

# Introduction to Health Information Analysis

## Introduction to Health Information Analysis

Key concepts, steps, and tools for turning health data into actionable insight.

### 1. What Is Health Information Analysis?

| Aspect        | Description                                                                                                                                                                          |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Definition    | Systematic examination of health-related data (clinical, administrative, public health, or patient-generated) to discover patterns, assess performance, and support decision-making. |
| Purpose       | <ul style="list-style-type: none"><li>Improve patient outcomes</li><li>Optimize resource use</li><li>Detect disease trends</li><li>Inform policy and research</li></ul>              |
| Typical Users | Clinicians, epidemiologists, health system managers, public health officials, researchers, insurers, and policymakers.                                                               |

### 2. Core Data Sources

| Source                               | Examples                                        | Typical Format                         |
|--------------------------------------|-------------------------------------------------|----------------------------------------|
| Electronic Health Records (EHRs)     | Diagnosis codes, lab results, medication orders | Structured (HL7/FHIR), semi-structured |
| Claims & Billing Data                | CPT/HCPCS codes, reimbursement amounts          | Structured (CSV, SQL)                  |
| Public Health Registries             | Cancer registries, immunization logs            | Structured, often CSV/JSON             |
| Patient-Generated Data               | Wearables, mobile app logs, surveys             | Time-series, JSON, CSV                 |
| Research Databases                   | Clinical trial repositories, genomics datasets  | Structured, high-dimensional           |
| Social Determinants of Health (SDOH) | ZIP code, income, education level               | Structured or linked external datasets |

# Introduction to Health Information Analysis

**Tip:** Verify data provenance and licensing before analysis.

---

## 3. The Analytical Workflow

### 1. Define the Question

- Clinical (e.g., “Which patients are at risk of readmission?”)
- Operational (e.g., “What is the average length of stay for cardiac surgery?”)
- Public?health (e.g., “Is there a seasonal spike in influenza cases?”)

### 2. Data Acquisition & Integration

- Extract from source systems (APIs, ETL pipelines).
- Merge datasets using patient identifiers, encounter IDs, or geographic keys.
- Apply **record linkage** techniques when IDs differ.

### 3. Data Cleaning & Preparation

- **Validate:** check ranges, data types, mandatory fields.
- **De?duplicate** records.
- **Standardize** coding systems (ICD?10, SNOMED CT, LOINC, RxNorm).
- **Handle missing values** (imputation, exclusion, flagging).
- **Create derived variables** (e.g., age groups, comorbidity scores).

### 4. Exploratory Data Analysis (EDA)

- Summary statistics (mean, median, prevalence).
- Visualizations: histograms, box plots, heatmaps, geographic maps.
- Correlation matrices and outlier detection.

### 5. Modeling & Inferential Analysis

- **Descriptive:** rates, incidence, prevalence.
- **Predictive:** logistic regression, random forests, gradient boosting, neural networks.
- **Causal:** propensity?score matching, instrumental variables, difference?in?differences.
- **Time?Series:** ARIMA, Prophet, survival analysis.

# Introduction to Health Information Analysis

## 6. Validation & Performance Assessment

- Split data (train/validation/test) or use cross-validation.
- Metrics: AUC, ROC, sensitivity, specificity, calibration plots, RMSE, etc.
- Conduct external validation when possible.

## 7. Interpretation & Communication

- Translate statistical results into clinical or policy implications.
- Use visual