BP801T · B.Pharm Semester VIII · Theory Notes

Ethical Considerations and Translational Applications of AI in Pharmacy

Unit IV — AI in Public Health & Real-World Data Analytics (6 hours)

Unit IV (6 hours)

Pills & Python Notes

From-zero teaching standard: every technical term is defined before use.

Syllabus coverage: Unit IV (6 hours).

PILLS & PYTHON

Pharmacy • Data • AI • Healthcare

Instagram @pillsandpython · pillsandpython.com · YouTube @PillsAndPython

CHAPTER 10: AI in Public Health & Real-World Data Analytics

1. Learning Objectives

After studying this chapter, you will be able to:

- Define **real-world data (RWD)** and explain how it differs from clinical-trial data.
- List the main sources of RWD: electronic health records (EHR), insurance claims, disease surveillance systems, disease registries, and wearables.
- Explain the concept of **outbreak prediction** using early-warning signals from RWD.
- Describe **population risk modeling** and **risk stratification** in simple terms.
- Explain how **regression** is used in epidemiology to study disease risk factors.
- Explain **vaccination forecasting** — predicting vaccine demand for immunization drives.
- Identify **data-quality and bias problems** in real-world data.
- Describe Indian examples: the **Integrated Disease Surveillance Programme (IDSP)**, the **CoWIN** platform experience, and the **Ayushman Bharat Digital Mission (ABDM)**.
- Connect public-health AI to the pharmacist's role in surveillance, reporting, and vaccine services.

2. Definitions

Read these first. Every later section uses these exact meanings.

- **Public health:** The science and practice of protecting the health of whole populations and communities (not just one patient at a time). Example: vaccination drives, sanitation, disease control.
- **Clinical trial:** A carefully planned experiment in which a small, selected group of patients is treated and observed under strict rules to test whether a drug works and is safe.
- **Real-world data (RWD):** Information about health collected during routine, everyday medical care and daily life — not inside a controlled experiment. Example: prescriptions written in a hospital, insurance claims, temperature reported by a wristwatch-like wearable device.
- **Clinical-trial data:** Information collected inside a clinical trial under strict rules, from a selected group of patients, in a fixed time period.
- **Electronic health record (EHR):** A digital version of a patient's medical chart — diagnoses, prescriptions, lab results, visit notes — stored in a hospital or clinic computer system.
- **Insurance claims data:** Records created when a hospital or pharmacy asks an insurance company (or a government health scheme) to pay a bill. They record which services and medicines were billed.
- **Disease surveillance:** Continuous, systematic watching of disease patterns in a population — collecting, checking, analysing, and sharing data so that action can be taken. Think of it as a "watchtower" for disease.

- **Disease registry:** A organised list/database of patients who share one condition (for example, a cancer registry or a tuberculosis registry), kept so that treatment outcomes and trends can be studied.
- **Wearable (wearable device):** A small electronic device worn on the body — smartwatch, fitness band, pulse sensor — that measures things like steps, heart rate, and sleep.
- **Outbreak prediction:** Using patterns in data to estimate that a disease outbreak (a sudden rise in cases) is likely to happen soon.
- **Early-warning signal:** A small change in data that appears before a big change — like dark clouds before rain. Example: rising fever-related medicine sales before an official flu spike is reported.
- **Population risk modeling:** Using mathematics to estimate which groups of people are most likely to get a disease, so help can be sent to them first.
- **Risk stratification:** Dividing a population into layers (strata) by risk level — high risk, medium risk, low risk — so that each layer gets the right level of attention.
- **Regression:** A mathematical method that finds the relationship between one outcome (e.g., "has diabetes or not") and one or more possible causes or risk factors (e.g., age, weight). It answers: "If this factor goes up by one step, how much does the outcome change?"
- **Epidemiology:** The study of how diseases spread in populations, who is affected, and why. An epidemiologist is like a detective for disease patterns.
- **Vaccination forecasting (demand forecasting):** Predicting how many vaccine doses will be needed, where, and when — so stock is ready before people arrive.
- **Data quality:** How correct, complete, timely, and consistent data are.
- **Bias (data bias):** A systematic (one-sided) error in data — for example, data that only represents rich city people and misses poor rural people — which makes AI results unfair or wrong.
- **Integrated Disease Surveillance Programme (IDSP):** India's national programme for disease surveillance — collecting reports of outbreaks and unusual illness from health facilities across the country.
- **CoWIN:** India's digital platform used for COVID-19 vaccination registration, scheduling, certification, and stock management.
- **Ayushman Bharat Digital Mission (ABDM):** India's national mission to create digital health infrastructure — unique health IDs, registries of facilities and health professionals, and consent-based sharing of digital health records.

3. Terminology Table

Term	Short meaning	Think of it as
RWD	Everyday health data from routine care	A diary written while living daily life
Clinical-trial data	Data from a strict experiment	A lab report from a controlled test
EHR	Digital patient medical chart	A digital notebook of every doctor visit
Insurance claims	Bills sent to insurers/schemes	A receipt that also says what was treated
Surveillance	Continuous disease watching	A watchtower scanning for disease

Term	Short meaning	Think of it as
Registry	Database of patients with one condition	A register of one disease, like a school attendance book for diabetes
Wearable	Body-worn sensor device	A personal health watchman
Early-warning signal	Small change before a big change	Dark clouds before rain
Risk stratification	Dividing people by risk level	Sorting passengers by ticket class — different care per class
Regression	Math to find "factor → outcome" relation	A weighing machine that tells which factor pushes most
Epidemiology	Science of disease spread in populations	Disease detective work
Bias	One-sided error in data	A scale that always shows +1 kg
IDSP	India's disease surveillance programme	India's disease watchtower network
CoWIN	India's COVID-19 vaccination platform	The digital ticket counter for vaccination
ABDM	India's digital health mission	A digital highway for health records

4. Explanation (from absolute zero)

4.1 What is public health?

When a doctor treats **one** patient in a clinic, that is individual care. **Public health** looks at **thousands or lakhs of people together** and asks: "How do we keep this whole community healthy?" Vaccination camps, clean-water projects, and controlling a disease outbreak are public-health activities.

Public health needs **data** — numbers and facts about who is falling sick, where, and when. Without data, officials are guessing. With data, they can act early and precisely.

