

# PUBH5215 Introductory Analysis of Linked Data

## Elective

**Credit points:** 6

**Semester:** 1&2 Late Intensive

**Delivery mode:** Block Mode

**Pre-requisites:** PUBH5018 and (PUBH5010 or BSTA5011) and (PUBH5211 or BSTA5004)

**Co-requisites:** Nil

**Unit coordinator:** Professor Judy Simpson, School of Public Health

### Summary

This unit introduces the very specialised topic of the analysis of linked health data. The types of study and research questions that linked data can be used to address will be discussed, including how to construct study populations, the special limitations and pitfalls associated with linked data, and methods of checking data and results. Students use de-identified extracts of linked files of routinely collected NSW administrative health data to learn how to manipulate the data in SAS to prepare them for analysis. The computing component of the unit assumes a basic familiarity with computing syntax used in SAS.

### Aim

This unit introduces the topic of linked health data analysis at an introductory to intermediate level. It fills a gap in research training opportunities by combining the principles of health care epidemiology with hands-on practical exercises in the implementation of computing solutions. The modular structure of the unit provides students with a theoretical grounding in the classroom on each topic, followed by a training session on the corresponding computing solutions. Students use de-identified linked NSW data files in the hands-on exercises. The computing component of the unit assumes a basic familiarity with computing syntax used in SAS and methods of basic statistical analysis of fixed-format data files.

### Learning objectives

Upon completion the participant will be able to:

- understand the theory of data linkage methods and features of comprehensive data linkage systems, sufficient to know the sources and limitations of linked health data sets, and in particular those for NSW;
- apply epidemiological principles to the design of studies using linked data;
- construct numerators and denominators for the analysis of disease trends and health care utilisation and outcomes;
- assess the accuracy and reliability of data sources;
- check data linkages and assure the quality of the study process, e.g. consistency of definitions, missing data;
- list the steps required to perform the manipulation of large linked data files;
- write syntax to prepare linked data files for analysis, derive exposure and outcome variables, relate numerators and denominators and produce results from statistical procedures.

## Content

Contents include: an overview of the theory of data linkage methods and features of comprehensive data linkage systems, sufficient to understand the sources and limitations of linked health data sets; construction of numerators and denominators used in the analysis of disease trends and health care utilisation and outcomes; basic statistical analyses of linked longitudinal health data; conceptualisation of manipulation of large linked data files; writing syntax to prepare linked data files for analysis, derive exposure and outcome variables, relate numerators and denominators and produce results from statistical procedures at an introductory to intermediate level.

## Instructional format

Block/intensive mode 5 days 9am-5pm

## Assessment

eWorkbook of the issues that arise, how they were resolved, and the key learnings, as you work through the daily exercises (30%)

1x assignment (70%)

Note: The assignment involves the analysis of (altered) linked health data sets provided by NSW Health to answer certain research questions. Due to the Confidentiality Agreement to be signed by each student, this assignment can only be done in the School of Public Health computing lab. It is due 10 days after the end of the unit.

## Learning materials

Required/ recommended/ prescribed texts:

Notes will be distributed in class and will be available via eLearning

## Notes

Students considering enrolling in this unit are advised to listen to the [soundbite](#).