for a moment. Unlike tape measures and weight scales, social and psychological measures are usually not self-evident. In research articles measurement information is summarized by stating of how many questions a scale in a questionnaire consists, what the response options are, what the range of possible scale scores is. An example item is often provided.
This is the minimum required information to interpret the results, but a full version of the instrument needs to available for checking and replication. The instrument itself, meaning its observation coding scheme or questionnaire items and response options should be available to others.

This requires for example recording what version of a standardized test was used, or documenting and providing the instrument itself, if it's not publicly or commercially available.

Besides information about instruments and materials the **research protocol** needs to be documented. The research protocol refers to the order and manner in which materials and information were presented to participants and the way procedures were implemented. A research protocol is a clear description of what happens to the participant from start to finish. This includes the moment of first contact, how a participant is recruited, with what information and by whom. The protocol also describes how consent is obtained, what instruction or information is given in what order, what help is provided if participants don't understand instructions or they behave in an unexpected manner. The research protocol also describes how participants are debriefed about the research after their participation.

Documenting materials is relatively easy, it simply means saving emails and documents. Writing a research protocol is more tedious work and is not always performed diligently. Of course this can become a big problem when other researchers want to replicate a study and can only use summarized information or have to rely on the original researcher's memory!

Another thing that is often badly documented or not even explicitly formulated at all is the **statistical plan**. Ideally a researcher should determine what statistical analysis will be performed *before* looking at the data. This is because the choice of statistics can sometimes determine whether the predictions are confirmed or disconfirmed.
Ideally the hypothesis and intended statistical analyses are **preregistered**, so that a researcher is not tempted to change either after seeing the results and realizing that a different approach would result in more favorable outcomes. Preregistration is customary in medical sciences and is gaining popularity in neuroscience, clinical psychology and other fields.

Finally, information about pilot studies often fails to be documented. A pilot study is a small, preliminary study where a newly developed instrument or manipulation procedure is tested. Such studies can contain useful information for other researchers developing or adapting similar instruments. A pilot study should therefore at least be mentioned in the published article, so that others are aware that the results are available.

6.02 Practice, Ethics & Integrity: Data management

After the data are collected, a researcher's responsibility to be transparent and to document actions and decisions does not stop. After data collection, data need to be stored, processed and analyzed using statistics. Together this is referred to as **data management**.

Once the data are collected, they need to be recorded. Measurements are stored in a computer file, usually in a data matrix where columns represent variables and rows represent participants. Such a data file is useless if it's unclear what the recorded data represent. If we have a column representing sex for example, but we don't know whether a zero means female or male, the information is useless.

Information or meta-data about what the variables mean, is stored in a **codebook**. The codebook specifies what property each variable measures and what the values mean, what the range of possible values is and what values are used to denote a participant did not supply relevant information, referred to as missing values.

Suppose we collect responses to ten items forming a depression questionnaire with three answer options. To correctly interpret these scores we need to know that for each item the minimum possible value is one, the maximum value is three and we denote a missing response with the value nine. Without this information we would not realize that something went wrong if we find a score of five. We could have made for example an error entering the raw data into the computer.

Because data entry errors are always made it is extremely important to always save the original data. With the original data, I mean the paper questionnaires filled out by participants, or the video material used for observational coding. Without the original data we cannot check whether the data were entered into the file correctly. When data are entered manually it's always a good idea to let someone else enter a random selection of the data again and check for consistency. If it turns out the original entry is inconsistent then all the data need to be entered again, but more carefully of course!

The original data and instrument information are also necessary to check the codebook. Sometimes a codebook can be confusing or seem to be wrong. For example when responses to an item are unexpectedly low, this could be a valid pattern, but it could also be an error in the codebook. It is possible for example the codebook wrongly indicates an item is positively worded, when it is in fact phrased negatively and therefor should be recoded.

The original data file is also very important and should be stored separately. With the original data file I mean the file that contains the raw data as they were entered originally, before any manipulation was performed. Data manipulation refers to computations such as recoding and computing aggregate scores like a sum score of depression. Another example is calculating age from date of birth.

Without the original data file, we cannot check whether we made any errors in manipulating the data. Suppose that, for a negatively worded depression item, I change the score of one to a score of three - and three to one - and then I accidently recode again changing threes back into ones and ones into threes. I end up with negatively scored items, without being aware of this, thinking they are scored positively. If I find unexpected results, I can simply check if I made a recoding error by comparing against the original data file, that's why it's important.

