BPSU Study guidance – Basic epidemiology for BPSU studies

Dr Simon Lenton on behalf of the Scientific Committee of the British Paediatric Surveillance Unit

Updated 07 01 19

BPSU parent bodies:

Public Health England

RCPCH Royal College of Paediatrics and Child Health *Leading the way in Children's Health*

UCL GREAT ORMOND STREET INSTITUTE OF CHILD HEALTH

with support from:

Great Ormond Street Hospital Charity

Contents

Introduction

This paper has been written for 'first time researchers' who want to develop their ideas for surveillance into a BPSU study. It is not intended to be an epidemiology textbook, but focuses on research and epidemiological topics that are important when designing a BPSU study.

Why is this guidance needed?

The intention of this guidance is to enable applicants to write clear and concise submissions which have a high probability of making good progress through the BPSU Scientific Committee.

Significant numbers of study applications take longer to process due to a necessary dialogue between the applicant and the scientific committee, who often have questions or require clarification about methodological issues within the proposed study.

Commonly, the questions revolve around the purpose and value, i.e. the intended benefits of the study; the surveillance and analytical definitions used for the condition; validity of proposed statistical analysis and the outcome measures used at follow-up.

BPSU studies

The unique aspect of BPSU studies is the engagement of almost all paediatricians in the UK in the identification of children with rare conditions or rare complications of common conditions, through its monthly surveillance reporting system.

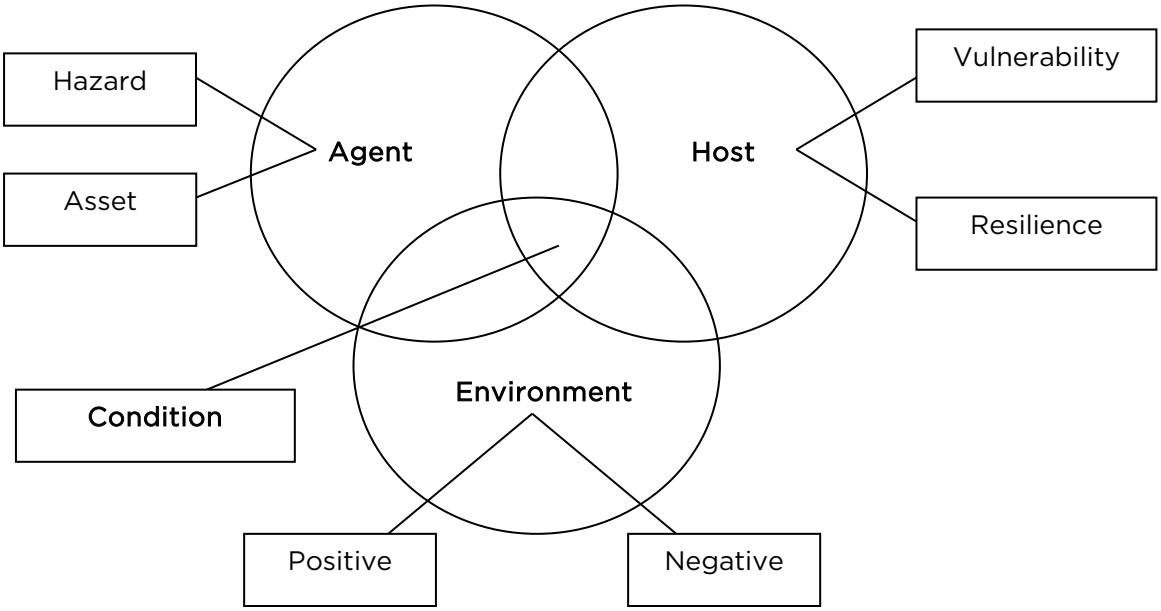
The unique aspect of studying rare conditions and outcomes also brings with it specific limitations to BPSU studies. The majority of studies are designed to analyse no more than 300 children, usually collected over a one-year period, but occasionally longer periods when incidence is low or where temporal trends are important.

Follow-up periods vary, ranging from zero for point/period prevalence studies up to three years, depending upon the outcomes of interest. Longer follow-up using record linkage are currently being explored.

Concepts of disease development

Traditionally public health has used the concepts of an interaction between an agent (causing the problem), the host (experiencing the problem) within the wider physical and social environment, as a model for understanding and then structuring interventions to tackle

health problems. This conceptualisation can be applied to both infectious diseases and non-communicable diseases.



