research • innovation • discovery

Guidelines for creating a comprehensive research protocol

This document was prepared by the Research Methods Unit (RMU) at Nova Scotia Health to guide new and early career researchers as they develop their research ideas into research proposals.

PROJECT TITLE

Project title should be/have:

- Concise, descriptive, and fully explanatory
- A short statement that describes the main topic or variables under examination
- Worded in terms of a functional relationship that indicate independent and dependent variables.
- Key words, for indexing purposes

PROTOCOL SUMMARY

This is a brief and comprehensive summary of the entire protocol (usually about ~200 words in length) and is often the first part of your study that people will read. Do not add information to the summary that is not discussed in the protocol.

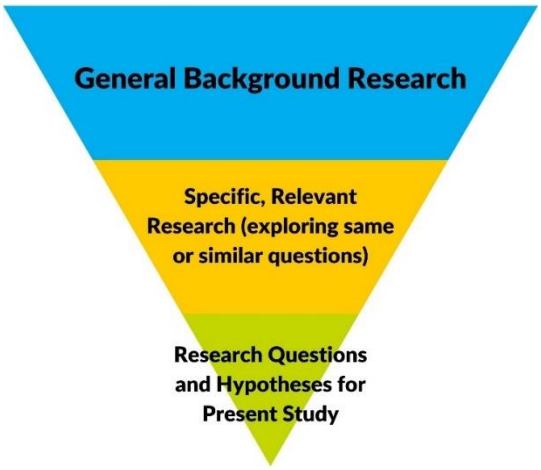
OR, IF PREPARING A MANUSCRIPT FOR PUBLICATION, USE:

ABSTRACT

Like the protocol summary, the abstract is brief and comprehensive. Word count varies depending on the publication.

A good abstract should consist of the main highlights of your protocol with respect to the problem under investigation, sample characteristics, essential features of your study's method, most important findings, and conclusion/implications. Do not add information that is not discussed in the protocol, but ensure it includes essential information – abstracts are indexed in search databases and is usually the first part of your study that people will read when doing a literature search.

BACKGROUND



The background section provides the reader with the necessary context for your research problem. This information should proceed from general to more specific information related to your topic. The overarching goal is to provide a theoretical and empirical justification for your research project.

When writing the background section, it may help to ask yourself:

- ***Why is the topic/problem important?*** *For example, is there a need to resolve differences in past research, extend research to new areas or populations, a neglected issue, enhancing treatment or practices, etc.?*
- ***What research has been done on this topic already – what do we know?*** *This does not have to describe all research ever done: summarize and critically evaluate representative major arguments and conclusions and provide new insights or conceptual frameworks.*
- ***What are the shortcomings of the current research?*** *This could be methodological issues, contradictions or inconclusive evidence in the literature, lack of generalizability of previous work etc.*
- ***How is your study different*** *and/or how does it add to the literature and the broader body of scientific knowledge on the topic (i.e., what gaps will your research address)?*

RESEARCH QUESTIONS

The background section should logically progress to the study's research question(s). There must be a clear link between the literature reviewed and the stated research questions. They must be fact-oriented, information-gathering questions, capable of being confirmed or refuted. Research question(s) can take many forms and are often followed up by more specifically defined objectives and aims. At a minimum, a single research question is needed. If the study is a pilot study or primarily a feasibility or exploratory study, this should be clearly indicated.

Building a research question

- Version #1. "does water kill?"
  We mean drinking water, not a tsunami or drowning by immersion.
- Version #2. Does drinking water kill?
  We mean fresh water, not salty water.
- Version #3. Does drinking fresh water kill?
  How much water? 1000 liters per day will kill you.
- Version #4. Does drinking a swig of fresh water kill?
  What is the source of the water? Tap, fountain, directly from the river…
- Version #5. Does drinking a swig of water from the Broad Street pump kill?
  Over which period?
- Version #6. Does drinking a swig of water from the Broad Street pump between August 31 and September 10 kill?
  Ok, compared with what? With drinking 3 liters of beer?
- Version #7. Does drinking a swig of water from the Broad Street pump between August 31 and September 10 kill compared with drinking all your water from other pumps?
  What about other factors that may affect the causal effect of interest?
- Version #8. Does drinking a swig of water from the Broad Street pump between August 31 and September 10 and not initiating a rehydration treatment if diarrhea starts kill, compared with drinking all your water from other pumps?