4.2 What is real-world data (RWD)?

Imagine two ways of collecting health information:

Way 1 — Clinical trial (a controlled experiment). Scientists choose 500 carefully selected patients, give them a new drug, watch them for 2 years under strict rules, and record everything. The data are clean and reliable — but they come from **only 500 special people**, in one hospital, under perfect conditions.

Way 2 — Real-world data. Millions of ordinary people visit doctors, get prescriptions, are admitted to hospitals, buy medicines, wear smartwatches. Every one of these actions leaves a small digital footprint. Put together, these footprints form **real-world data**: health information collected during **routine, everyday life and care**, not inside a planned experiment.

Key point: clinical-trial data tell us "does this drug work in ideal conditions?" RWD tells us "what is actually happening in real life?"

4.3 The five main sources of RWD

1. Electronic health records (EHR). When you visit a hospital, the doctor writes your diagnosis, orders lab tests, and prescribes medicines — all in the computer. This digital chart is the EHR. When millions of such charts are studied together (with names removed), they reveal patterns: which diseases are rising, which medicines cause side effects in real patients.

2. Insurance claims data. When a hospital bills an insurance company or a government scheme (like Ayushman Bharat), it must write codes for the disease treated and medicines given. These claim records are tidy and standardised, and they cover huge populations — useful for studying treatment costs and disease frequency.

3. Disease surveillance systems. These are dedicated networks — like India's IDSP — where clinics and labs regularly report cases of specific diseases (fever with rash, acute diarrhoea, etc.). They are built for one job: catch outbreaks early.

4. Disease registries. A registry is a long-running database of patients with one specific condition — e.g., a tuberculosis registry or a cancer registry. It follows the same patients over years, so researchers can study long-term outcomes.

5. Wearables and sensors. Smartwatches and fitness bands continuously measure heart rate, steps, and sleep. This is data from **healthy daily life**, not hospitals — it can catch early signs of illness before a person even visits a doctor.

4.4 Outbreak prediction — how AI can warn us early

An **outbreak** is a sudden rise in disease cases. Waiting until hospitals are full is too late. AI can look at many data streams together and spot **early-warning signals**:

- Sudden rise in fever-related medicine sales at pharmacies
- More people searching for "high fever remedies" online
- A cluster of school absences in one area
- Surveillance reports of similar symptoms from several clinics

Each signal alone is weak; together, an AI model can estimate: "There is a high chance of an outbreak in this district in the next 1–2 weeks." That early warning lets officials stock medicines, alert doctors, and prepare hospitals **before** the wave hits.

4.5 Population risk modeling and risk stratification

A district has 10 lakh people but only 500 health workers. Who should be visited first? **Population risk modeling** uses data (age, existing diseases, past hospital visits) to estimate each group's chance of getting seriously ill. **Risk stratification** then sorts people into layers:

- **High-risk layer:** elderly with diabetes → home visits, priority vaccination
- **Medium-risk layer:** adults with one risk factor → regular reminders
- **Low-risk layer:** healthy young adults → general awareness messages

Resources go where risk is highest. This is the heart of "smart public health."

4.6 Regression in epidemiology — finding risk factors

Regression is the main mathematical tool for asking: "Which factors raise disease risk, and by how much?" Example: researchers collect data on 1,000 people — age, body weight, smoking, and whether they have diabetes. A regression model studies all these together and reports something like: "Smoking raises the odds of diabetes by 2 times, even after accounting for age and weight." This helps doctors know **which risk factors to target** — stop smoking, control weight.

4.7 Vaccination forecasting

A vaccine drive needs the right number of doses, at the right place, at the right time. Too few doses → people turned away. Too many → vaccines expire and are wasted. **Vaccination forecasting** predicts demand: the AI model looks at population size, past vaccination rates, festival seasons (people travel), and disease trends, and estimates how many doses each centre will need each week. Pharmacists involved in cold-chain and vaccine stores directly use these forecasts.

4.8 Data-quality and bias cautions

RWD is powerful but messy:

- **Missing data:** A patient visited three hospitals; records exist at only two.
- **Wrong or delayed entry:** A diagnosis typed incorrectly, or reported weeks late.
- **Not everyone is represented:** Wearable data comes mostly from wealthy city users — models trained on it may work poorly for poor or rural populations. This is **bias**.
- **Correlation is not causation:** RWD shows patterns ("people who carry umbrellas get wet less") but cannot always prove cause. A rigorous study design is still needed.

Rule of thumb: **AI on bad data produces bad decisions faster.** Data quality checks and bias checks are not optional — they are part of ethical AI.

4.9 AI in pharmacovigilance — finding side effects hidden in RWD

Pharmacovigilance means watching medicines for side effects after they are sold. Clinical trials are too small and too short to catch rare side effects — a trial of 3,000 people cannot detect a reaction that happens in 1 in 50,000 users. But RWD covers crores of people. AI can scan EHR and claims data for **signals**: for example, "patients who started Medicine X this month are visiting doctors for liver complaints 3 times more often than similar patients who did not take X." Such a signal does not *prove* the medicine caused it — but it tells drug-safety experts exactly where to investigate. Pharmacists' ADR reports are the highest-quality seeds of these signals, because a pharmacist's report already links a specific drug to a specific suspected reaction with clinical judgement attached.

