# INTRO TO CLINICAL DATA STUDY GUIDE

## MODULE 7 –CLINICAL DATA ETHICS

### INTRODUCTION TO RESEARCH ETHICS AND AI

Any discussion of the role of artificial intelligence in healthcare must carefully distinguish the *development* of models, algorithms, predictive tools, and so on, from the routine uses of those tools in medical *practice*, whether as a decision aid, or a screening tool.

- **Research**
  - Development of new AI tools, models and algorithms
  - An activity that is usually structured and has as its primary goal, the production of generalizable knowledge
- **Clinical practice**
  - Application of validated or accepted AI tools, models and algorithms into different aspects of standard clinical practice or operations
  - An activity that aims to benefit individual patients. The purpose of the activity is to help make a diagnosis, prevent disease, or provide treatment to individual patients

**History of Research Ethics**

- 1947 Nuremberg Code in response to Nazi research atrocities
- 1932-1972 Tuskegee Syphilis Study
  - African American men were denied antibiotics and misled into thinking that lumbar punctures and other data collection measures were treatments for “bad blood”
- 1966 Henry Beecher NEJM article
  - Demonstrated that most published clinical trials had made no attempt to obtain informed consent from research participants
- 1964 Declaration of Helsinki
  - First major update since Nuremberg Code and included a requirement for independent review of formal research protocols
- 1978-9 The Belmont Report
  - 1974, US National Research Act created a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, the final report becoming “the Belmont Report”

    - Significant overlap with Helsinki, though the Belmont Report included a very influential ethical framework that provided a foundation for those recommendations
- 1981 45 CFR 46 "The Common Rule" adopted

**Belmont Report Ethical Framework**

1. **Respect for Persons**: To respect the autonomy of agents who have the capacity to make their own decisions. Protect individuals with diminished capacity
2. **Beneficence**: A fundamental principle of medical ethics that has been widely seen as applicable to biomedical research: primum non nocere or "do no harm". Possible benefits should be maximized and possible harms should be minimized as much as possible
3. **Justice**: A fair distribution of the benefits and burdens of research.

Together, these principles and their application are the keys to understanding the requirements for the ethical development of new AI tools.

## THE BELMONT REPORT: A FRAMEWORK FOR RESEARCH ETHICS

The three principles that are articulated in the Belmont report are the framework for the regulatory system governing human subjects research and have been enormously influential in understanding the ethics of research on human subjects. Research to develop new AI tools or to develop new knowledge or findings from human data will fall under these regulations and should be broadly consistent with this ethical framework.

### 1. RESPECT FOR PERSONS

Respect the autonomy of agents who have the capacity to make their own decisions

#### Terminology

- **Autonomy**: The ability of individuals to self-legislate in light of reason
- **Autonomy (ethics)**: Recognition of the right of individuals to make choices that reflect their values and interests
- **Autonomy (research)**: Research participants should make an informed decision about whether they want to participate in the activity

Informed Consent:

- Requires that the individuals be capable of providing consent
- Provide enough information to make an informed choice about whether to participate
- Free of coercion

The Belmont report (and the regulations in the U.S. that were based on them) does explicitly allow research to take place without consent, even for individuals who have capacity. The Belmont report recognized that sometimes, the requirement that potential participants provide informed consent could impair the validity of the research itself.

Waiving the Normal Informed Consent Requirements:

1. Incomplete or non-disclosure is necessary for the research goals to be achieved
2. Risks of the research are minimal
3. Later plan for dissemination of information about the research
4. No other rights of the individual will be violated or other harms to the participant will take place as a result of the research

These requirements are particularly important for the data collection required to develop AI.

Respect for persons also required that individuals with diminished autonomy be respected.

## 2. BENEFICENCE

Minimize harms and maximize benefits to participants of research.

Risks to research participants must be reasonable in relation to anticipated benefits of the research, including the value of the generalizable knowledge to be gained.

## 3. JUSTICE

Fair distribution of risks and benefits of research:

- Aristotle's formal principle of justice: treat similar cases similarly
- Distributive fairness
- Procedural fairness

Fair procedures could still produce unfair distributive outcomes as a result of "social, racial, sexual, and cultural biases in society".