Not only should we record the original data and data file, we should also record any processing, selection or computations we perform. Otherwise we might not be able to reproduce processed data that are used to formulate the final conclusions.
For example when I select a subset of my sample, say only people who completed the depression questionnaire within a month, then my results might be different from results obtained from the entire sample. If I don't record my selection criteria, then in a year from now I will probably have forgotten the exact criteria and will not be able to reproduce my own results.

Both the processing of data, for example recoding and computing sum scores, selection of data and the statistical analyses are generally recorded in a syntax file. The syntax file is like a simple programming file that can be used to reproduce all the computations and statistical analyses at the push of a button. This is very useful for checking and replicating results, not just for other researchers but also for the original researcher.

6.03 Practice, Ethics & Integrity: Unethical studies

Research ethics do not just concern research integrity; they also concern ethics towards participants. To understand the current thinking on ethics it's important to be aware of some of the studies that led to the formation of the ethical guidelines that we use today.

In the first half of the twentieth century several highly unethical studies were performed. When these studies were addressed in the popular media, outrage about the unethical nature of these studies led to the formation of a special committee.
The committee's report, referred to as the Belmont report, formed the basis for our current ethical guidelines. I will discuss the study that gave direct rise to the Belmont report, referred to as the **Tuskegee syphilis study** and I'll discuss a related study, the **Guatemalan syphilis study**.

The aim of the ***Tuskegee study*** was to assess the natural progression - without treatment - of syphilis, a sexually transmitted infectious disease. Participants were six hundred black, African-American men from Tuskegee Alabama. Four hundred of the men had syphilis, two hundred did not. Participants were offered health care for minor illnesses, free meals on examination days and free burials if they consented to being autopsied.

Participants were given "medical treatment" for "bad blood", a local term for different illnesses. The treatments weren't treatments at all, they actually consisted of spinal taps performed without anesthetics. The four hundred syphilis patients were never told that they had syphilis.

The study started in nineteen thirty-two, when a treatment for syphilis was still being sought. The study consisted of periodic measurements performed over a period of forty years. Half way through the nineteen-forties, around the end of the second world war, it had become clear that penicillin was an effective treatment for syphilis. And at the start of the nineteen-fifties, penicillin was widely accepted as the preferred treatment.
But until the end of the study in 1972 none of the participants were informed of this fact and no actual treatment was ever provided. Participants that were diagnosed with syphilis elsewhere were actively denied treatment or lied to and treated with placebos. A large number of participants directly or indirectly died from syphilis. Many infected their spouses; several children were contaminated in the womb and were born with the disease.

Now the Tuskegee study was not secret or performed illicitly. The study was performed by the U.S. Public Health Service. It involved many different researchers who periodically reported procedures and results in scientific medical journals. Of course it is obvious that the Tuskegee study is extremely unethical. Participants were lied to and seriously and unnecessarily harmed. They were a vulnerable group, consisting of poor, often illiterate people facing difficult economic times. These vulnerabilities were exploited in a horrific fashion.

But if you think this is the most horrendous study ever performed in peacetime you would be wrong. Just after the Second World War, the U.S. public health service performed a study in ***Guatemala***. With the cooperation of Guatemalan health services officials they actively infected people with syphilis and gonorrhea without their consent and without treatment.
Subjects were prisoners, soldiers and mental patients, coerced into participation. It seems researchers were fully aware of the unethical nature of this experiment. They knew the study would not be accepted in the U.S. Also, at the time of the study the trials at Nuremberg were taking place. These trials concerned the experiments performed by the Nazis in the concentration camps and could not have escaped the researchers' attention.

6.04 Practice, Ethics & Integrity: Research ethics

The Tuskegee study I discussed earlier, led to the formation of ethical guidelines for reviewing research proposals. Institutional Review Boards or IRB's now assess and approve research proposals involving human participants. Three ethical principles are generally distinguished: **Respect**, **beneficence** and **justice**.

Respect refers to respect for the participant's ***autonomy***. The decision to participate in research should always be made by participants themselves and should be voluntary. **Voluntary consent** can be contrasted with **coercion**. Coercion can be subtle. It can consist of offering an extremely large financial reward for participation. If the financial gain is large, then for people who have very little money, it becomes almost impossible *not* to participate. The same applies if the benefits consist, for example, of access to an experimental medical treatment that offers hope to terminal cancer patients.
A very specific form of coercion happens in most universities that offer psychology programs. In many cases first year psychology students are required to participate in a certain number of experiments for course credit. This is presented as part of their training. Alternatives to participation are offered to students, but these generally consist of very unattractive writing assignments.