5. Classifications

5.1 Classification of RWD sources

```
REAL-WORLD DATA SOURCES
├── 1. Clinical data (EHR)
│   ├── Diagnoses, prescriptions, lab results, discharge summaries
│   └── Strength: rich medical detail
├── 2. Administrative/financial data (insurance claims)
│   ├── Billing codes, service dates, medicine costs
│   └── Strength: standardised, large populations
├── 3. Public-health surveillance data
│   ├── IDSP reports, lab confirmations, outbreak alerts
│   └── Strength: built for early warning
├── 4. Disease registries
│   ├── TB registry, cancer registry (long-term follow-up)
│   └── Strength: long-term outcome tracking
└── 5. Patient-generated data
    ├── Wearables (heart rate, steps, sleep), symptom apps, surveys
    └── Strength: continuous, real-life behaviour
```

5.2 Classification of public-health AI uses

Category	What it does	Example
Outbreak detection & prediction	Spots or predicts disease rises	Early warning for dengue in a district
Risk stratification	Sorts population by risk	Priority list for vaccination
Resource/demand forecasting	Predicts needed supplies	Vaccine dose demand per centre
Risk-factor analysis (regression)	Finds which factors raise disease risk	Smoking → heart disease link
Pharmacovigilance signal detection	Finds drug side effects in real data	Rare side effect of a new drug appearing in EHR
Behavioural/health-messaging targeting	Predicts who needs which health message	Hesitant groups needing counselling

5.3 Classification of RWD problems

```
DATA PROBLEMS IN RWD
├── Quality problems
│   ├── Missing (incomplete records)
│   ├── Wrong (entry errors)
│   └── Late (delayed reporting)
├── Representation problems (bias)
│   ├── Some groups missing from data (rural, poor, elderly)
│   └── Data collectors' habits differ across hospitals
└── Interpretation problems
    ├── Correlation mistaken for cause
    └── Hidden factors (confounders) not recorded
```

6. Mechanisms

6.1 How an outbreak-prediction model works (step by step)

1. **Collect data streams:** surveillance reports, pharmacy medicine sales, weather, search trends, hospital admissions.
2. **Clean and align:** remove errors, put everything on the same weekly time scale for each district.
3. **Create early-warning signals:** e.g., "fever medicine sales this week vs. last 4 weeks' average."
4. **Train a model:** feed several past years of data to a supervised ML model; the "label" is whether an outbreak followed within 2 weeks.
5. **Set an alert threshold:** e.g., model score above 0.8 → district flagged as "watch."
6. **Human review and action:** an epidemiologist reviews the flag; officials stock medicines and alert clinics.

Key takeaway: The model does not "see the future" — it recognises patterns that, in the past, appeared just before outbreaks.

6.2 How regression finds a risk factor (mechanism in simple steps)

1. **Gather data:** for each person, record outcome (e.g., diabetes yes/no) and possible factors (age, weight, smoking).

2 **Feed to regression:** the method looks at all factors at the same time.

3 **It calculates a "weight" for each factor:** how much the outcome changes when that factor changes by one step, while the others are held steady.

4 **Result:** "Smoking raises diabetes odds 2-fold, after adjusting for age and weight." Now we know smoking matters independently — not just because smokers happen to be older.

7. Diagrams / Flowcharts

Diagram 10A — From raw RWD to a public-health action

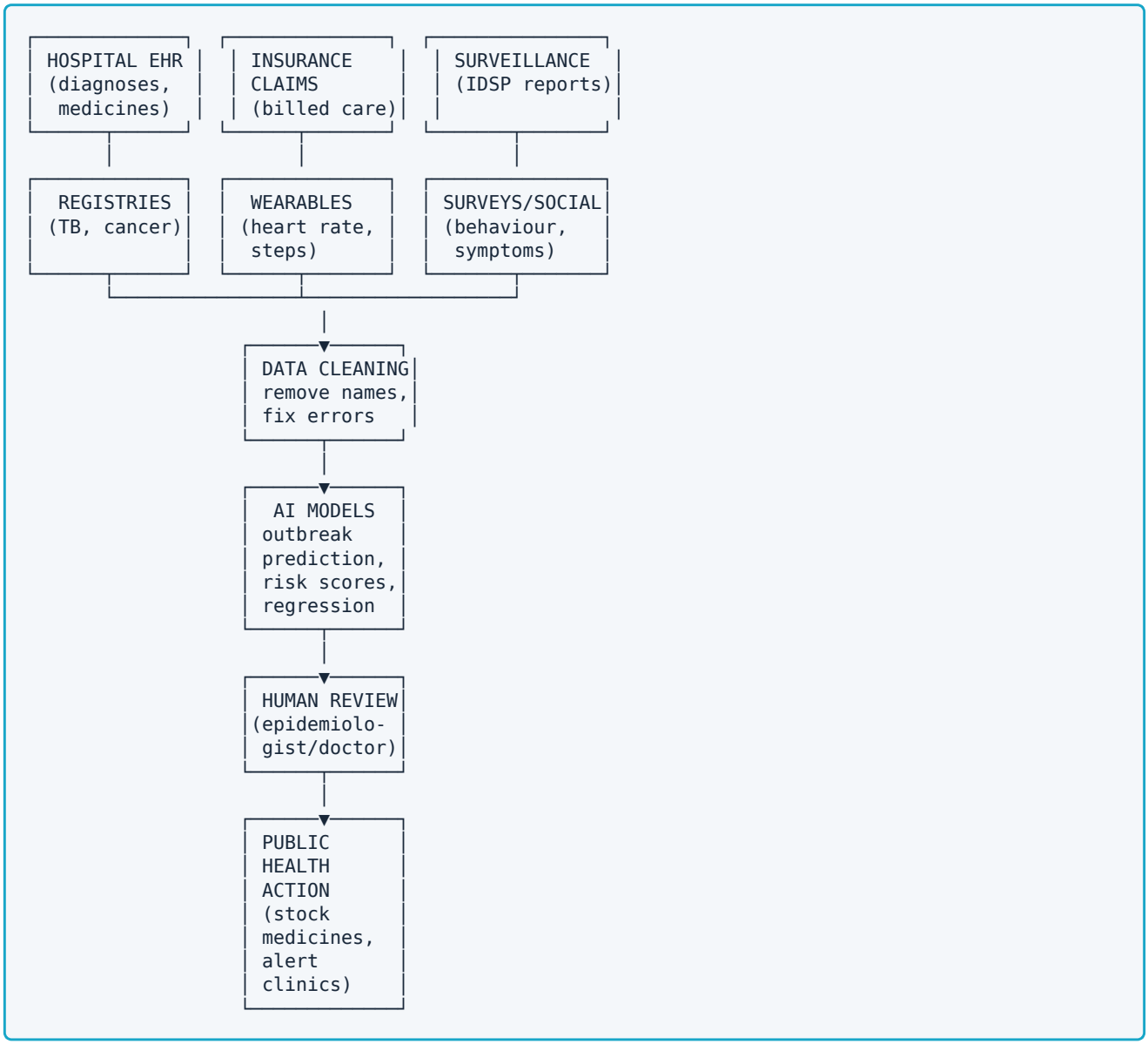
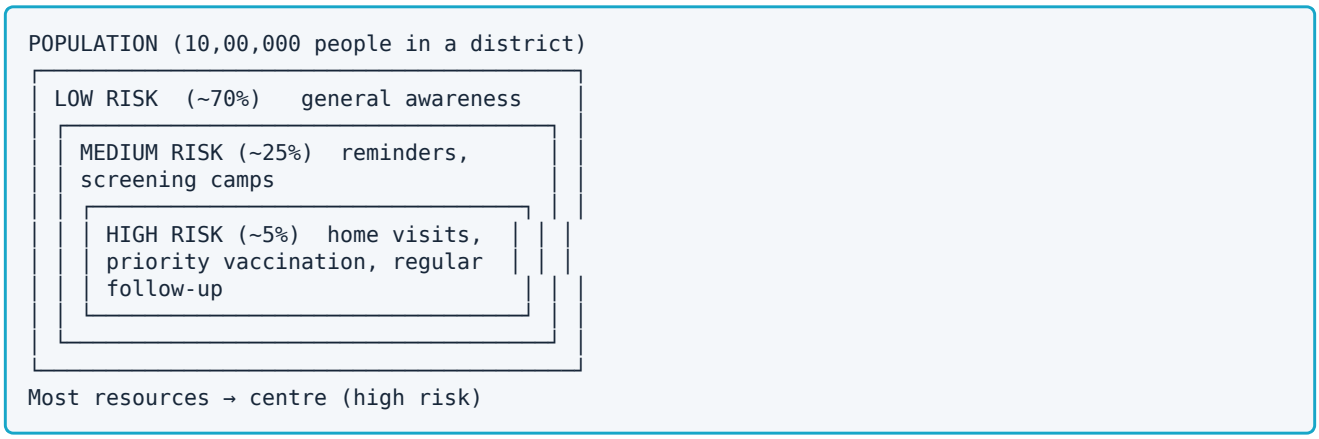