There are many critics of the Belmont Report, most importantly that it does not provide any guidance about how to deal with trade-offs between these principles, since there is no rank ordering among them. However, it is a critical starting point that governs how Institutional Review Boards (IRB)/Research Ethics Committees (REC) evaluate research.

## ETHICAL ISSUES IN DATA SOURCES FOR AI

**Example of AI Research:**

- Researchers use deep learning to try to identify and understand novel pharmacogenomic variants
- Novel algorithms and models have been developed to better predict which patients will benefit from aggressive palliative radiation
- Tools have been developed to improve identification of tumors in radiological images
- Tools have been developed to improve health insurance claims processing

In order for new tools to be developed, AI requires data for it to learn from. AI has often been linked to the concepts of Big Data and Precision Health. The vision of creating more precisely tailored interventions for individual patients requires that large amounts of data from different sources be used to understand why patients respond differently to the same medications or have different side effect profiles.

**Types of Data for AI Research:**

1. Research repositories
    a. NIH "All of Us Research Program"
    b. UK "100,000 Genomes" project
    c. Geisinger biobank
2. Secondary uses of data collected for other purposes
    a. Electronic health records (EHR) data from patient encounters
    b. Insurance claims data
    c. Newborn blood spots
    d. Census data
3. Consumer data collected outside of healthcare
    a. Wearables
    b. Mobile health
    c. Direct to consumer (DTC) genetics

d. Computer usage data
e. None of the oversight or regulatory mechanisms that govern apply

### Research Repositories

- Data collected through interactions with a person engaged in an activity for the purpose of research
- Consent is normally required and obtained
- Informed consent for each use of the data is not usually possible
- Broad consent is an option

### Requirements for Broad Consent – Participants need to be told:

- Of any known risks
- Of any anticipated benefits
- Details about how confidentiality is maintained
- Participation is voluntary
- If genetic analysis will be done
- Whether data may be used to generate profit and whether shared

### Research Repository Issues

- Informed consent
- Security efforts to ensure privacy as much as possible
- Justice
    - Research repository recruits have tended to be middle class or affluent individuals of European ancestry
    - Fair procedural process may still result in inequality in outcome as a result of historical mistrust

Creating fair access to the results of AI research will require great effort to repair trust in under-represented populations, to build relationships between under-represented communities and researchers, and to succeed in improved recruitment and engagement. There is some promising improvement in some of these efforts, but it remains an open question whether research repositories can ever be truly representative of the public.

## SECONDARY USES OF DATA

A great deal of AI research makes use of data that was collected for purposes other than research; EHR data, Insurance claims data, Newborn blood spots, Census data.

Obtaining informed consent for the secondary uses of data collected for other purposes is often (though not always) impracticable, particularly if it is required to meet regulatory requirements for informed consent for research on human subjects.

### Pathways to Use Without Consent:

- If the "research" activity is regarded as "Quality Assurance (QA) and Improvement"
    - Not considered research under regulations
    - QA typically does not require consent
    - Limited guidance about how to determine whether something is "QA" rather than research
        - Rule of thumb: are findings generalizable to other institutions?
    - Note: intent to publish does not make the activity research
- Secondary use of de-identified data
    - HIPAA standards of de-identification
        - Expert determination standard
        - Safe harbor standard: Strip 18 identifiers
          "does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is the subject of the information"
    - NIH sees genomic information as identifiable; researchers continue to find new ways to identify individuals in databases that were previously regarded as de-identified
    - The idea that de-identification is sufficient to remove the duty to obtain consent remains ethically controversial
- Waiver of consent

### Waiver of Consent Requirements:

1. The research must present minimal risk to participants
2. The research could not be carried out successfully if consent was required
3. The waiver of consent will not adversely affect the rights or welfare of the participants
4. Afterwards, the participant will be informed of their participation if it is appropriate

"If the research involves using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format"

**Ethically Controversial:**

- Concern by many in bioethics that unconsented data use will undermine trust and demonstrates a lack of respect
- Surveys shows that potential research participants want to provide consent for uses of their data
- Surveys also demonstrate high levels of support for research uses of their data
- Less studied is how potential participants view the trade-off between these two preferences
- Requiring consent will likely lead to fewer participants, less data, and more biased data
- Pronounced tendency for consented data in research repositories to leave out minority populations, the poor and those in rural areas
- Are there alternative (and possibly better) ways of demonstrating respect for participants

## RETURN OF RESULTS

AI research on samples has the potential to discover information about individuals in the data.

