BSMS Statistics Handbook

September 2025

brighton and sussex medical school

Contents

1. Statisticians at BSMS

Dr Chris Jones
Principal Research Fellow in Medical Statistics

Department of Primary Care and Public Health

Dr Saskia Eddy
Assistant Professor in Medical Statistics

Department of Primary Care and Public Health

Dr Vasso Anagnostopoulou
Senior Research Fellow in Medical Statistics

Department of Global Health and Infection

Department of Primary Care and Public Health

BSMS statisticians are based in the Department of Primary Care and Public Health and work on projects across BSMS, particularly with the Brighton and Sussex Clinical Trials Unit.

2. Accessing statistical resources at BSMS

Statisticians at BSMS provide free advice through Statistics Clinic and are costed into grant applications as Co-Is.

Note that we do not independently verify analyses performed by AI (see section 6 for issues relating to the use of AI). Instead of using AI we recommend getting in touch, so we can assist you with any analyses.

2.1 Statistics Clinic

Statistics Clinic is open to BSMS staff and research students free of charge. It is often the first point of contact between researchers and statisticians. 1 hour one-to-one appointments can be held either online or in-person and can be requested using the online booking form. Advice can be provided for quantitative projects at any stage.

2.1.1 Getting the most out of your stats clinic appointment

- We may not be familiar with your area of research, so be prepared to give a jargon-free general overview of your project.
- Please provide as much detail as possible when filling out the form so we can best prepare for your appointment.
- You can send data ahead of the meeting (or bring it with you if meeting in-person).
- Data can be in Excel, SPSS, Stata or GraphPad formats, or as a plain-text file (.csv files etc.).
- For sample size calculations, we need to know what outcome variable the calculation will be based on, have estimates of its expected values (e.g. mean or proportion) and variability (e.g. standard deviation), as appropriate, and how many groups there will be (e.g. treatment and control). If your study is longitudinal, you will likely need to factor in attrition over time. Please consider these quantities before your appointment.
- When discussing results, the methods and/or code use to create them results should also be provided.

2.2 Grant applications and funded time

For large quantitative projects, it is advisable to include a statistician as a Co-I. Statistical input cannot be costed generically as a "service" as it is required throughout the project's lifespan and is unique to every project.

Grant reviewers will assess whether the proposed analyses are appropriate and that the study team have the skills required to perform the analyses. Inclusion of a statistician in the team is a strong indicator that analyses have been carefully considered and are achievable.

Statisticians will also contribute to the grant application (such as study design, analyses, sample size) and write relevant sections that will pass scrutiny by statistical reviewers. When included as part of the project team, statisticians will be able to invest time, at no cost, to develop the application.

Where inclusion of a statistician is not feasible, advice on grant applications and analyses can be sought through stats clinic, but this will be limited by capacity.

2.2.1 BSMS Statistician roles and costing guidance

During the course of a study, a statistician may undertake any number of the roles below. Not all roles will apply to every study, and there may be additional roles required for specific studies. Roles in bold require substantial amounts of time.

Prior to study setup and start

- **Input into funding application and writing sample size/analysis sections.**
- **Input into study protocol and writing sample size/analysis sections.**
- Writing an early version of the statistical analysis plan.
- Input into ethics applications.
- **Input into database design and review of test exports.**
- **Generation of randomisation list and set up of randomisation system.**
- **Liaising with co-applicants/study management group to address to address statistical queries.**

While the study is running

- **Preparation of the Statistical Analysis Plan(s).**
- **Production of reports for study management meetings.**
- Attendance at study management meetings.
- **Production of reports for study Steering Committee and Data Monitoring Committee meetings.**
- Attendance at study Steering Committee and Data Monitoring Committee meetings.
- Draft final analysis code.
- **Liaising with study management groups (TSC, DMEC) to address statistical queries.**

Towards the end of study recruitment/follow up and at the end of the study

- **Perform interim analyses as detailed in the SAP and production of interim analysis report.**
- **Perform final analyses as detailed in the SAP and production of final analysis report.**
- Input into reports to funders.
- **Input into study publications (writing methods/results sections, producing graphs and tables) and study publication review process.**
- **Production of final study datasets for use within the study team, archiving, and/or to be made publicly available (which may include anonymisation of data).**

Normally, either a single senior statistician will be costed into a project, or a senior statistician will be costed in for a small amount of time along with more time for a junior statistician. Study design and input into grant applications is uncosted. In the table below, the indicative Full Time Equivalent (FTE) amounts are for a single statistician, but higher amounts could be divided amongst more than one.