Diagram 10B — Risk stratification layers



8. Infographic Concepts

Infographic 1 — "Two worlds of data." Left side: a small tidy laboratory with 500 selected patients labelled "Clinical trial — perfect conditions, small group." Right side: a busy Indian market-street scene with a hospital, pharmacy, insurance office, and a person wearing a smartwatch, labelled "RWD — real life, millions of people, messy but powerful." An arrow between them: "Trials ask IF a drug works. RWD asks what is ACTUALLY happening."

Infographic 2 — "The outbreak early-warning clock." A circular clock face where the hand points from "dark clouds" (early signals: medicine sales rise, fevers reported) → "storm arrives" (outbreak). Text under: "AI watches the clouds so health teams can act before the storm."

Infographic 3 — "Five rivers of RWD." Five rivers (EHR, claims, surveillance, registries, wearables) flowing into one big lake labelled "Public-health AI," with small boats on the lake labelled "outbreak alerts," "risk scores," "vaccine forecasts."

9. Tables

Table 10.1 — RWD vs clinical-trial data

Feature	Clinical-trial data	Real-world data (RWD)
Where collected	In a planned experiment	During routine care/daily life
Number of people	Small (hundreds–thousands)	Huge (lakhs–millions)
How patients are chosen	Strictly selected (inclusion/exclusion rules)	Whoever visits the doctor
Follow-up	Fixed, complete	Irregular, often incomplete
Data cleanliness	Very clean	Messy, missing entries
Main question answered	"Does the drug work under ideal conditions?"	"What is happening in real life?"
Example	A 2-year drug trial in one hospital	Prescriptions of 2 crore patients across India
AI use	Testing treatment effect	Outbreak prediction, risk scoring, safety monitoring

Table 10.2 — RWD sources: strengths and weaknesses

Source	Strengths	Weaknesses
EHR	Rich medical detail	Different hospitals use different formats; errors
Insurance claims	Standard codes, huge coverage	Only billed care; misses uninsured patients
Surveillance (IDSP-type)	Built for outbreak catching	Depends on timely reporting by staff
Registries	Long-term follow-up of one disease	Only one disease; expensive to maintain
Wearables	Continuous real-life data	Only wealthy/tech users; accuracy varies

Table 10.4 — Early-warning signals: fast proxies and what they catch

Fast signal (available in days)	What it may catch early	Needs cross-check with
Pharmacy sales of fever/cold medicines	Flu-like outbreaks	Clinic case reports
ORS/antidiarrhoeal sales spike	Water-borne disease outbreaks	Lab confirmations
School absenteeism clusters	Respiratory illness in children	Surveillance reports
Wearable resting-heart-rate rises (population-level)	Community infection waves	Hospital admissions
Search trends for symptom remedies	Emerging public concern	Official case counts

Key takeaway: Fast signals buy *time*; official data buys *certainty*. An early-warning system needs both.

Table 10.3 — Common RWD biases and their fix

Bias	What goes wrong	Simple fix
Representation bias	Data only from big-city hospitals; rural patients missing	Collect from PHCs/rural centres too; report who is missing
Measurement bias	One hospital records "fever" loosely, another strictly	Standard case definitions; train staff
Selection bias	Only insured people appear in claims data	Combine multiple data sources
Reporting delay bias	Outbreak seen late because reports arrive weekly	Digital daily reporting; use fast proxy signals (medicine sales)

10. Worked Toy Examples (all datasets are illustrative toy datasets)

Toy Example 10.1 — Early-warning signals before an outbreak (ILLUSTRATIVE TOY DATASET)

District "Toyapur" reports weekly fever cases and pharmacy fever-medicine sales (numbers are invented for teaching only).

Week	Fever cases reported	Fever-medicine packs sold
1	40	120
2	45	130

Week	Fever cases reported	Fever-medicine packs sold
3	42	125
4	90	300
5	210	620
6	480	1,400

Observation: Medicine sales (300 packs in week 4) jumped **before** the official case count exploded (week 5–6). The early-warning signal = "sales rising 2× above the recent average." An AI model learns: *sales spike → outbreak likely within 2 weeks.* This is exactly how early-warning systems combine "fast" signals (sales) with "slow" official reports.

> **Key takeaway:** Fast signals give time; official data give confirmation. Together they form an early-warning system.

