Computing for Medicine
Quiz - 1

Time - 60 minutes

Date - 28th August 2025

Name:

Max. Marks - 100

4*15 = 60

Multiple Choice Questions (MCQs)

1. Two hospitals exchange allergy information for a patient using a FHIR standard syntax. One system interprets "penicillin allergy" as a drug intolerance instead of an allergy. What is the most likely issue?
   a) Human failure
   b) Semantic interoperability failure
   c) Process Interoperability failure
   d) Organizational failure

2. Your friend is a medical intern who thinks that an electronic health record (EHR) is just a scanned PDF of medical reports. What will you tell them to correct their thinking?
   a) EHR is paperless storage
   b) EHR is a repository in computer-processable form with a standardized model
   c) EHR is a vendor's private software for health data
   d) EHR is a software for administrative billing

3. Your clinician friend has shared a dataset using a shared Google Drive link named "data_final_use_this.xlsx". You see that
   I. No DOI or persistent identifier is given.
   II. Metadata is missing.
   III. The link often breaks when ownership changes.
   Which FAIR principle is MOST violated here?
   a) Findable
   b) Accessible
   c) Interoperable
   d) Reusable

4. You are doing a study that needs to identify patients with diabetes from data of two hospitals. One hospital's electronic health record system HR uses ICD-10 codes, another uses SNOMED CT. What technique is essential before data integration?
   a) Data encryption
   b) Ontology mapping
   c) De-identification
   d) Natural language processing

5. Simpson's paradox is a phenomenon where the overall group effects are opposite to that of subgroups. This is caused due to:

a. Data duplication
b. Confounding variables
c. Measurement error
d. Random noise

6. A retrospective study design differs from a prospective study design because it looks at data that were collected prior to initiation of the study. What is NOT true as a limitation of this design?

a) It suffers from a high risk of selection bias and information bias due to reliance on existing records
b) Limited control over the quality and completeness of the collected data
c) It is more expensive as compared to prospective studies
d) It suffers from difficulty in establishing temporal relationships between exposure and outcome

7. Which of the following is an example of confounding?

a. Coffee drinking is associated with lung cancer because coffee drinkers are more likely to smoke.
b. A new vaccine reduces infection rates after randomization.
c. A measurement device consistently overestimates blood pressure.
d. Patients are lost to follow-up in a trial.

8. Which is NOT true of human physiological systems? They

a) exhibit self-organizing behavior
b) are highly optimized
c) have non-linear properties
d) are adaptive

9. In a DAG representing Smoking → Lung Cancer ← Asbestos Exposure, the variable Asbestos Exposure is:

a) A parent node
b) A child node
c) A confounder
d) An exposure

10. A radiology AI system predicts tumor presence with high accuracy on training scans. But in practice, its performance drops because new scans are stored in a different file format and resolution. This is an example of:

a) Interoperability challenge
b) Data leakage
c) Simpson's paradox
d) Random noise

11. Which of the following is an example of structured healthcare data?

a) X-ray images
b) Doctor's narrative notes
c) Blood pressure readings
d) MRI scans

12. A policymaker sees a rise in obesity rates alongside increased smartphone use and concludes that Smartphones cause obesity. This is an error due to:
a) bad data
b) random sampling
c) overfitting
d) confounding

13. We think about interoperability to enable high-fidelity exchange of data, meaning, workflow, and action. Workflows are a part of:
a) Semantic Interoperability
b) Human Interoperability
c) Syntactic Interoperability
d) Process Interoperability

14. Which of the following best explains why EHRs are better than paper records?
a) They are cheaper to maintain
b) They allow structured storage, interoperability, and analytics
c) They require no security measures
d) They eliminate the need for doctors

15. What is the biggest challenge in using unstructured healthcare data?
a) It is always inaccurate
b) It lacks a predefined organization
c) It cannot be digitized
d) It is not useful for AI

Short- Answer Questions **10*4 = 40**

1. What is an ontology(1)? How is it different from a simple data dictionary(1). Explain with an example and create a small ontology graph depicting your understanding of any four medical terms using is-a, has-a process.
2. SNOMED CT concepts are more than just terms. They carry certain essential properties as discussed in class. List two essential properties of a SNOMED CT concept. (2) What do you understand by Fully Specified Name?(2)
3. Phrases such as "Possible fracture of arm" and "Patient is recovering well" are not valid SNOMED CT concepts. Why? Use your imagination to break it down using Post-coordinated valid concepts (need not be actual SNOMED-CT concepts).
4. You are building a predictive model for ICU patients. The data available include:

- blood pressure readings
- X-ray, CT and MRI images
- Text notes capturing patient history, progression and treatment

a) Classify each of the above as structured, semi-structured, or unstructured.(1)

b) Which type of data is most directly usable for traditional statistical models, and why? (1)

c) What will be your approach to use unstructured data (e.g., notes) for predictive modeling?(2)