Venn diagram illustrating the interaction of factors contributing to a disease or condition.

A hazard is anything that has the potential to cause harm, for example an infectious disease or exposure to lead. An asset is anything that promotes health or well-being, for example good nutrition or exercise. Vulnerability includes factors that make children less resistant to hazards, whereas resilience protects health. The environment covers both physical and social environments which both can be either positive or negative. BPSU studies may include the collection of factors which represent agent, host, or environmental domains depending on the condition being studied.

Public health surveillance

Health surveillance is generally defined as the ongoing systematic collection and analysis and dissemination of information to support decision-making and action. It is the public health (population-based) equivalent of clinical diagnosis.

Epidemiology

Epidemiology is the systematic study of the host distribution (frequency, patterns), the determinants (causes, risk factors) and the environment in which they occur, to understand the development of

health conditions in specified populations (communities). What then follows is the implementation of evidence-based interventions to control, ameliorate the problem, or improve quality of life.

Epidemiological studies contribute to the understanding and control of diseases/conditions often through a sequential process of investigation including:

1. ecological studies
2. cross-sectional (incidence/prevalence) surveys
3. case–control studies
4. cohort studies
5. experimental/intervention studies

Epidemiological methods, relevant to BPSU studies, can describe the presenting symptoms, clinical features, investigation, management, the natural history, external (environmental) factors or factors (host) related to the child associated with onset and outcomes of the condition. BPSU studies can inform power calculations for later intervention studies.

Occasionally there is the opportunity to evaluate the impact of an intervention when only a proportion of the cohort/study group received the intervention. Proving causality and intervention studies are outside the scope of BPSU studies either because they require explicit consent or information on comparisons with children without the condition (i.e. controls).

Descriptive epidemiology

Descriptive epidemiology covers the questions "who, why, what, where and where?" sometimes distilled down to "time, place and person". BPSU studies generally are only descriptive but often collect associated factors which might be used later to explain causation.

Analytic epidemiology

Case–control studies, cohort studies and experimental/intervention studies all fall under the category of analytic epidemiology because they involve a comparison or control group and the purpose of the study is to identify and quantify the relationship between an exposure to a hazard or asset and a particular health outcome. In the experimental study the investigator controls the exposure, but in the observational study individuals are exposed in natural conditions. Differences in outcomes between exposed and unexposed individuals are analysed to conclude whether exposure is associated with outcomes. BPSU studies rarely involve this type of analysis

because they lack a control group. Causation is a separate issue requiring fulfilling the use of causation criteria.

Triangulation (cross verification)

Because BPSU studies investigate rare conditions and are dependent on paediatrician's recollection there is always the possibility of underreporting. Occasionally conditions are routinely reported through other systems, for example, obstetric reporting systems, hospital episode statistics or pathology systems, especially for conditions with specific ICD-10 codes. If the data is available from other sources it is wise to first check the likely frequency of the condition under study, and the possibility for cross verification of the BPSU dataset with other sources. Indeed if data is available from other sources check whether a BPSU study is required.

Capture-recapture

Capture-recapture is a specific epidemiological method to estimate the size of the population when incomplete data is available from different sources. It is essential to know the size of the population from which each sample has been derived and be able to correctly identify common individuals and then apply appropriate analysis (ecological models; log-linear models and sample coverage approach). If capture-recapture methodology is proposed it must be fully described in the BPSU application, meet all the criteria for capture-recapture analysis, including the requirement that each population sample must be independent of each other.

See Knowles, R. L., Smith, A., Lynn, R., & Rahi, J. S. (2006). Using multiple sources to improve and measure case ascertainment in surveillance studies: 20 years of the British Paediatric Surveillance Unit. *Journal of Public Health*, 28(2), 157-165.

BPSU research design

BPSU studies are based on epidemiological principles and so study design should be no less rigorous than any other academic or research study and the following issues are of particular importance:

Clarity of purpose of the study

Explaining the purpose and objectives behind your study is important and then how it adds value to patient care or public health understanding is a critical first step to a successful study. Generally BPSU studies are good for:

1. conditions that come to the attention of paediatricians Review BPSU studies similar to your proposal that have been successful in the past, this will give you some ideas about outcomes that are practical.
2. the incidence or prevalence of a disease or condition, often with some element of follow-up outcomes over one or two years
3. intervention incidence (e.g. how often a procedure is used) again with appropriate follow-up

Outcomes can be measured in many different ways including direct clinical observation, physiological measures, or the burden of disease from the perspective of either the individual, family or community. Sometimes costs (reflected by use of services) of NHS support can be captured from clinical records. Occasionally there are opportunities for quality assurance, for example, asking whether best practice followed?

