Guidance on protocols involving retrospective reviews of secondary data
Undergraduate Research Ethics Committee (UREC)

What are secondary data?
Secondary data refers to data that were collected by someone other than the *current* researcher. These are data that were originally collected for other research purposes, usually as part of a previous, larger study, or data in medical records that were collected as part of standard patient health care. A study that uses secondary data is called a 'retrospective study'. The word 'retrospective' refers to events that happened in the past, such as the collection of data during a previous research study or as part of standard medical care.

Advantages of using secondary data
- The use of these existing data provides a viable option for researchers who may have limited time and resources.
- From an ethics perspective, retrospective studies are typically considered minimal risk as they do not involve direct interaction with human participants. However, they still fall under the category of "research involving human participants" as they include data from or about individuals.

Disadvantages of using secondary data
- Keep in mind that each research project is undertaken with a view to answering a specific research question. How well the research question and objectives of the *current* study align with those of the original study must be carefully considered. The data may have been originally collected for a different purpose and therefore may not be optimal for the research problem under consideration in the *current* project.
- If the secondary data are from medical records or hospital databases, some data may be incomplete or missing and this could limit the overall quality of your dataset and ultimate findings. You should consider this issue in the limitations section of your protocol.
- When using medical records or hospital databases, there are important ethical considerations regarding consent. Patient data are usually collected as part of routine medical care, and not for research purposes. This means that patients would not have provided consent for the use of their medical or other personal information for research.

Guidelines on writing the methods section of your protocol

A: For studies using data collected as part of previous research

Important note: The information provided in the protocol should pertain to the *current* question, aims, and objectives, as well as the data and variables under analysis in the *current* study. If the *current* study is a retrospective review of existing data from previous research, it is not necessary to repeat all the details of the original study. The methods section of the protocol should mostly focus on the *current* study's methods, although some reference to the original study should be provided where appropriate.

Below are some guidelines about what to include in different sub-sections of the 'methods' section of your protocol:

Study design
State that the current project forms part of a larger project or uses data that were gathered as part of another study. The researcher should provide the following regarding the original study: 1) the ethics approval number; and 2) the purpose (aim/objectives) of the original study. If the original study has been published, you can provide a reference to the appropriate journal article(s).

Example:
"This study will involve the analysis of data from participants recruited as part of the Biological Endophenotypes of HIV and Childhood Trauma (BEHCT) study conducted at Stellenbosch University, Cape Town (HREC reference N19/00/001). The aim of the BEHCT study was to assess the genetic, cognitive, and neuroimaging outcomes in HIV-positive women. This study utilizes an adult female cohort (18–65 years),

recruited between 2008 and 2013, with samples genotyped in 2017. The study design, data collection methods, and the inclusion/exclusion criteria for the BEHCT study have been described previously.27 The main procedures applicable to the current study are briefly outlined below."

Study setting
The setting of the original study should be described because this provides important context for the *current* study. You should look back at the protocol for the original study and summarise the details provided there.

Participants
The characteristics of the original study participant sample should be described, as well as the sampling strategy and final sample size. If the *current* study uses data from a subset of the original participants, the researcher must explain the participant characteristics that make them eligible for inclusion in the current study. You should motivate for why some patients will be excluded i.e. motivate for why only a subset and not the full sample, will be the focus of the *current* research.

Data collection
The researcher must clearly detail, in a data extraction sheet or table, what data will be extracted and recorded from the larger/original dataset, medical records, or database, including any demographic or personally identifying information. The variables can be listed or placed in a table format in the data extraction sheet. The protocol should include a clear rationale for the collection of these specific data. For example, if you collect information relating to ethnicity, you need to strongly motivate for why this information would be relevant to your study.

Data analysis
The researcher should explain how their data (i.e. the data used in the *current* study) will be analysed. Note that in cases where data from a previous study are used, this analysis plan may be different from the data analysis section in the original protocol, especially if the current study aims to use only a subset of the original data.

Ethical considerations

With respect to secondary use of stored data or samples, the steps regarding consent are outlined in the Department of Health (DoH) (2015) ethics guidelines (pp. 43-44, 3.3.7) and in the Health Research Ethics Committee (HREC) Standard Operating Procedures (SoPs) (2019) (pp. 56-60, 11.5) as follows:

1. **Request for proof of original consent** to store and use the data/samples for future research – i.e. the original informed consent document that was used in the original study must be submitted together with the current ethics application.

2. If the original consent **covers permission to store data for future use *and* covers the objectives of the current research project**, then no new consent is required.

