Guidelines for reproducible and reliable experimental design and statistical analysis involving human brain tissue

These guidelines have been assembled in order to provide a reference for researchers who are involved in the design and/or assessment of studies using human brain tissue.

Human brain tissue is a precious research resource and the Sydney Brain Bank (SBB) will only approve those studies that are deemed to be meritorious and also not wasteful, although sufficient numbers of samples for the major conclusions are required. It is expected that the results arising from studies that use human brain tissue supplied by the SBB be of a quality that can be published in peer-reviewed journals. It is therefore important to consider the requirements of peer-reviewed journals in the preliminary stage of study design, including the tissue application phase.

A common set of principles and guidelines for biological research has recently been agreed on by major journals, funding agencies and scientific leaders (www.nih.gov/about/reporting-preclinical-research.htm). In addition to this, a framework for improving the design and analysis of laboratory-based research has also recently been published [1]. Many of the principles outlined are applicable to all biological research, but use of human brain tissue introduces some unique elements that need to be considered when designing/assessing a study.

The purpose of these guidelines is to raise awareness of issues related to the study of human brain tissue, and thereby increase experimental reproducibility and reliability.

Considerations

Reproducibility and reliability

An essential element of the scientific process is ensuring reproducibility of data. However, good reliability must be balanced with an awareness of case availability and quality for research purposes. Often larger numbers of cases are required to achieve statistical significance, especially when the effect size is considered small and/or experimental variability is high. In the case of human tissue studies, experimental variability is usually high with factors such as age, gender, post-mortem delay, genetic variability and disease stage all playing a role. In addition, the source of the tissue can also contribute to such variability and factors such as tissue processing, characterization and storage all impact on potential outcomes. Variation within the population could be considered to be ~30-35%.

Major sources of variability include:

1. Diverse genetic and environmental background. Collections of donated human brain tissue are inherently heterogeneous in nature as they are

composed from a population with diverse genetic and environmental backgrounds. This is in contrast to animal or cell-based studies where both of these factors are well controlled. Therefore, even when variables such as age and gender are taken into account, there usually is significant deviation in any measured quantity. This represents the real world and needs to be factored into any human tissue research study if the effect size of the research measurement is thought to be small.

2. Disease stage. A measured effect may vary significantly according to the disease stage. Standardising for disease stage and/or including disease stage in the study design as a factor to be evaluated and/or restricting the cases to a specific stage may reduce the numbers required.
3. Population bias. Population effects may be present due to the limited collection area of most brain banks. Adding to this may be differences in tissue characterisation, processing and storage protocols. Therefore, if it is important to establish the generality of a measured effect, it may be important to obtain material from a variety of sources.

***Effect size*:**

Despite these potentially large sources of variability, robust, disease-specific effect sizes can be large [2, 3], and therefore it is important to have an idea of what the effect size might be in the human in order to ensure that the minimum numbers are used to obtain a reliable but generalisable result and that using unnecessarily large numbers of cases is avoided (it is better to replicate in different population collections). If the aim of the study is to prove a negative result, then this can often mean the inclusion of more substantial numbers and sampling in order to account for the sources of variability associated with the use of human brain tissue.

If the effect size is not known, a preliminary proof of effect study may be appropriate. The SBB can provide limited “trial” tissue for the purpose of developing new techniques or gaining preliminary data for power analyses.

Statistical analysis

Once the sources of variability have been accounted for/minimised and an estimate of the effect size established, then a power analysis is highly recommended as part of the justification for the requested case numbers. The type of power analysis employed will depend on the statistical analysis tests/models that will be employed in the research project. A power analysis consists of 4 variables: the sample size, effect size, significance level and power level. Inputting any 3 of these variables will determine the remaining one. Therefore if the expected effect size (based on preliminary observations or published data) is known along with the pre-determined acceptable significance level and power level, then the appropriate case numbers can be determined and justified.

A key recommendation arising from the RIPOSTE framework [1] is to encourage early discussion of the study design and analysis within a multidisciplinary team including

statisticians. If the researcher has insufficient experience with selecting the appropriate statistical tests and/or power analysis, then consultation with a statistician is prudent.

Power analysis example

The most common type of power analysis carried out when designing a study is termed "a priori", where the researcher is interested in determining the sample size given that they know the effect size, significance level and power level. The significance level and power are usually set to 0.05 and 0.8, respectively. The effect size is the expected effect and can be inferred from pilot study results or published findings from a similar study.

In this example the effect size was determined to be 1.0 (considered large) from a previous study. The type of test to be used is a t-test, means: difference between two independent means (two groups). The software used was G*Power (available as a free download).

1. Select the type of power analysis (a priori)
2. Select the type of statistical test that will be used (t-test, means: difference between two independent means (two groups))
3. Input the significance level (0.05)
4. Input the level of power (0.8)
5. Input the effect size (1.0)

The sample size required would be 28 (14 per group).

If the level of significance were increased (to 0.01), then the sample size required would be 44 (22 per group).

If the power required were increased (to 0.9), then the sample size required would be 36 (18 per group).

If the effect size were less (0.8), then the sample size required would be 42 (21 per group).

References

1 Masca NG, Hensor EM, Cornelius VR, Buffa FM, Marriott HM, Eales JM, Messenger MP, Anderson AE, Boot C, Bunce C, Goldin RD, Harris J, Hinchliffe RF, Junaid H, Kingston S, Martin-Ruiz C, Nelson CP, Peacock J, Seed PT, Shinkins B, Staples KJ, Toombs J, Wright AK, Teare MD. RIPOSTE: a framework for improving the design and analysis of laboratory-based research. Elife 2015; 4

2 Hardman CD, Halliday GM, McRitchie DA, Morris JG. The subthalamic nucleus in Parkinson's disease and progressive supranuclear palsy. J Neuropathol Exp Neurol 1997; 56: 132-42

3 Murphy KE, Gysbers AM, Abbott SK, Tayebi N, Kim WS, Sidransky E, Cooper A, Garner B, Halliday GM. Reduced glucocerebrosidase is associated with increased alpha-synuclein in sporadic Parkinson's disease. Brain 2014; 137: 834-48