For clinical studies, particularly trials that involve long recruitment or follow-up periods, the statistician(s) will be most intensely involved at the start during study set up and towards the end (preparing for and conducting the final analyses). For trials involving the BSCTU, statistician time will be included in the BSCTU costing.

Note that University of Sussex costing guidance (Sussex login required) from February 2025 onwards sets the minimum Co-I academic time to "*A minimum of 5% FTE per person over the lifetime of the project for each Co-I involved*".

FTE (%)	Example statistician workload
2.5	***Note that as this level is below the 5% minimum it requires specific justification and approval*.** Attendance at some study team meetings. Supervision of very limited statistical analyses performed by someone else, e.g. a research fellow. Review into statistical sections of manuscripts etc.
5	Attendance at study team meetings. Review of statistical sections of manuscripts etc. Supervision of statistical analyses performed by someone else, e.g. a research fellow.
10	Attendance at study team meetings. Writing statistical sections of manuscripts etc. Supervision of substantial statistical analyses performed by someone else, e.g. a research fellow OR performing limited statistical analyses ourselves.
20	Attendance at study team meetings. Input into study reporting. Performing the statistical analyses. Writing statistical sections of manuscripts etc. Preparation of analysis datasets for sharing.
30+	Attendance at study team meetings. Production of study monthly reports. Performing substantial statistical analyses. Writing statistical and results sections of manuscripts etc. Preparation of complex analysis datasets for sharing (e.g. those requiring anonymisation).

2.3 PhD supervision

For PhD projects involving substantial quantitative analyses, or align with a particular statistician's research interests, it may be possible for a statistician to join the supervisory team. This can be discussed directly with the statistician.

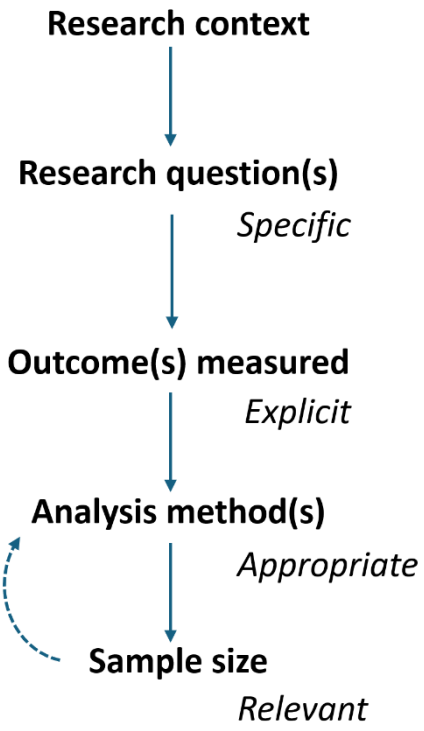
3. Study design and statistical considerations for grant applications

Research is conducted to expand human knowledge, allow us to understand complex systems and make informed decisions. To be useful, it must be conducted in a coherent, robust way that allows us to ask a question, figure out how to answer the question, collect and analyse the required data, and reach appropriate conclusions.

We aim for our results and conclusions to be correct, unbiased and reproducible.

3.1 Underlying structure of any study design

The following steps are required for a coherent study design. The first three follow a linear order. Analysis methods to answer the research questions may need iterative consideration depending on external limitations to potential sample size.