Toy Example 10.2 — Regression to find risk factors (ILLUSTRATIVE TOY DATASET)

A toy survey of 8 people (invented data):

Person	Age	Smoker?	Diabetes?
A	30	No	No
B	32	No	No
C	55	No	Yes
D	58	No	Yes
E	60	Yes	Yes
F	62	Yes	Yes
G	35	Yes	No
H	38	Yes	No

Regression studies all factors together and would conclude (on this toy data): **age is the main driver** — older people have diabetes regardless of smoking; smoking alone (young smokers G, H) does not show an effect here. This shows regression's power: it separates the effect of one factor while holding others steady. *(With real data, thousands of people and careful methods are needed.)*

> **Key takeaway:** Regression answers "which factor matters, after accounting for the others?"

Toy Example 10.3 — Vaccine demand forecast (ILLUSTRATIVE TOY DATASET)

A toy immunisation centre plans next month's doses:

Input	Value
Eligible children in area	5,000
Past coverage rate	80%
Expected dropouts/no-shows	10% of covered
Festival week (lower turnout)	−5% adjustment

Forecast = 5,000 × 0.80 × (1 − 0.10) × (1 − 0.05) = 5,000 × 0.80 = 4,000 → 4,000 × 0.90 = 3,600 → 3,600 × 0.95 = **3,420 doses** needed.

A real forecasting model adds trends, past no-show patterns, and neighbouring districts' data — but the logic is the same: **base population × behaviour patterns = demand.**

Key takeaway: Forecasting turns population numbers and behaviour patterns into an order quantity.

11. Analogies

1. **RWD vs trial data = street food survey vs lab recipe test.** A lab test checks a recipe in perfect conditions; a street survey asks thousands of real eaters what they actually ate and how they felt. Both are useful; they answer different questions.
2. **Early-warning signals = dark clouds before rain.** You cannot stop the rain, but dark clouds tell you to carry an umbrella. Medicine-sales spikes are the dark clouds of an outbreak.
3. **Risk stratification = airport security lanes.** Everyone flies, but lanes differ by need — crew, priority, general. Similarly, high-risk patients get "priority lanes" of care.
4. **Regression = a weighing machine for causes.** Put several possible causes on it, and it tells you which one pushes the scale most.
5. **Surveillance = a watchtower, not a camera.** It does not record everything; it watches for specific dangers and raises the alarm.

12. Pharmacy Relevance

Pharmacists are **front-line data generators and users** in public-health AI:

- **Dispensing data as early-warning signals:** A community pharmacist notices a sudden rush for antipyretics or ORS. Reported systematically, this is an early-warning signal — pharmacists are the "sensors" of the surveillance system.
- **ADR reporting feeds RWD:** Every adverse-drug-reaction report a pharmacist files becomes part of real-world safety data.
- **Vaccination services:** Pharmacists in immunisation centres use demand forecasts to order stock and manage cold chains.
- **Antimicrobial stewardship:** Surveillance of antibiotic sales helps detect irrational use and resistance trends — the pharmacist is the gatekeeper.
- **Health education:** Risk-stratified lists (e.g., high-risk diabetics) help pharmacists target counselling and medication-adherence support.

In short: pharmacists both **produce** RWD (dispensing records, ADR reports, vaccination logs) and **act on** AI outputs (forecasts, risk lists, alerts).

A pharmacist's weekly RWD routine (what this looks like in practice):

- **Monday:** Review weekend dispensing — any unusual spike (e.g., antidiarrhoeals)? Note it for the surveillance register.
- **Wednesday:** File ADR reports for any suspected reactions seen — each report strengthens national safety signals.
- **Friday:** Check the vaccine stock against the demand forecast — indent early if a camp is scheduled.
- **Ongoing:** When a risk list arrives (e.g., "high-risk diabetics due for review"), prioritise counselling calls for those patients.

You do not need to build the AI models — but understanding what feeds them and what their outputs mean makes you an informed, trusted link in the chain.

13. India Context

- **IDSP (Integrated Disease Surveillance Programme):** India's national disease-surveillance network. Health facilities report syndromes (like fever with rash, acute diarrhoea) and lab-confirmed cases upward through districts and states. It is the backbone for outbreak detection — AI models can sit on top of this reporting flow to add prediction.
- **CoWIN platform experience:** During the COVID-19 vaccination drive, India used CoWIN for beneficiary registration, slot booking, vaccination certificates, and stock visibility. It showed how a digital platform can manage a population-scale immunisation programme — including demand visibility that feeds forecasting. (No statistics are quoted here; refer to official programme reports for figures.)
- **ABDM (Ayushman Bharat Digital Mission):** The mission to build India's digital health backbone — unique health IDs (ABHA), registries of health facilities and professionals, and consent-based sharing of health records. As ABDM adoption grows, the volume and quality of Indian RWD (especially EHR data) available for ethical, consent-based public-health AI will grow with it.
- **Pharmacist's Indian context:** In India, community pharmacists are often the first point of contact for fever, diarrhoea, and respiratory symptoms — making pharmacy sales and pharmacist observations especially valuable surveillance inputs. Pharmacists also staff government immunisation sessions and report ADRs through the Pharmacovigilance Programme of India (PvPI).
- **Caution:** India's RWD is uneven — strong in big cities and private chains, thin in rural and informal care. Any public-health AI built on Indian RWD must account for this gap or it will serve cities and ignore villages.

14. Key Points

- RWD = health data from routine care and daily life; clinical-trial data = data from a controlled experiment. They answer different questions.
- Five RWD sources: EHR, insurance claims, surveillance systems, registries, wearables.
- Outbreak prediction uses early-warning signals (fast proxies like medicine sales) plus official reports.
- Risk stratification sorts populations into high/medium/low risk so resources go where needed most.
- Regression finds which risk factors matter, holding other factors steady.

- Vaccination forecasting converts population + behaviour patterns into dose orders.
- RWD is messy: missing, delayed, biased. Bad data → bad AI decisions.
- India: IDSP (surveillance), CoWIN (vaccination platform), ABDM (digital health backbone).
- Pharmacists generate RWD (dispensing, ADR reports, vaccination logs) and act on AI outputs (forecasts, alerts, risk lists).

