

# DATA201

## Week 12 – Lecture 1

What is Good Data Practice

Previous Lecture(s)

# Good data Practice

- What is good data practice?
- How do we do good data practice?

# The ASA Ethical Guidelines for Statistical Practice

- Many different frameworks, guidelines, charters, developed in various countries to guide ethical use of data, algorithms and statistics in their country (once you've seen one, you've seen them all...).
- American data Association (ASA) Guidelines developed to help data practitioners to make ethical decisions when formulating, collecting, analysing, modelling, and interpreting data and outputs.
  - Currently - 59 recommendations + 12 for Organisations and Institutions
  - Eight (plus one) different categories
  - Guidelines can be found at: <https://www.amstat.org/your-career/ethical-guidelines-for-data-practice>
- Note: I have made a slight change in the ASA recommendations to say “data practitioners” rather than “statistical practitioner”.

# The ASA Ethical Guidelines for Data Practice

- A. Professional integrity and accountability
- B. Integrity of data and methods
- C. Responsibilities to stakeholders
- D. Responsibilities to research subjects, data subjects, or those directly impacted by data studies
- E. Responsibilities to multidisciplinary teams
- F. Responsibilities to fellow data practitioners and the profession(s)
- G. Responsibilities of leaders, supervisors and mentors in data practice
- H. Responsibilities regarding professional misconduct
- I. Responsibilities of organisations/institutions

# A. Professional Integrity and Accountability

Professional integrity and accountability require taking responsibility for one's work. Ethical data practice supports valid and prudent decision-making with appropriate methodology. The ethical data practitioner represents their capabilities and activities honestly and treats others with respect.

This includes:

- Uses data and methodology that is valid, relevant, and appropriate
- Does not knowingly conduct practices that exploits or perpetuates biases
- Opposes efforts to predetermine or influence results
- Promotes the fair treatment of all people

## B. Integrity of Data and Methods

The ethical data practitioner seeks to understand and mitigate known or suspected limitations, defects, or biases in the data or methods and communicates potential impacts on the interpretation, conclusions, recommendations, decisions, or other results of data practices.

This includes:

- Communicates data sources, collection processes, and data processing and transformation procedures, including missing data handling
- Transparent about assumption made in the execution and interpretation of outputs
- Communicates the purpose and intended use of data practices

# C. Responsibility to Stakeholders

Those who fund, contribute to, use, or are affected by data practices are considered stakeholders. The ethical data practitioner respects the interests of stakeholders while practicing in compliance with these guidelines.

This includes:

- Seeks to establish what stakeholders hope to obtain from a specific project
- Does not use practices to mislead stakeholders
- Uses practices that are appropriate to exploratory and confirmatory phases of a project
- Protects the use and disclosure of data, including conforming to confidentiality requirements for data, data outputs, and the reuse of data



## D. Responsibility to Research/Data Subjects, or Those Directly Affected by Data Practice

The ethical data practitioner does not misuse or condone the misuse of data. They protect and respect the rights and interests of human and animal subjects. These responsibilities extend to those who will be directly affected by data practices.

This includes:

- Adhering to applicable rules and guidelines for the protection and welfare of human and animal subjects
- Makes informed decisions regarding sample size and methodology to avoid inadequate or excessive numbers
- Protect subjects' privacy, and collect data where consent has been granted
- Considers the impact the research can have on individuals, groups, and society at large

## E. Responsibilities to Members of Multidisciplinary Teams

Research and data practice is often conducted in teams made up of professionals with different professional standards. The data practitioner must know how to work ethically in this environment.

This includes:

- Respecting other professions' ethical standards and processes
- Where ethical guidelines conflict, upholds ethical data practices where data is concerned
- Promote transparency of all data practices
- Regardless of pressure within research teams, upholds good data practices

## F. Responsibilities to Fellow Practitioners and the Profession

Data practices occur in a wide range of contexts. Irrespective of job title and training, those who practice statistics have a responsibility to treat data practitioners, and the profession, with respect. Responsibilities to other practitioners and the profession include honest communication and engagement that can strengthen the work of others and the profession.

This includes:

- Recognising different practitioners may come to different conclusions, so hash out differences with constructive dialogue
- Help strengthen and not undermine other's work through peer-review processes
- Promotes reproducibility and replication of their work where possible

# G. Responsibilities of Leaders and Mentors

Practitioners leading, supervising, and/or mentoring people in data practice have specific obligations to follow and promote these ethical guidelines. Their support for, and insistence on, ethical data practice are essential for the integrity of the practice and profession of statistics, as well as the practitioners themselves.

This includes:

- Advocate for a working environment that supports responsible data practice
- Promote a safe and inclusive working environment
- Identify and create opportunities for team members to develop professionally and maintain their proficiency

# H. Responsibilities Regarding Potential Misconduct

The ethical data practitioner understands that questions may arise concerning potential misconduct related to data, scientific, or professional practice. At times, a practitioner may accuse someone of misconduct or be accused by others. At other times, a practitioner may be involved in the investigation of others' behavior. Allegations of misconduct may arise within different institutions with different standards and potentially different outcomes. The elements that follow relate specifically to allegations of data, scientific, and professional misconduct.

This includes:

- Knows the definitions and procedures relating to misconduct
- Avoids condoning data, scientific, or professional misconduct
- Develops fair and transparent processes and procedures for misconduct

# I. Responsibilities of Organisations/Institutions

Whenever organizations and institutions design the collection of, summarize, process, analyze, interpret, or present data or develop and/or deploy models or algorithms, they have responsibilities to use data practice in ways that are consistent with these guidelines, as well as to promote ethical data practice.

This includes:

- Supporting practices that are objective and transparent
- Avoid practices that exploit vulnerable communities
- Objectively, accurately and efficiently communicating a team's or practitioners work through the organisation

# Conclusions

- Good data practice is done through integrity, and an awareness of responsibility to others
- For guidance on what integrity and responsibility is (in the context of data), ethical frameworks can give good guidance
- These are generally not exhaustive – in Aotearoa we would like to combine the ASA guidelines with I/MDSov if we are working particularly with Māori data