Some research on data specifies that no results will be returned:

- Sharp boundary between research and clinical practice means no duty of care

Raised cautionary flags about the idea of "Therapeutic Misconception"

- "**Therapeutic Misconception**": The confusion between research and practice by participants

**Ethical challenges to this approach:**

1. Participants often want information
2. Ancillary duty to warn may still exist
3. Grounded in Beneficence (maximize benefit)

**Permissive Approach:**

If participants want any information that is discovered about them in the course of research on their samples or data, they have a right to that information and therefore any findings should be offered.

- Return all information that is discovered about them that is linked to them

    - Grounded in respect for patients and their autonomy
    - Individuals differ in what information they consider important or relevant
- Challenges
    - Findings may not be valid or even interpretable
    - Some findings will not be actionable
    - Poorly defined, unvalidated information could cause harm and fails to minimize risks
    - Potential wide range of information makes consent about risks impossible

**Range of Intermediate Positions:**

Intermediate positions between these two extremes may be more plausible, but there are a range of views about precisely where to draw the line.

- Duty to return at least some results - maximize benefits
    - Analytically and clinically valid
    - Significant medical impact
    - Creates opportunity to take action to treat, mitigate or avoid impact
- Findings that should not be returned - minimize harms
    - Uncertainty or lack of interpretation
    - Lack of validity
    - Risk of harm
- Grey area in between
- Variation exists in practice about what to include in each category as well as what to do about grey area

There is a tradeoff between respecting the autonomy of participants (or parents who provide permission for their children) and the obligation to minimize harm to participants, which is grounded in Beneficence. There remains a challenging balancing between these two considerations leading to variation in practice about the best place to draw the line on what should be returned.

For findings that are neither required to be returned nor impermissible to be returned, there is a grey area with variation in practice. Researchers would not be required to return these results, though it would be permissible to do so.

The issue is somewhat more complicated in the context of secondary research uses of de-identified clinical or public health data, if there has not been any consent. There is arguably a heightened duty to return actionable results because the information was collected in a clinical context, rather than a research context. The practical challenges of actually returning information when people are unaware of the research being done with their data are quite difficult. This problem would be much more

manageable if patients were more aware of how their data can be used to improve quality, as well as to conduct research to improve healthcare.

## AI AND THE LEARNING HEALTH SYSTEM

A potential solution to the problems presented by the regulations can be found in the concept of the Learning Health System. AI research, and in particular the data collection needed for research, presents a difficult tradeoff between fully informed consent and the ability to produce valid, representative, scientifically and medically important research.

For secondary uses of data collected for clinical or public purposes, it can sometimes be a challenge to obtain meaningful consent at all. For uses of data in research repositories, it will typically not be possible to obtain consent for each use of the data, requiring a broad consent rather than anything approximating fully informed consent. The fact that findings from this research may sometimes be returned makes the consent issues even more complicated.

### Challenges to Distinction Between Research and Practice

The starting point is the distinction between research and practice. Some research is not aimed at trying to develop new drugs or interventions, but instead aims to improve different aspects of medical practice.