3.2 Good and bad examples of study design steps

Step	Good examples	Poor examples
Research Context	A new intervention may improve patient outcomes. Understanding the pathway of a biological process.	
Research question(s)	Does the new intervention improve quality of life for patients compared to the control group? Does a mutation in gene X affect expression of gene Y, compared to wild-type?	We want to examine quality of life, life expectancy, side effects, costs to the NHS, etc. for patients receiving the new intervention, and see if these effects are greater or lesser for older patients, men/women, or those with previous history of….
Outcomes measured	EQ-5D-5L index score (continuous). Expression of gene Y (continuous).	"Quality of Life". Difference in expression of gene Y.
Analysis method(s)	Multiple linear regression. t-test. (Any appropriate analysis method).	(Any inappropriate analysis method, including analyses performed by AI).
Sample size	Based on appropriate statistical test for outcome and analysis method, or justification based on precision (width of 95% confidence interval around an estimate).	Any excuse that leads to "we don't need to consider sample size/power", e.g. "This other paper was published with n=3 in each group, so we need n=3 in each group". Or something based on the wrong statistical test.

3.3 Common research study types

Study type	Description	Sample size and power	Example uses	Appropriate analyses	Relevant protocol/reporting guidelines
Pilot/Feasibility	A feasibility study is used to inform whether it is possible to conduct a larger future study or a particular aspect, or aspects, of it. It should aim to resolve the key uncertainties in the design and conduct of a future such study. Feasibility studies may lack some features of the full study, such as some arms or randomisation. Pilot studies are a type of feasibility study designed as small-scale version of a planned larger study and are conducted before the larger study to ensure the planned protocol is feasible. They will typically have all the design features of the full study, such as all arms and randomisation.	No formal sample size calculation, but sample size should be justified in terms of the feasibility objectives e.g. estimating the proportion retained to final follow up, to a specified level of precision. Typically, tens of participants depending on objectives. Could be smaller for less variable populations, e.g. model organisms).	Testing processes (e.g. recruitment methods). Assessing multiple possible outcomes in search of a good primary outcome for a full study. To collect initial data to inform sample size calculations for future studies. Checking that the selected primary outcome can be collected as expected and is an appropriate measure for the study population (e.g. has no obvious floor or celling effects). Estimating attrition/loss to follow up To check assumptions made for sample size calculations for future studies are realistic and adequately precise.	Descriptive statistics and confidence intervals (no hypothesis testing of clinical/efficacy outcomes). Exploratory statistical testing of secondary/clinical outcomes *may* be acceptable in larger feasibility studies, but is generally discouraged. While within participant effects may sometimes be of interest in feasibility studies, these must be interpreted appropriately: any effects observed attributed to the intervention and are not indicative of between group effects to be tested in the full study.	Clinical trials: CONSORT Extension for Pilot and Feasibility trials. Guidelines for the reporting of protocols for *randomised* pilot and feasibility trials Guidelines for the reporting of protocols for *non-randomised* pilot and feasibility studies
Exploratory	A study without pre-defined outcomes, analyses or sample size. Or studies technically powered to observe a difference between groups, but with insufficient participants to allow analysis models with appropriate complexity to be fitted.	Indicative sample size calculations may be performed. Power for given situations, or detectable effect size may be considered. (e.g. 40 participants would provide Y power to detect an effect size of X).	Situations necessitating compromise between study design and cost/availability of participants. Small studies technically powered to detect a stated effect size but with insufficient numbers to allow appropriately complex modelling (e.g. a study comparing disease occurrence between two groups but without sufficient numbers to include important demographic characteristics in the modelling).	Descriptive statistics and confidence intervals. Hypothesis testing within simple models, with *cautious* interpretation of the results due to high risks of type I and type II errors, and the subsequent risk of biased conclusions. If strong conclusions are sought, a confirmatory powered study should be conducted instead.	