See BPSU follow-up assessments paper: https://www.rcpch.ac.uk/sites/default/files/2018-04/bpsu_study_guidance-follow-upassessmentmeasures_230317.pdf

Definition of the condition

Definitions are critical to successful epidemiological studies. Rarely are diagnoses purely categorical (yes or no) and often there is a spectrum of disease and defining the critical cut-off point in plain English using data easily extractable from clinical notes can be a challenge. Using internationally defined criteria is advantageous in order to be able to compare conditions across the globe; however there are times when studying a subgroup of the condition, or a complication of a condition may be more appropriate. Once again, definitions and clarity over severity or the complication to be investigated is essential.

Selection of subjects – the surveillance and analytic criteria

All BPSU studies include a surveillance definition (effectively a reporting definition for probable cases) which is broader than the analytic or case definition for confirmed case to be included in the analysis. The purpose of the surveillance definition is the capture all of the potential cases i.e. to have a high sensitivity, in contrast the analytic definition must exclude those cases which do not meet the study criteria i.e. have a high specificity. Both must be clearly defined as must be the process of selection of "cases" from "reports". If complex definitions are used, this sifting process may require independent reviewers and a clinical algorithm which describes the process of deriving true positives (cases) from false positives

(reports). A description of this process is essential to a successful application.

Defining determinants

Determinants may be derived from the basic model of disease development (see above), for example:

Host

- preterm infant
- immunocompromised child

Agent

- infectious
- non-communicable agent

Environment

- poverty (socio-economic)
- geographical distribution

Each will require clear definitions to enable clinicians to accurately extract the relevant data from clinical notes and these definitions are best included in the questionnaire. It is only worth collecting information that is likely to always be recorded in the clinical notes (or electronic patient record).

Outcome measures

The definitions of outcome measures are equally as important as the definition of the condition or complication itself. Often the limiting factor in BPSU studies is the limited availability of data from routine clinical records. This process is evolving with the development of electronic clinical records and in the future it may be possible include data from other parts of the health service and potentially even education and social care. The development of data linkage opens up the possibilities of measuring outcomes currently not available, for example, using the ASQ (Ages and Stages Questionnaire) at 2 1/2 years of age for neonatal follow-up studies.

Sometimes a pilot study is required to demonstrate that outcome data are available, recorded consistently in clinical notes and that it is feasible to extract this data by paediatricians completing the form.

Data collection and storage

See: https://www.rcpch.ac.uk/sites/default/files/2018-04/4._system_level_security_policy_guide_-_22052017_2.pdf

Data analysis

Sophisticated analysis using multivariable analysis is rarely possible within a sample size of less than 300 (usually the maximum number of the BPSU study). However simple descriptive analysis is usually all that is required. Where more sophisticated statistical analysis is used, justification for the methods and possibly a power calculation will be required in the application.

It is expected that all the data collected will be used. To aid investigators in this, a data analysis plan is required which should clearly indicate, where appropriate, the numerator and the denominator for each derived piece of information, how this information will be analysed and how these analyses map onto the study objectives.

BPSU resources

https://www.rcpch.ac.uk/sites/default/files/1._The_BPSU_Study_Application_Handbook__-_Phase_1_-_20-05-2017__1__0.pdf

Appendix

Definitions

In epidemiology definitions are of critical importance because they determine the numerator and denominator when estimating, for example, incidence and prevalence, or, rates and proportions. (Remember that the numerator is the upper portion of the fraction and the denominator is the lower portion of the fraction).

The relationship between prevalence and incidence

There are two main measures of disease frequency, which are prevalence and incidence and then there are further subdivisions within each. The proportion of the population that has a disease at a point in time (prevalence) and the rate of occurrence of new disease during a period of time (incidence) are closely related.

Prevalence depends on:

1. The incidence rate (I)
2. The duration of disease (T)

and the relationship between incidence and prevalence can be expressed as

$$P = I \times D$$

where P=Prevalence, I=Incidence Rate and D=Average duration of the disease.