Ok, back to voluntary consent: A decision to voluntarily participate can only be made if all relevant information is available to the participant. A participant should not only give consent, this consent should be ***well-informed***. An informed consent form should always be provided and signed beforehand, informing participants about the nature of the study. Of course revealing the purpose of the study conflicts with the finding that participants can react differently to the experiment if they are aware of the purpose of the study. Often, some form of **deception** is necessary to control for reactivity and demand characteristics. A review board decides whether this deception is necessary and does not cross ethical boundaries.

There are different forms of deception. Deception can occur by ***omission***: The goal of the study is not stated or formulated in very general, vague terms. Deception can also be ***active***: a cover story is provided that is entirely different from the actual purpose of the study. Or participants are given ***false feedback***. For example, participants are provided with a bogus intelligence test and they are told that they scored extremely low. The purpose could be to temporarily lower the participants self-esteem: a manipulation to see how lowered self-esteem affects people's ability, for example, to negotiate a salary.
A dangerous consequence of providing such false feedback is what's known as a ***perseverance effect***. This means participants are still affected by the deception, even after they are debriefed and the deception is revealed and explained to them. This can happen because participants might believe the researcher is just lying about the deception to make them feel better about their low scores.

If deception is deemed necessary and not harmful, then a review board might approve an informed consent form using a cover story, combined with an extensive debriefing afterwards. In all studies, participants should be made aware that they can withdraw their consent at any time during, or right after a study, and ask for their data to be removed.

Ok, the second ethical principle is ***beneficence***. Beneficence means that ***participants should not be harmed***. This principle is not as simple as it sounds. Sometimes participation carries a risk of doing harm but also a potential for doing good. The cost should always be weighed against

potential benefits. This applies at the individual level, for example when a patient participates in a study on a new cancer treatment that is a potential cure, but also has severe side effects.
But the cost-benefit analysis also applies on a broader level, for example for all cancer patients or even society as a whole. The missed benefits of not doing a study, not learning about a new cure for cancer or the cause for a societal problem, should also be weighed.

A type of harm that is perhaps less obvious is the invasion of a participant's ***privacy***. Participants should know what their data will be used for and who will have access to them. Anonymity of data should only be promised if identifying information is deleted and not even the researcher can retrace the data back to the participant.
Otherwise a researcher should be clear about the confidentiality of the data: who will have access and what will it be used for? Issues concerning confidentiality and use of data are becoming more important as more and more information of our behavior is recorded automatically.

Finally, the third principle of **justice** means that the costs and benefits of research should be divided reasonably, fairly and equally over potential participants. Specific groups should not be given preferential treatment. And reversely, vulnerable groups should not be exploited, as was the case in the Tuskegee study.

6.05 Practice, Ethics & Integrity: Research integrity

Research integrity can become compromised due to several reasons. The most serious violations of research integrity are **fraud**, **plagiarism**, **conflicts of interest** and undue influence of the researcher's **personal values**. All these threats to research integrity concern an abandonment of the basic scientific principles of openness, transparency and critical and systematic empirical testing.

Fraud refers to cases where data were either fabricated or falsified to provide false support for the researchers' hypothesis. Fabrication means the data were never collected, but were made up. Falsification means existing data were illegitimately altered. When data are fabricated or falsified to support a researcher's claims, science is damaged in several ways.

First of all, the scientific literature is contaminated with false empirical data, this alone holds back scientific progress. Researcher might pursue a line of research that *seems* promising but is in fact baseless. Honest researchers see their projects fail to replicate promising results. They might incorrectly conclude this is due to their personal lack of research skills. The project might be abandoned without the failed replications ever coming to light.

Secondly, precious funding resources that could have been spent on valid avenues of research are spent on fraudulent research and related projects that build on it, but are in fact less promising than they seem.
Thirdly, once fraud is exposed, the reputation and credibility of the field, including the majority of researchers who do have integrity, is severely damaged. This can result in more difficulty to obtain funding in a general area where fraud has been exposed. This puts honest researchers who have the bad luck to be in a related field at a disadvantage.

Fraud cases are invariably accompanied by reluctance or unwillingness to share data and research information with others. Unless researchers are challenged or even required to be open and transparent, fraud can be tempting in an academic climate were publishing positive, confirmatory results is held in high regard.
Preregistration of the research hypothesis and design, and documentation of materials, data and data manipulation form a good way to discourage fraud. If researchers know their procedures, data and analyses can be checked at any time, the risk of fraud being exposed is much more evident. Unfortunately guidelines on how to document and preregister research proposals are not always implemented and if they are these guidelines vary greatly between, even within scientific fields.

Let's move on to ***Plagiarism.*** Plagiarism is a different violation of research integrity. Plagiarism means that a substantial scientific contribution is presented as one's own by copying original text, concepts or data of others without referring to the original source.