15. Common Confusions

1. **"RWD replaces clinical trials."** No. RWD shows real-life patterns; only well-designed studies (including trials) prove that a drug causes an effect. They complement each other.
2. **"Prediction means certainty."** No. An outbreak "prediction" is a probability ("70% chance in 2 weeks"), not a guarantee. Human experts must review every alert.
3. **"More data automatically means better AI."** No. More *biased* data just makes confidently wrong models. Representativeness matters more than size.
4. **"A risk score is a diagnosis."** No. A risk score says "this group needs attention"; a doctor makes the diagnosis.
5. **"Surveillance means spying on patients."** No. Surveillance tracks disease patterns with de-identified (name-removed) data for community protection, under legal and ethical rules.

16. Common Mistakes

1. **Using medicine-sales spikes alone as proof of an outbreak** — a sale or discount can also spike sales. Always cross-check with clinical reports.
2. **Forgetting reporting delays** — this week's "low cases" may just be reports not yet submitted.
3. **Ignoring who is missing from the data** — building a model on city-hospital data and applying it to tribal areas.
4. **Treating correlation as causation** — "areas with more pharmacies have more disease" (they also have more people!).
5. **Skipping consent and de-identification** — using patient data for AI without proper approvals and name-removal.

17. Mnemonics

- **Five RWD sources — "Every Indian Should Read Widely"**: **E**HR, **I**nsurance claims, **S**urveillance, **R**egistries, **W**earables.
- **Outbreak response steps — "CLEAN-ACT"**: **C**ollect streams, **L**ink & clean, **E**arly signals, **A**nalyse (model), **N**otify threshold, **A**ssess by expert, **C**ommunicate action, **T**rack results.
- **Risk layers — "High-Medium-Low, Help Most the High"**: resources flow to the highest-risk layer first.
- **Data-quality dangers — "MiD-BiC"**: **Mi**ssing, **D**elayed, **Bi**ased, **C**onfounded.

18. Exam Summary

Real-world data (RWD) is health information collected during routine care and daily life, unlike clinical-trial data collected under strict experimental rules. The five main RWD sources are electronic health records (EHR), insurance claims, disease surveillance systems, disease registries, and wearables/patient-generated data. AI uses RWD for outbreak prediction (early-warning signals such as rising medicine sales before official case spikes), population risk modeling and risk stratification (sorting people into high/medium/low risk layers), regression analysis in epidemiology (finding which risk factors raise disease risk while holding others steady), and vaccination forecasting (predicting dose demand per centre). RWD suffers from quality problems (missing, wrong, delayed data) and biases (representation, measurement, selection bias); correlation must not be mistaken for causation, and consent plus de-identification are mandatory. In India, IDSP provides disease surveillance, CoWIN demonstrated population-scale digital vaccination management, and ABDM is building the digital health record backbone. Pharmacists generate RWD through dispensing records, ADR reports, and vaccination logs, and act on AI outputs such as forecasts, alerts, and risk lists.

19. One-Minute Revision

RWD = real-life health data (not trial data). Sources: **E**HR, **I**nsurance, **S**urveillance, **R**egistries, **W**earables — "Every Indian Should Read Widely." AI uses: outbreak early warning (dark clouds before rain), risk stratification (High-Medium-Low), regression (weighing machine for causes), vaccine forecasting (doses = population × behaviour). Watch out: missing/delayed data, bias (who is missing), correlation ≠ causation. India: IDSP watches disease, CoWIN managed vaccination, ABDM builds digital records. Pharmacists are both data sources and data users.

20. MCQs (with Answer Key)

Q1. Real-world data (RWD) is best defined as: a) Data collected inside a randomised clinical trial b) Health information collected during routine care and daily life c) Data published only in research journals d) Data collected from laboratory animals **Answer: b**

Q2. Which of the following is NOT a main source of RWD? a) Electronic health records b) Insurance claims c) A randomised controlled trial dataset d) Wearable devices **Answer: c**

Q3. An "early-warning signal" in outbreak prediction is: a) A confirmed laboratory diagnosis b) A small data change appearing before a big change, e.g., rising medicine sales c) A government order to close markets d) A published research paper **Answer: b**

Q4. Risk stratification means: a) Diagnosing each patient's disease b) Dividing a population into high, medium, and low risk layers c) Removing risky patients from records d) Randomly selecting people for surveys **Answer: b**

Q5. Regression in epidemiology is mainly used to: a) Count the total number of patients b) Find the relationship between risk factors and a disease outcome c) Replace clinical trials d) Encrypt patient data **Answer: b**

Q6. Vaccination forecasting predicts: a) Which vaccine brand is best b) How many doses are needed, where, and when c) The chemical structure of vaccines d) Patient names for vaccination **Answer: b**

Q7. Which bias occurs when RWD comes only from big-city hospitals and misses rural patients? a) Measurement bias b) Representation bias c) Reporting bias of doctors d) Publication bias **Answer: b**

Q8. IDSP in India stands for: a) Indian Drug Safety Programme b) Integrated Disease Surveillance Programme c) International Disease Study Project d) Indian Diagnostic Services Portal **Answer: b**

Q9. A community pharmacist's sudden rush of antipyretic sales can serve as: a) Proof of an outbreak by itself b) A possible early-warning signal to be cross-checked with clinical reports c) A reason to increase medicine prices d) A clinical trial endpoint **Answer: b**

Q10. "Correlation is not causation" warns that: a) RWD is useless b) Patterns in RWD do not by themselves prove cause and effect c) Correlation can never be calculated d) Clinical trials have no correlation **Answer: b**

21. Viva / Application Questions

1 Define RWD and give two pharmacy-related examples of RWD you generate daily.

2 Explain the difference between clinical-trial data and RWD in two sentences.

3 Name the five RWD sources and one strength of each.

4 What is an early-warning signal? Give one pharmacy-based example.

5 Explain risk stratification to a village health worker in simple words.

6 What does regression tell us that a simple comparison cannot?

7 Walk through the steps of forecasting vaccine demand for one immunisation centre.

8 List three data-quality problems in RWD and one fix for each.

9 Why is representation bias dangerous for public-health AI in India?

10 Describe the roles of IDSP, CoWIN, and ABDM in one line each.

11 A district model flags "outbreak likely" — what should the human expert do before acting?

12 Give one example each of how a pharmacist generates RWD and acts on AI output.

22. References

- World Health Organization. Guidance on ethical and effective use of digital health and AI-related public-health tools — as per the latest guidance, confirm the current edition.
- Government of India. Integrated Disease Surveillance Programme (IDSP) — programme documents and reporting guidelines, as per the latest guidance, confirm the current edition.
- Government of India. CoWIN platform — official platform documentation, as per the latest guidance, confirm the current edition.
- National Health Authority, Government of India. Ayushman Bharat Digital Mission (ABDM) — mission framework documents, as per the latest guidance, confirm the current edition.
- Standard epidemiology and biostatistics textbooks on regression methods and surveillance principles, as per the latest editions — confirm the current edition.