Study type	Description	Sample size and power	Example uses	Appropriate analyses	Relevant protocol/reporting guidelines
			Studies where many outcomes are being measured (e.g. gene expression studies) and it is not clear how many will show differences between groups.		
Confirmatory **(May also be referred to as 'Full scale', 'definitive', 'main' etc.)**	A sufficiently powered (i.e. usually at least 80%) confirmatory study with pre-defined outcomes, analyses, sample size calculation (for specified power), and stated minimum biologically/clinically important difference(s).	Sample size determined by formal calculation relevant to primary outcome, design and thence analysis method, taking into account attrition/loss to follow up. Could be tens to tens of thousands of participants.	Experimental studies (including clinical trials). Large observational studies.	Descriptive statistics and confidence intervals. Pre-defined hypothesis testing and estimation within appropriately complex modelling. Some secondary exploratory analyses may be appropriate.	Clinical trials: SPIRIT for protocols and CONSORT for the results publication. Observational studies: STROBE guidelines. Lab studies: RIPOSTE guidelines.
Secondary analysis of pre-existing or routinely collected data	Studies involving the secondary analysis of pre-existing datasets that were not collected specifically for the current study, or data not collected primarily for research purposes.	Maximum sample size set by size of pre-existing dataset. Consideration of calculations for power and/or detectable effect size should be presented to show objectives are realistic and allow for missing data.	Examining relationships between exposures and disease in routinely collected patient data. Re-analysing previously published data with new objectives or methods. Studies involving Machine Learning methods. Clinical audits.	Descriptive statistics and confidence intervals. Pre-defined hypothesis testing within appropriately complex modelling. Some exploratory analyses may be appropriate.	Routinely collected data: RECORD guidelines.
Systematic reviews and meta-analyses	Studies that collect and analyse published results from the literature.	Availability of published results is determined by the extent of the literature.	Used to summarise extent and quality of available literature. Meta-analyses combine evidence from similar studies.	Descriptive statistics and confidence intervals (in particular, forest plots).	Protocol: PRISMA-P guidelines for protocols Reporting results: PRISMA guidelines. Risk of bias tools Cochrane Handbook for Systematic Reviews of Interventions.

Study type	Description	Sample size and power	Example uses	Appropriate analyses	Relevant protocol/reporting guidelines
Methodology studies	Studies that describe or analyse methods (design, conduct, analysis or reporting) in published (or unpublished) literature.	For studies examining the literature, such as methodological systematic reviews, the sample size is determined by the extent and availability of the literature. A formal sample size calculation, or confidence interval approach (precision based calculation) could be performed depending on the research question.	Pre-cursors to reporting guidelines, to gain an understanding of current practice and possible gaps. Factors associated with reporting practices Appraise quality, quantity, completeness, accuracy and consistency of health research.	Descriptive statistics and confidence intervals	Reporting results: MISTIC checklist (under development)

The Library of Guidance for Health Scientists (LIGHTS) collates methods guidance papers and can be used to identify published guidance in a specific area/for a specific design.

4. Statistical, data collection and visualisation software

Software	Sample size calculations	Data collection/ database	Statistical analysis	Publication quality presentation of results	Comments
Epi Info	(✓)	✓	(✓)	(✓)	Freely available, but soon to be unsupported.
G*Power	✓	x	x	x	Freely available.
GraphPad Prism	x	(✓)	(✓)	✓	Available on some BSMS computers. Beginner friendly. Limited contingency table or regression functionality. Excellent at graph/figure creation. Excellent user guide explaining statistical tests and analyses.
REDCap	x	✓	x	x	Hosted by Sussex, free to use for BSMS staff. For access: https://bsmsredcap.sussex.ac.uk/
Microsoft Access	x	✓	x	x	Available on all BSMS computers.
Microsoft Excel	x	x	x	x	Excel is **not recommended** for data collection, analysis, or visualisation. Available on all BSMS computers.
R	✓	x	✓	✓	Freely available with extensive data manipulation and analysis tools. GUIs available, e.g. RStudio
Python	✓	x	✓	✓	Freely available with extensive data manipulation and analysis tools. Jupyter is useful for organising analysis code and results.
SPSS	x	(✓)	✓	(✓)	Available through University of Sussex Software Hub. Can be used for data collection for smalls studies.
Stata	✓	x	✓	(✓)	BE version available through University of Sussex Software Hub SE version requires purchase of a license, but BE version is sufficient for most analyses.
Inkscape	x	x	x	✓	Freely available.
Adobe Illustrator	x	x	x	✓	Full Adobe Creative Cloud available on BSMS computers. UoB and UoS licenses allow use on personal computers.

5. General statistical advice

- Statistician(s) should be consulted or involved, as appropriate, from project design through to final publication.

- Don't leave anything to the last minute. We are unlikely to be able to advise you at short notice.

- Statistical methods used across research areas are often similar, despite substantial differences in the language used to describe them.