Besides the obvious infringement on someone else's intellectual property, plagiarism contaminates the scientific literature with redundant information. If a study is plagiarized and presented as a separate, independent study, this could create the impression that a finding is more robust than it really is. Unknowing readers might interpret the plagiarized study as a successful replication, when the two studies are in fact one and the same.
Of course plagiarism often takes the form of copying small elements of other people's work, not entire studies. This is still a problem because it prevents people from having access to relevant related information in the source document, because this document is not referred to.

The large pressure on researchers might be to blame for a relatively new type of plagiarism called ***self-plagiarism***. This might seem like a contradictio in terminis, how can you plagiarize yourself? Well presenting a substantial scientific contribution that was already published elsewhere as an original contribution contaminates the literature with redundant information and makes it harder to gather relevant information if the original source is not referred to. Also when the original contribution was made with the help of co-authors, self-plagiarism means that these co-authors are not credited for their original work.

A ***conflict of interest*** is a violation of research integrity that is most frequent in the medical sciences. Researchers are funded for example by pharmaceutical companies that have a huge interest in showing a drug is effective. Whether

consciously or unconsciously, researchers can be swayed to present results more favorably. This is in part because it is also in their best interest to show positive results. Conflicts of interests cannot always be avoided, but they should at least be explicitly stated in a publication, so readers can judge for themselves what the credibility of a study is.

A final outright violation of research integrity is formed by **undue influence of personal values**. Strong conviction or personal values can blind researcher to their data and valid critiques. If researchers do not adhere to the principle of objectivity and are unwilling to accept critique or discuss plausible counterarguments based on logic and empirical evidence, then the researcher places his research outside the realm if science.

6.06 Practice, Ethics & Integrity: Questionable research practices

Fabrication and falsification of data, plagiarism and unreported conflicts of interest are considered outright violations of research integrity. But there is also a grey area of practices referred to as questionable research practices or QRP's.
Questionable research practices refer to practices that are acceptable if they are implemented objectively and responsibly, but they can abused to obtain more favorable results. These practices generally refer to selective manipulation or massaging of data and selective reporting of results.Different types of QRP's can be distinguished. I'll discuss **harking**, **p-hacking**, **cherry-picking** and **selective omission** here.

Harking is short for 'hypothesizing after results are known'. This means the hypothesis is adapted to fit the observed data. Of course researchers are allowed to formulate new hypotheses based on the data they collected. This is basically what droves scientific progress forward.
Harking becomes a *questionable* research practice if the adapted hypothesis is presented as the original hypothesis without referring to the true original hypothesis. This is a highly questionable thing to do, because results that independently confirm an a priori hypothesis can be considered relatively strong form of support for a hypothesis. But here the hypothesis was a posteriori, formed after the fact.
And hindsight is 20-20, meaning that it is easy to find an explanation that fits a specific result. Prediction is much harder. Also, adaptation of the hypothesis based on the results means that the original hypothesis was not confirmed. This failure to support a hypothesis forms useful information for other researchers who are investigating the same phenomenon. This information is lost if the original hypothesis is never reported.

Let's turn to the questionable research practice of ***p-hacking*.** A statistical test is often used to determine whether an effect - a difference between groups or a correlation between variables - is large enough to be considered a confirmation of the hypothesis.
In most cases a probability called a p-value is used to decide this issue, hence the term p-hacking. **P-hacking** refers to data manipulation or selection that makes the results - in effect the p-value - more favorable. Data manipulation could

consist of transforming scores. Data selection could consist of selecting only certain items in a questionnaire, using only certain variables or removing one of several experimental conditions.

These selection and manipulation methods can be harmless if they are performed for good reasons, for example because scores are heavily skewed, questionnaire items are unreliable or certain variables or conditions show too much missing values to provide valid results. However, sometimes these methods are employed just because it produces a more favorable p-value. The confirmation or rejection of the hypothesis thereby depends on arbitrary choices of data selection and manipulation.
The golden rule of p-hacking is that as long as data selection and manipulation are reported and arguments are provided, the reader can judge for himself whether these choices are justified. P-hacking becomes a serious, questionable problem when the data 'massaging' is not or incompletely reported.

A special form of p-hacking is **cherry picking**: reporting only results that are favorable and significant, for example only one out of three experimental conditions and only one of two dependent variables. The opposite of cherry picking is **selective omission**. Selective omission refers to the omission of non-significant results but also omission of results that contradict the hypothesis.