CHAPTER 11: Case Studies in AI for Public Health (Illustrative)

⚠ **IMPORTANT — READ FIRST:** All three case studies in this chapter are **ILLUSTRATIVE CASE STUDIES — toy data, not real findings.** The numbers, district names, and "results" are invented purely for teaching. They show you HOW an analysis would be done, not what any real study found. Never quote these toy results as facts.

1. Learning Objectives

After studying this chapter, you will be able to:

- Walk through a complete AI analysis workflow: problem → data → method → result → interpretation → limitation → ethical caution.
- Explain how survey or social-media data could be modelled to find vaccine-hesitancy segments (illustrative).
- Explain how a toy regression trend could project diabetes prevalence for a district (illustrative).
- Explain how toy time-trend data on antibiotic resistance would be read and interpreted (illustrative).
- List the limitations and ethical cautions of each case study.
- Summarise cross-case lessons about AI in public health.

2. Definitions

- **Case study:** A detailed, worked example of how a problem was (or could be) tackled, shown step by step so you can copy the method.
- **Vaccine hesitancy:** Delay in accepting, or refusal of, vaccines even when vaccination services are available. It is not the same as being anti-vaccine — hesitancy has many reasons (fear of side effects, rumours, mistrust, inconvenience).
- **Hesitancy segment:** A group of people who share similar reasons for hesitating — e.g., "worried mothers," "busy daily-wage workers who cannot spare a day," "rumour-believers."
- **Survey data:** Information collected by asking people questions (in person, by phone, or online).
- **Social-media data:** Public posts, comments, and trends on social platforms, analysed for themes and sentiment (overall mood: positive, negative, neutral).
- **Prevalence:** The fraction of a population that has a condition at a given time. Example: "diabetes prevalence of 12%" means 12 out of every 100 people have diabetes.
- **Trend projection (forecasting a trend):** Extending a past pattern into the future to estimate what comes next — like extending a road's direction on a map.

- **Antibiotic resistance (AMR):** When bacteria change so that antibiotics no longer kill them. A resistance percentage of 40% for a drug-bacteria pair means 40 out of 100 tested samples were not killed by that drug.
- **Sentiment analysis:** An AI technique that reads text and classifies its mood — positive, negative, or neutral.
- **Clustering/segmentation:** An AI technique that groups similar people or responses together without being told the groups in advance.

3. Terminology Table

Term	Short meaning	Think of it as
Case study	Step-by-step worked example	A solved question paper
Vaccine hesitancy	Delay/refusal despite availability	Standing outside an open shop, unsure whether to enter
Hesitancy segment	Group with similar reasons	Customers grouped by why they hesitate
Prevalence	Share of population with a condition now	How full the bus is right now
Trend projection	Extending past pattern forward	Extending a road's direction on a map
AMR	Bacteria surviving antibiotics	Thieves learning to pick a new lock
Sentiment analysis	AI reading mood of text	A thermometer for public mood
Segmentation	Grouping similar people	Sorting mixed lentils into bowls by type

4. Explanation (from absolute zero)

4.1 How to read this chapter — the seven-step analysis workflow

Every case study below follows the same seven steps. Learn this pattern; it is the skeleton of every AI public-health project:

1 **Problem:** What decision needs help? (Be specific.)

2 **Data (toy):** What information would we use? (Here: invented toy data.)

3 **Method (concept):** Which AI/statistical idea applies, and why?

4 **Toy result:** What did the analysis "find" on the toy data?

5 **Interpretation:** What does the result mean in plain words for a health official?

6 **Limitation:** Why might the result be wrong or incomplete?

7 **Ethical caution:** What harm could occur if this were used carelessly?

> **Key takeaway:** A result without limitations and ethics is not a finished analysis — it is a half-told story.

Why the order of the seven steps matters: Beginners often jump straight to data ("we have a spreadsheet, let's model it!"). Professionals start at step 1 — the decision. A model built without a decision

in mind usually answers a question nobody asked. Similarly, interpretation (step 5) must come *before* limitations (step 6): first say what the result seems to mean, then attack your own conclusion with "but what if I'm wrong?" That self-attack is what makes the final recommendation trustworthy. Ethics (step 7) comes last in the list but first in importance — it constrains everything above it: which data you may collect, which groups you may single out, and which actions you may recommend.

4.2 A note on toy data

Real public-health datasets are large, messy, and protected by privacy rules — you cannot paste them into a textbook. So we use **toy datasets**: tiny, invented tables that behave like real data in miniature. They let you practise the *logic* (how a method reads data) without needing real patient records. Every table below is labelled; treat its numbers as classroom props, not facts.

Think of it this way: when you learned titration in the chemistry lab, you first practised with a known, safe solution before handling unknown samples. Toy data is the "known solution" of data science — the answers are simple on purpose, so you can check whether you understood the method. Once the method is clear in miniature, you can imagine it scaled to lakhs of real records handled by professionals under privacy rules.

4.3 Why three different question types?

Public-health AI is not one tool but a toolbox. The three cases were chosen because they cover the three most common question shapes you will meet:

- **"Why?" questions (behaviour):** Why do people hesitate? Why do patients stop medicines? These need segmentation and survey/text analysis.
- **"How many, when?" questions (trends):** How many diabetes cases next year? How many vaccine doses next month? These need trend projection and forecasting.
- **"Is it getting worse?" questions (change over time):** Is resistance rising? Is maternal mortality falling? These need time-trend reading.