3. If this original, future-use consent **does not cover the scope of the current research, then new consent may be required.**

4. **Failing to demonstrate consent** as described in the above three steps, and with justification as to why new consent cannot be obtained, **UREC may consider a strongly motivated waiver of consent request** (see HREC SoPs section 11.5.8 pp. 59-60)**.** Guidance on applying for waivers of consent can be found here (and see HREC SoPs section 11.4 p. 55). In such cases, the request should demonstrate not just the standard waiver of consent conditions of anonymised data at the point of collection and aggregated findings, but also justification as to why this research is believed to not pose the risks outlined in 3.7.7.iii) (p. 44) in the DoH guidelines. The researchers should also highlight the social value of the *current* research study.
   UREC has previously given feedback to this effect as follows:

UREC will consider whether granting a waiver of informed consent for use of these samples is ethically appropriate given the unique circumstances and potential social value argued for by the applicant. If UREC were to grant a waiver of informed consent, it would be based on consideration of whether the requirement for informed consent (respect for autonomy) could in this case be exceeded by the potential social value (anticipated benefit to society) of the research as well as any potential benefits to patient management/care that are anticipated. The researcher is encouraged to submit convincing evidence of these anticipated benefits and social value in the revised protocol.

5. In this request for a waiver of consent, the researchers should also specify whether the data are stored in an **anonymised** or a **de-identified** database (see HREC SoPs pp. 78-79) and, if de-identified, what, if any, access the researchers will have to the personally identifying data. **NB**: For a waiver of consent to be considered, data must be anonymised at the point of data collection in the *current* study: i.e. when the current researchers extract the data, *no* personally identifying information (e.g. file numbers) should be recorded in any way.

B: For studies using data collected as part of standard health care

We have used the protocol from the published paper by Botha et al. (2020), and others where appropriate, for the examples in this section.

The aim of this study was: *"To describe the demographic and clinical profile of all patients admitted to New Beginnings between 01 January 2011 and 31 December 2015."*

Study design
This section provides an overall 'snapshot' of your investigation. There are four important elements to include: 1) your study design; 2) the key characteristics of your sample; 3) your study setting; and, 4) the time period under investigation.

Example:
"We will conduct a retrospective audit of patients admitted to New Beginnings between 01 January 2011 and 31 December 2015."

Study setting
This refers to the clinical, physical, social, political and/or cultural context in which the researcher conducts the study. A reviewer should be able to understand how your setting is similar or different from other settings. Important elements to consider for this section: 1) the type of institution e.g. "large university hospital", "small outpatient clinic"; 2) the level of care i.e. general population, primary care, secondary care, etc.; 3) financial accessibility (i.e. public vs. private); 4) standard of care in this setting; 5) the kind of geographic region from which that institution draws its patients; and 6) the socio-demographic status of the population served.

Example:
"New Beginnings is situated in Bellville, Cape Town, and was launched in 2008 because of the increasing need for acute psychiatric services. It is an intermediary care facility focused on psychosocial rehabilitation and accommodates 40 patients in a step-up or step-down setting. Patients discharged from the acute inpatient wards at Stikland Hospital and needing further psychosocial rehabilitation may be referred. These patients are stabilised on psychotropic medication and referred to New Beginnings as voluntary patients. New Beginnings is regarded as a ward of Stikland Psychiatric Hospital from a management perspective and is funded by the Department of Health. All medication comes from the Stikland Hospital budget and patients go to the Stikland outpatient department on a monthly basis for review and treatment by a medical officer. The main focus of the programme is not only on stabilisation of acute episodes of serious mental illness (in comparison to acute inpatient services) but also on empowerment, skills acquisition and structuring to individual patient needs. The staff complement of New Beginnings includes a facility manager, occupational therapist, occupational therapist technician, social worker, operational manager, administrative management team consisting of two administrative clerks, a facility driver and a housekeeper, six professional nurses, seven enrolled nurses, eight

nursing assistants, two cooks, four cleaners and three security guards (day and night). The mental health staff are trained professionals with experience in psychosocial rehabilitation and psychiatry. A medical officer (based at Stikland Hospital) provides care as needed and there is access to a psychiatrist if more input is required. State psychiatrists are joint appointments, two-thirds clinical and one-third academic. New Beginnings also make use of contracted home-based carers who work on rotational shifts and guide the patient programme in terms of activities of daily living (self-care, chores, meal preparation, clothing care, escorting patients to appointments and use of leisure exploration activities). Criteria for admission to New Beginnings are summarised in Box 1."

Participants
This section should include 1) inclusion criteria; 2) exclusion criteria; 3) the estimated size of the total population; and, 4) the expected size of the final sample.

Example
"We will use a convenience sample of all patients (both male and female) admitted to New Beginnings between 01 January 2011 and 31 December 2015. No exclusion criteria have been identified. The total number of patients admitted during this period was 730 people."

Data collection
Here you need to describe the process of data collection in a logical sequence and include the following details: 1) From where will the data be retrieved? 2) How will the data be extracted? 3) Who will extract the data? 4) What data (variables of interest) will be extracted? 4) How will diagnoses (if any) be made?