For example, if the incidence of a condition is low but the duration of disease (i.e. time until recovery or death) is long, the prevalence will be high relative to the incidence. An example of this would be diabetes.

Conversely, if the incidence of a disease is high and the duration of the disease is short, the prevalence will be low relative to the incidence. An example of this would be influenza.

Prevalence

Prevalence is a useful measure of the burden of disease in a population at a given point in time, for example, when planning health services.

Point prevalence measures the frequency of condition in a defined population **at a single point in time**.

$$\text{Point Prevalence} = \frac{\text{No. of cases in a defined population at one point in time}}{\text{No. of persons in a defined population at the same point in time}}$$

Period prevalence is the number of individuals identified as cases *during a specified period of time*, divided by the total number of people at risk of the condition in that population (generally the midpoint population number). The majority of BPSU studies which consider prevalence, collected over a year, so therefore determine an annual period prevalence.

Incidence

Incidence is a measurement of the number of new individuals who develop a condition during a particular period of time and is expressed as a percentage, or, if small, as per 1000 persons. It is particularly useful when new conditions are emerging or established conditions are changing, for example HIV incidence or CMV. Incidence can also be used for outcomes for example how many children die (the incidence of death also known as the mortality rate) from a defined condition during a specified time period.

There are two main measures of incidence:

Cumulative incidence (sometimes called incidence risk)

Cumulative incidence refers to the occurrence of events, such as disease or death in a group studied over time. It is the proportion of individuals in a population initially free of disease or condition but who develop it within a specified time interval.

$$\text{Incidence Risk} = \frac{\text{Number of new cases of disease in a specified period of time}}{\text{Number of disease-free persons at the beginning of that time period}}$$

It is important to note that the denominator is the total number of people who were free of disease at the start of the study period. This is defined as the population at risk. The incidence risk assumes that the entire population at risk at the beginning of the study period has been followed for the specified time period for the development of the outcome under investigation. An example would be number developing HIV in the first year of life (the numerator) with the denominator being number of live births.

Incidence rate

This measurement that seeks to account for varying time periods of follow up.

Incidence rates also measure the frequency of new cases of disease in a population, but takes into account the sum of the time that each participant remained under observation and at risk of developing the outcome under investigation.

$$\text{Incidence rate} = \frac{\text{Number of new cases of disease in a given time period}}{\text{Total person-time at risk during the follow-up period}}$$

Summary

Measure	Numerator	Denominator
Incidence proportion	Number of new cases of condition during specified time interval	Population at start of time interval
Incidence rate	Number of new cases of condition during specified time interval	Average population during time interval
Point prevalence	Number of current cases (new and pre-existing) at a specified point in time	Population at the same specified point in time
Period prevalence	Number of current cases (new and pre-existing) over a specified period of time	Average or mid-interval population

The relationship between rates, proportions and ratios

Rates. A rate is a measure of the frequency with which an event occurs in a defined population in a defined time (e.g. number of cases per hundred thousand population in one year).

Proportion. The number of people with a condition divided by the total population that does not have the condition.

Ratios. The value obtained by dividing one quantity by another, for example the male to female ratio. A ratio often compares two rates (the 'rate ratio'), for example comparing death rates for girls and boys at a given age.

The important difference between a rate and a ratio is that for a rate, the numerator is included in the denominator e.g., number of new cases of a condition divided by the total population. In a ratio, the numerator and denominator are usually separate and distinct quantities, neither being included in the other e.g., the ratio of males to females in the class.

A risk ratio (RR), also called relative risk, compares the risk of a health event (disease, injury, risk factor, or death) among one group with the risk among another group. It does so by dividing the risk (incidence proportion, attack rate) in group 1 by the risk (incidence proportion, attack rate) in group.

References

https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/numerators-denominators-populations

https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/section1.html

PDF textbook can be downloaded here:
https://www.cdc.gov/ophss/csels/dsepd/ss1978/SS1978.pdf

Knowles, R. L., Smith, A., Lynn, R., & Rahi, J. S. (2006). Using multiple sources to improve and measure case ascertainment in surveillance studies: 20 years of the British Paediatric Surveillance Unit. *Journal of Public Health*, *28*(2), 157-165.
https://academic.oup.com/jpubhealth/article/28/2/157/1569751