From: Hernán MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol. 2016;26(10):674-680. doi:10.1016/j.annepidem.2016.08.016

Below is a schematic of a research question that leads to more specific research objective:

"The goal of the current study is to [determine / examine / evaluate / establish / identify / compare / analyze / measure / etc.] the [difference / relationship / impact / effect / etc.] of X and Y. More specifically, this will be achieved by …"

METHODS

The methods section describes how the study will be conducted. It consists of major subheadings of participants, measures/materials, procedure, and statistical analysis plan. Methods should be rich in detail in order to allow for evaluation of the appropriateness of the methods, to assess the reliability and validity of the results, and to allow future researchers to replicate the study. Be sure to include enough information so that your readers can replicate your experiments or interventions.

This section should provide details of how the study will proceed step by step from the beginning to the end in chronological order. The information below is provided as a general

guide, but the information included will depend on the nature of the study. Include some information related to the following sub-sections:

Study design

- Is the study a cohort, case-control, retrospective, prospective, experimental, quasi-experimental, observational, cross-sectional, or something else? Justification/ reasoning for your choice of study design, and the potential advantages it provides in your research, should be included (e.g., the outcome is very rare, therefore we used a case-control study design).
- Does the study have between-subjects, within-subject, or mixed-design features?
- If the study is not a randomized controlled trial, details must be provided on how differences between the exposed group and comparison group will be controlled for at the study design phase.

Participants

- This section should provide details to allow examination of the appropriateness of your sample, particularly for generalizing the results. If there is prior knowledge of the total potential population size from which to sample from (e.g., a patient registry, all patients with condition X at hospital, all registered pharmacists in Ontario), include this information.
- It is good practice to include information to describe your sample (e.g., descriptive statistics). General sociodemographic, baseline, and clinical variables are appropriate so long as they are relevant to describing the sample and/or outcomes of interest. Also include study inclusion and exclusion criteria.
- The experimental/ treated/ exposed and the control/ comparison group should be introduced here. If the study is not a randomized controlled trial, details on how the comparison group was selected and how the researcher expects them to differ from the exposed group needs to be included.

Sampling procedures

- Method for selecting participants. For example, random sampling vs non-random sampling (e.g., quota, purposive, convenience) approaches. Details on what strategies the researcher will employ to obtain a representative sample of their target population, assess the representativeness of their sample, and mitigate loss to follow up should be included here.
- Payments to participants, how potential ethical concerns will be addressed, and safety monitoring procedures.

Specification of the variables under investigation

Importantly, the selection of any measure/material should be directly informed by the research questions and background literature review; if a measure is included, there must be a theoretical or empirical rationale in the study background section to justify its inclusion.

- Which variables are the independent and dependent variables?
- Which variables are predictor or outcome variables?
- Does theory or past research point to any confounding (e.g., nuisance, noise) variables that need to be controlled?
- As mentioned previously ensure there is a theoretical or empirical link between each variable and its inclusion in the study. This avoids casting too wide of a net, including irrelevant variables, and underpowered analyses.

Measures/Materials

- This section should provide details of the materials and measures needed for the study. Common examples include questionnaires and clinical / mechanical apparatuses.
- Include a clear but concise description of the questionnaires (e.g., number of items, rating scales, subscales) instruments, key settings and parameters, and scoring rules. Include information regarding their psychometric and biometric properties (i.e., reliability and validity).

Data collection

- Settings and locations in which the data will be collected and securely stored.
- Timelines for when data will be collected including start dates, end dates and number of collection time points if applicable.
- How will data be securely transferred from one location to another
- How participants will be recruited (e.g., in person, online, via chart review)
- For chart reviews, describe the database and how information will be accessed and extracted.
- How will the data be recorded? Will it be transposed directly from a chart to a data file? Are subjective ratings to be made by research assistants (if so, describing coding scheme)? How will interview data be handled?

If manipulations or interventions will be performed

- Who will carry the intervention and what is their level of training?
- How will conditions be manipulated, or interventions performed? e.g., number of sessions or events, duration, settings, in groups or individually

Describe procedures to enhance data quality

- Training for research assistants
- Reliability of the observers/coders (e.g., inter-rater reliability) and established criteria for successful reliability (e.g., intraclass correlation coefficients, Kappa statistic)
- Use of multiple observations.
- Data reliability checks

Systematic reviews or meta-analyses

Please note that guidance of systematic reviews and meta-analyses is beyond the scope of this protocol template. Given the importance of systematic reviews and meta-analyses in synthesizing bodies of research, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines should be followed (www.prisma-statement.org). Peer-reviewers and funders are increasingly adopting PRISMA (or similar) reporting standards guidelines to assess the completeness and transparency of a systematic review or meta-analysis.