Example
"The primary researchers (HB) will retrieve patient files from the New Beginnings archives storing unit. Other data will be extracted from the Clinicom Application Manager (a Western Cape hospital database keeping record of patient demographics, outpatient appointments and hospital admissions). All data will be collated on a data collection sheet (Appendix A) and later inputted to Microsoft Excel spreadsheet with cross checking by YV (a co-researcher). Data extracted from the patient charts will include demographic information (age, gender, residential area, language, highest level of education, marital status, number of children, employment, primary carer and disability grant) and clinical information (assertive community treatment [ACT] patient, primary and secondary diagnosis, number of psychiatric inpatient admissions and total days in hospital, number of New Beginnings admissions and total days in New Beginnings, medication, comorbid substance use and treatment for medical conditions). Data extracted from Clinicom will include all past psychiatric hospital admission dates. Diagnoses were made in the clinical setting, based on the Diagnostic and Statistical Manual of Mental Disorders IV, fourth edition, text revision (DSM-IV-TR) or the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) diagnostic criteria.[20]"

Data analysis
For descriptive studies, include only details about how you will summarise numerical and categorical data (i.e. descriptive statistics).
For analytical studies, include: 1) descriptive statistics; 2) the inferential statistics you will use to test your hypothesis; 3) the software used for analyses; and 4) the p-value cut-off (usually $p > 0.05$).

Example 1:
Descriptive study (Botha et al., 2020): *"Continuous variables will be summarised as mean and standard deviation, while nominal variables will be summarised as counts and percentages. As this is a descriptive study, no inferential statistics will be employed."*

Analytical study (Davis et al., 2020): *"Due to the small sample size, data will be expressed as medians with interquartile ranges for continuous data while categorical data is expressed as frequencies with the number of participants indicated. Differences between PET/CT positive and PET/CT negative patients will be analysed with a Mann-Whitney U test. Statistical significance will be considered at $p<0.05$. The data will be analysed using Statistica version 12.0 (StatSoft Inc, Tulsa, OK, USA)"*

Ethical considerations

In any research involving human participants – including information about or linked to human participants such as medical records or other health-related data – written informed consent is generally required. UREC and/or its parent committee, HREC, may waive the requirement to obtain informed consent provided that:

- The research involves no more than minimal risk to the participants,
- The waiver or alteration will not adversely affect the rights and welfare of the participant(s),
- The research could not practicably be carried out without the waiver or alteration, and
- Whenever appropriate, the participants will be provided with additional pertinent information after participation.

A formal waiver of consent request must be submitted together with your ethics application. The **minimum conditions that must be met for a waiver of consent to be considered are** that the data must be anonymised at the point of data collection/extraction (i.e. *no* personally identifying details, such as hospital file numbers) may be collected, and the data should be reported in aggregated form (i.e. no single cases described). Guidance on how to apply for a waiver of consent can be found on the Undergraduate Research Office ethics FAQs page.

Example:
"Ethical clearance will be obtained from the Health Research Ethics Committee of Stellenbosch University, as well as from the Head of Establishment at Stikland Hospital. We will request a waiver of informed consent for this retrospective study. All data will be anonymised to ensure privacy and confidentiality of patients' personal information, with each participant assigned a unique identifier."

References

Botha, H.F., Koen, L., Niehaus, D.J., Vava, Y., Moxley, K. and Botha, U., 2020. Demographic and clinical profile of patients utilising a transitional care intervention in the Western Cape, South Africa. *South African Journal of Psychiatry*, 6(0), a1523.

Davis, J.H., Burger, M.C., Pienaar, G. and Lambert, R.G.., 2020. 18F-FDG PET/CT as a modality for the evaluation of persisting raised infective markers in patients with spinal tuberculosis. *SA Orthopaedic Journal*, *19*(1), pp.23-27.

Hox, J.J. and Boeije, H.R., 2005. Data collection, primary versus secondary. In Kempf-Leonard, K. (ed), Encyclopaedia of Social Measurement, Elsevier Inc., pp. 593 - 599.

Jacobs, S., Moxley, K., Womersley, J.S., Spies, G., Hemmings, S.M. and Seedat, S., 2018. HPA-axis genes as potential risk variants for neurocognitive decline in trauma-exposed, HIV-positive females. *Neuropsychiatric disease and treatment*, *14*, p.2497-2504.

Compiled by:

Dr Debbie Marais, Undergraduate Research Office, Faculty of Medicine and Health Sciences, Stellenbosch University

Dr Karis Moxley, Registrar Research Support Office, Faculty of Medicine and Health Sciences, Stellenbosch University

Dr Phoebe Runciman, Department of Sport Science, Faculty of Medicine and Health Sciences, Stellenbosch University

Date last updated: 25 November 2020