5.1 Data collection

- Data collection should be planned and performed carefully. If not, extensive data cleaning will likely be required before analysis; this can be extremely time consuming.

- Handling dates and missing data can cause unexpected issues.

- Keep detailed records of decisions made during the project (e.g. what to do with an unexpected measurement) so that their effect on the analysis can be considered.

- Even seemingly simple questions can require a considerable data manipulation. It is important to consider how data will be used when designing collection tools/databases.

5.2 Analyses

- Analyses should be planned at the study design stage as appropriate methods are essential to ensure research questions can be answered, and must be known for a sample size/power calculation to be performed.

- Complexity of analyses performed should correspond to the amount of data available and complexity of the system being modelled. Lab data can often be analysed with relatively simple models (e.g. t-tests, ANOVA etc.) as many variables can be controlled in the experiment. Studies on humans or human samples require more complex models to adjust for variables that cannot be controlled directly (e.g. multivariable mixed effects regression).

- Descriptive analyses are always appropriate. Sometimes only descriptive analyses are appropriate.

- Never decide if data should be analysed with parametric or non-parametric methods based on normality tests. Distributional tests check if data have a specific distribution, but for parametric analyses, the relevant assumption is that data are from an *approximately* normal distribution. The decision to use parametric or non-parametric methods should be based on descriptive summaries, visualisations, and knowledge about what the distribution of the outcome is expected to be.

- Be aware of multiple testing. The more tests performed, the greater the risk you will observe a chance result. Collecting data, performing analyses, collecting more data, performing more analyses is multiple testing.

- Analyses should be planned in detail in a Statistical Analysis Plan **before** any analyses take place.

- Performing unplanned analysis can lead to:
    - Misleading conclusions due to bias and multiple testing – choosing analysis methods/options based on them producing favourable results.
    - Wasted effort – analyses should only be performed once, not repeatedly.
    - Lack of analysis boundaries – if the analyses to be conducted are not agreed beforehand, it's easy for a supervisor to ask "but what if we do it this way…?".

- A Statistical Analysis Plan template is available here. It is based on guidance for items required in a SAP for clinical trials, but can be adapted for other types of study.

- Post-hoc power calculations are misleading and should not be performed.

5.3 Interpretation and presentation

- Do not use "statistical significance" language. p-values should be interpreted on a continuous scale as weight of evidence against the relevant null hypothesis, not dichotomised into "significant" or "not significant" around an arbitrary threshold. The following language is more appropriate, but should be adapted to the research context:
    - $p \leq 0.001$: very strong evidence against the null hypothesis
    - $0.001 < p \leq 0.01$: strong evidence against the null hypothesis
    - $0.01 < p \leq 0.05$: (some) evidence against the null hypothesis
    - $0.05 < p \leq 0.1$: weak evidence against the null hypothesis
    - $0.1 < p$: little/no evidence against the null hypothesis.

- It is not always necessary or appropriate to present p-values. 95% Confidence intervals convey more information and should **always** be presented alongside p-values. They are particularly useful for visual comparisons.

- Interpretation of an effect should include consideration of the Minimally Clinically/Biologically Important Difference (MCID). Statistical evidence for an effect is irrelevant if the effect is too small to matter.

- Visualisation of data can make results considerably easier to interpret and is recommended where possible.

5.4 External statistical resources

- A list of free and paid statistical courses and resources from other institutions can be found here.

6. Issues with the use of AI for statistical analyses

Generative AI cannot be relied upon to perform statistical analyses and produce credible results, for numerous reasons:

- Using AI for analyses encourages a 'black box' approach where necessary details of the analysis are not provided or not recorded, or the researcher running the analysis does not understand what is being done.

- Code produced by AI can be difficult to interpret and may contain errors that are difficult to spot.

- Exactly repeating analyses is not possible as code produced and results presented will differ even when the same prompt is used with the same AI model.

- AI is often used without any contextual information for the research being provided, which may lead to it making inappropriate analysis decisions.

- AI may produce *simulated* results from the code it produces, rather than actually running the code.

- AI is known to distort information in multiple ways (https://www.nature.com/articles/s41599-024-03811-x).

- AI tends to focus on "statistical significance" of results, which is language that should not be used.