- Quality Improvement Research: Goal is to improve patient care

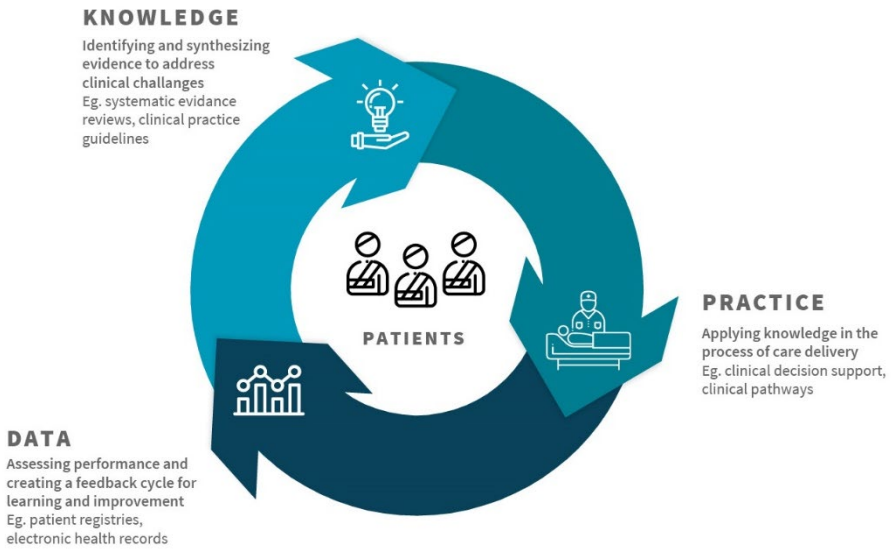
#### Terminology

- **ROMP:** Research on Medical Practices
- When an institution makes a commitment to systematically carry out ROMP using all of its data and clinical care practices as a way of improving care, it is sometimes called "the Learning Health Care System."

- Comparative Effectiveness Research: Goal is to determine which interventions within the range of usual care is best
- Precision Health: Goal is to determine which interventions within the range of usual care is best for particular cohorts of patients (or individuals)

### The Learning Health System:

- Institute of Medicine (now the National Academy of Medicine) defined LHS as "one in which knowledge generation is so embedded into the core of the practice of medicine that it is a natural outgrowth and product of the healthcare delivery process and leads to continual improvement in care"
- Challenges the traditional line drawn between clinical practice and research
- Every patient is a research participant whose data is an opportunity to learn
- Learning improves patient care
- Requires setting up systems to learn from patient data and to translate learning to improved patient care
- AI will be key to both feedback loops




- AI is likely to play a critical role, both in the learning process from the data and then in the development of methods to translate new knowledge into altered practice

## Learning Healthcare System Ethics Framework

Critics of the current research ethics paradigm derived from the Belmont Report have developed an alternative ethical framework for the Learning Health System. Their framework is based on recognition of a set of obligations or duties held by different stakeholders in the learning health system.

- Respect for Persons: This is captured in the LHS by the duty to respect the rights and dignity of patients
- Beneficence: This is captured in the LHS in three ways
    - Obligation to provide optimal care for each patient

    - Duty to avoid imposing nonclinical risks and burdens
    - Obligation to improve clinical care through learning
- Justice: The duty to address and reduce unjust inequalities
- Obligation to respect clinician judgment
- Patient obligation: Duty to contribute to knowledge, particularly when there is little risk or effort involved. Pass on the benefits that they themselves have received from contributions of others

### Remaining challenge

This framework does not offer details about how to balance the tradeoffs that exist between the obligations:

- How to balance duties and obligations (such as consent as a way of demonstrating respect) with duties of justice and beneficence
- Is a form offering broad consent for minimal risk activities that improve patient care the best way to demonstrate consent?
- Patients and participants want to give consent, but they also want the fruits of research
- Large majority of patients and the public were willing to accept less elaborated (or no) consent if necessary for research to take place

### Alternative Measures to Demonstrate Respect

Faden and Kass have suggested that there may be better ways of demonstrating respect for patient/research participants than the standard informed consent forms.

- Active engagement with patients about ongoing learning activities
- Transparency about the types of learning taking place
- Accountability in translating the results to improve patient care

It is worth noting that something like this has already taken place in regard to individual rather than systems learning:

### Analogous Learning at Individual Level

- Teaching hospitals require learning for training individual clinicians
- Patients required to receive some of their care from trainees
- Benefit is care at teaching hospital
- No consent option (required)

- Fairly broad public expectation
- Transparency about fact
- Accountability to minimize risks and to ensure adequate training

### The Learning Health System

The research regulations and the Belmont Report are flexible enough to accommodate this system as long as IRB's or Research Ethics Committees are willing to utilize appropriate mechanisms, such as waiver or alteration of consent, when it is appropriate.

- Analogous to learning at individual level
- Systems level learning a similar duty
- Greater potential benefit to participants than individual learning