A last specific type of p-hacking I want to mention is **data snooping**. Data snooping refers to the collection of data exactly until results show a favorable p-value. This practice is problematic because the choice to stop is arbitrary. Suppose the results are significant - the p-value is small enough - after collecting data from seventy-nine participants. It is entirely possible that the results will be unfavorable if the data for two more participants are included. Confirmation could be based on a fluke or extreme data from one participant.
The confirmation of a hypothesis should not depend on inclusion of data from an arbitrary participant that happens to pull the results far enough in the hypothesized direction. Sample size should be determined beforehand, based on non-arbitrary estimates of the expected effect size of a treatment and the required confidence level.

Again the golden rule in all these cases is that as long as choices are reported, they can be discussed and their influence on the results can be evaluated. If choices are not reported these practices can result in serious misrepresentation and misinterpretation of results, with no way to correct these errors.

6.07 Practice, Ethics & Integrity: Peer review process

Results of empirical studies are generally disseminated, or shared through specialized scientific journals. These journals publish articles that relate to a specific research field. Traditionally these journals charge universities and others high subscription fees. However, researchers are not paid for the articles that are published and cheaper, online publishing is replacing printing of journals.

This has led people to ask why the subscription fees need to be so high. Universities pay for the subscriptions to journals that contain articles that the universities pay their employees to write. According to many, scientific knowledge should be available freely to all, without making commercial publishing companies a lot of money. In the relatively new **open access** approach, researchers, or actually the universities, are asked for an author fee if a manuscript is accepted for publication. This allows open access journals to provide the content freely to all.

In either case, closed or open access, journals need to assess the quality of a submitted manuscript and decide whether it is good enough to publish. Submitted manuscripts in reputable journals are therefor subjected to **peer-review**. This means and editor asks two to four experts in the field to review the manuscripts. In most cases the author and the reviewers remain anonymous.
Reviewers do their review work for free. Reviewing is considered a professional responsibility. For some high-profile journals the review process takes several weeks. But in most cases the process takes several months, sometimes even a year or more.

Reviewers can reject a manuscript outright, which happens often, or they can accept it 'as is', which almost never happens. Most of the time reviewers will consider acceptance if the manuscript is revised. Revisions can entail rewriting the manuscript, but also extending or redoing a part of the study. The review and revision process can be repeated one, two or sometimes even three times before a manuscript is definitively accepted or rejected.

Journals obviously want to publish high quality research articles. But how do you determine whether reviewers did a good job and accepted good manuscripts? Journals determine their quality according to the number of times their articles are cited by other articles. These citation scores are used to determine the **impact factor** of a journal. Of course publishing in a high impact journal will increase a researcher's chances of being **cited**, because people expect high quality research in a high impact journal.

For researchers it is important to publish high quality articles, measured using the number of times their articles are cited. Of course it helps to publish more papers, because this increases the chance of being cited. Unfortunately the focus on measurable research output and the increasing competition for funding and faculty positions has led to an enormous **pressure to publish**, where more is better. This trend is often described using the phrase ‘publish or perish’.

It nicely illustrates the importance of publishing, especially for researchers who are starting out and are appointed temporary faculty positions until they have proven themselves extremely successful both in publishing their research results and obtaining grants to fund their research.

6.08 Practice, Ethics & Integrity: Dissemination problems

The number of publications, citations and the impact factor determine the reputation and success of both researchers and journals. This system has the unfortunate side effect of favoring the publication of new and exciting – significant – results. This is referred to as **publication bias**, a preference to publish confirmatory, positive results.

Replications, especially non-significant ones, are generally considered less interesting and have a greater chance of being rejected. This creates a problem: Non-significant results are underrepresented in the scientific literature. This distorts our understanding of the world around us.

This is made even worse by the **file drawer problem**. Writing a manuscript and submitting it for publication is an effortful time-consuming process. Researchers often feel it is not worth the time and risk to write up and submit a manuscript that describes a non-significant result, because the time and cost involved will generally result in rejection, or at best publication in a low impact factor journal. It is more rewarding to do a new study and hope for positive results and leave the null result study *in a file drawer*.

These negative side-effects of the current system for review and publication can resolved by reforming the process on several points. A first improvement is to require preregistration of the research question, design and statistical plan. This eliminates harking, cherry-picking and selective omission.
A second improvement can be made by basing the decision to accept or reject a manuscript on preregistered research proposals, instead of completed studies. This will eliminate publication bias and the file drawer problem, even if a study produces non significant results, it is accepted and published based on the hypothesis and quality of the research design, not the outcome.
Finally, if the data and statistical analysis are made publicly available this will reduce p-hacking and hopefully reduce fraud.

Not everyone is willing to change and become more open and transparent and freely share data and materials that cost a lot of time and effort to collect. But slow steps are being taken to reform the current review and publication process and to make the dissemination of research results more open, and transparent and more publicly available.