DATA ANALYSIS PLAN

Describe the proposed data analyses and their importance in understanding how the research question and the data/measures are linked. The most appropriate analysis is often the simplest analysis (quantitative or qualitative) that is related to the study design and that will most effectively answer the research question. Include primary and secondary analyses and describe which were pre-specified or exploratory.

Research studies/questions can be exploratory in nature, but **an analysis plan should always be pre-specified**. Even with exploratory research, there is always a plan of which variables to analyze and which tests to use.

- Indicate how data will be screened for inaccuracies and statistical model assumptions.
- Will any baseline, demographic, or clinical descriptive statistics be provided?
- Assume your reader has knowledge of common statistical method (e.g., t-tests, correlations, ANOVA models, regression models, chi-square). Uncommon methods may require some citations supporting their relevance, appropriateness, and robustness

- Each statistical model should include a complete description of what variable will be predictor vs outcome variable, between-subjects and within-subject factors and levels in the case of ANOVA models, and so on. The assumed form of the outcome should always be specified (e.g., continuous and linear).

- For multivariable analyses, the analysis plan should include a detailed plan of how variables will be selected into the model. For an overview of options and practical implications see: *Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J. 2018 May;60(3):431-449. doi: 10.1002/bimj.201700067. Epub 2018 Jan 2. PMID: 29292533; PMCID: PMC5969114.*

- Methods to account for confounding, loss to follow-up and missing data should be included in detail.

- State your alpha level (i.e., p = .05; or justify any modifications such as Bonferroni or Benjamini-Hochberg corrections), and state whether confidence intervals (e.g., 95% or 99% confidence interval) or measures of effect size (e.g., Cohen's d, odds ratios) will be reported.

- For qualitative research describe how you will recruit participants and collect information, type of qualitative analysis, how the analysis will be conducted (e.g., on NVIVO), etc.

Sample size: Report the intended sample size and how this sample size was determined (e.g., power analysis, published sample size recommendations). You can use power calculators to help, such as gpower (http://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html)

If the study has different conditions or groups for comparison, report the intended sub-sample sizes. If a power analysis is performed, include all input parameters so that it can be evaluated in its entirety and replicated. If the study is inadequately powered, indicate any design elements or efforts to mitigate the lack of power and/or the potential effect on conclusions and interpretations. Studies that are known to be underpowered should have an explicit disclaimer that the study is not adequately powered and that to interpret results and conclusions with caution. Power analyses are not necessary for qualitative projects.

Statistical analysis plans are among the most important elements of a research protocol. It is of upmost importance that an analysis plan be specified in complete detail before collecting or analyzing any data. This avoids the ever-growing problems of p-hacking (conscious or non-conscious changing of analysis and/or methods to produce ideal results, cherry-picking only significant results) or HARKing ("hypothesizing after the results are known", aka, peeking) in

research. Even in the case of exploratory research, the details of an analysis plan should be provided. One way to protect against p-hacking, HARKing, and related problems, is to specify your analytic plan beforehand in detail and to not deviate from it. In some cases, deviations from an approved analysis plan are unavoidable; for example, if the obtained sample is smaller than anticipated or if large amounts of missing data occur. Revised analysis plans should be treated with extreme caution, discussed with the research team and if possible, reviewed by an independent person or committee.

Limitations

The limitations of the study should be outlined with a focus on the sources of potential bias or imprecision. A discussion around both direction and magnitude of any potential bias should be included. There can be many sources of potential bias or imprecision including confounding, selection bias, misspecification or outcomes or exposure, measurement error, modelling assumptions etc.

KNOWLEDGE TRANSLATION PLAN (based on guidelines from CIHR)

A brief description of how the authors plan to disseminate their findings should be included. Conference presentations and publications in peer-reviewed journals should be used for research at the early stages of discovery, when the knowledge has more relevance to academics who are contributing to a body of evidence that is not yet appropriate for application.

When there are potential knowledge-user audiences beyond the research community, knowledge translation activities could benefit from emphasis on nonacademic modes of communication such as websites, creative media (e.g. film, theatre, art), meetings/workshops with patient organizations, or online technologies (e.g. podcasting, webinars, YouTube). To disseminate more broadly to the general public, media such as television, radio and print may be engaged.

CONCLUSION

The protocol should conclude with a short summary of the purpose of the project, what is hoped to achieve, how this will benefit the population (patients and/or staff) or the health system, and its potential implications for theory, research, and practice. Benefits and implications need to be in line with the level of evidence provided by the study (e.g., results from a pilot study will not change clinical practice).