If you can recognise which shape a question has, you already know which toolbox drawer to open. That recognition skill is more valuable than memorising any single algorithm's name.

5. Classifications

5.1 Classification of the three case studies

```
THREE ILLUSTRATIVE CASE STUDIES
├── A. Vaccine hesitancy analysis  → BEHAVIOUR question
│   ("why do people hesitate?")
├── B. Diabetes prevalence forecasting → TREND question
│   ("how many cases next year?")
└── C. Antibiotic resistance trend analysis → CHANGE-OVER-TIME question
    ("is resistance rising, and where?")
```

5.2 Classification of methods used across the cases

Method family	What it does	Used in case
Segmentation (clustering)	Finds groups with similar behaviour	A — hesitancy segments
Sentiment/theme analysis	Reads mood and topics in text	A — social-media rumours

Method family	What it does	Used in case
Regression / trend projection	Extends a past trend into the future	B — diabetes prevalence
Time-trend analysis	Reads change across years	C — resistance percentages

6. Mechanisms

6.1 Mechanism — how segmentation finds hesitancy groups (concept)

1 Collect survey answers (age group, main worry, trusted information source).

2 Convert answers to numbers (e.g., worry about side effects: yes = 1, no = 0).

3 The clustering algorithm measures "distance" between people — people with similar answers sit close together.

4 It forms groups (clusters) of nearby people.

5 We inspect each group and give it a plain-language name ("rumour-exposed youth," "time-pressed workers").

6 Each group then gets a **different** message strategy — because one message does not fit all.

> **Key takeaway:** Segmentation does not judge people; it sorts reasons, so that communication can be targeted instead of generic.

6.2 Mechanism — how a trend projection works (concept)

1 Plot past prevalence values against years on a graph.

2 Fit a trend line through the points (regression finds the best-fitting line).

3 Extend the line into future years — the extension is the projection.

4 Add a caution band: projections assume the future behaves like the past.

> **Key takeaway:** A projection is an "if nothing changes" estimate, not a prophecy.

A note on uncertainty bands (concept): Honest forecasts never give a single number — they give a range. Instead of "14.9% in 2027," a professional report says "14.9%, likely between 13.5% and 16.3%." The band widens the further you project, because uncertainty grows with distance. If anyone shows you a forecast without a range, ask for it — a point estimate without uncertainty is a half-truth.

6.3 Mechanism — how resistance time-trends are read (concept)

1 For each year, compute % of tested bacterial samples resistant to each antibiotic.

2 Plot each drug's percentages over years.

3 Read the slope: rising steeply = urgent; flat = stable; falling = improving.

4 Compare drugs and regions; flag drug-bacteria pairs crossing concern thresholds.

Key takeaway: The trend's *direction and speed* matter more than any single year's number.

7. Diagrams / Flowcharts

Diagram 11A — The seven-step workflow (used by all three cases)

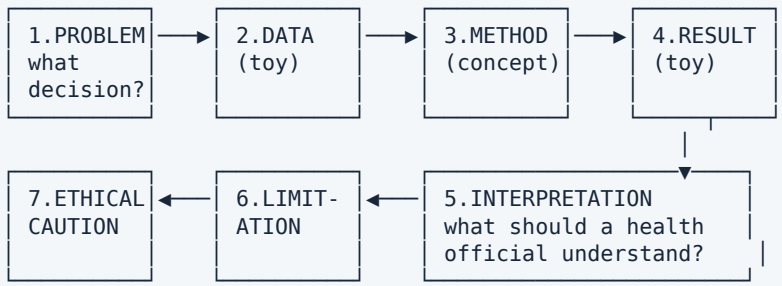
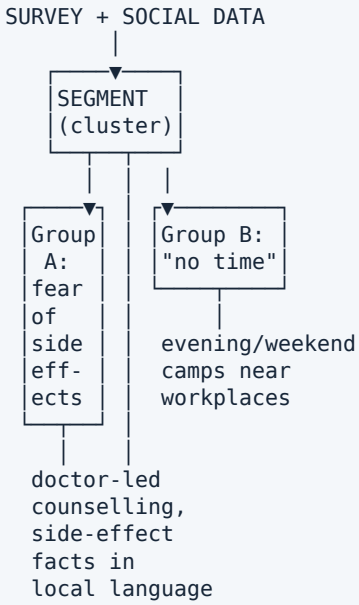


Diagram 11B — From hesitancy segments to targeted messages



8. Infographic Concepts

Infographic 1 — "Three detective stories." A three-panel comic: Panel 1, a detective with a magnifying glass over speech bubbles ("why hesitate?"); Panel 2, a detective extending a rising line on a graph into the future ("how many next year?"); Panel 3, a detective watching coloured bars grow taller year by year ("is resistance rising?"). Caption: "Same workflow, three questions — problem → data → method → result → interpretation → limitation → ethics."

Infographic 2 — "One message does not fit all." A megaphone shouting one generic message at a crowd where people cover their ears, versus three smaller speech bubbles each matched to a group (a worried mother hears a doctor, a worker hears about weekend camps, a student hears myth-busting facts). Caption: "Segment first, then speak."

9. Tables

Table 11.1 — The three case studies at a glance (all illustrative)

	Case A: Hesitancy	Case B: Diabetes forecast	Case C: AMR trends
Question type	Behaviour (why?)	Trend (how many later?)	Change over time (rising?)
Toy data	200-person survey + toy social posts	6 years of toy prevalence	5 years of toy resistance %
Method (concept)	Segmentation + sentiment	Trend-line regression	Time-trend reading
Toy result	3 hesitancy segments	~15% projected in 2 years	Drug X resistance rising fast
Main limitation	Survey may miss the truly hesitant	Assumes future = past	Lab testing patterns change
Main ethical caution	Do not label or shame groups	Do not alarm without context	Do not name-and-shame hospitals

10. Worked Toy Examples — the three case studies

★ ILLUSTRATIVE CASE STUDY A — COVID-19 vaccine hesitancy analysis

(ILLUSTRATIVE CASE STUDY — toy data, not real findings)

Step 1 — Problem. A district health officer wants to raise COVID-19 booster uptake. Posters and announcements are not working. The real question: *WHY* are different people hesitating, so messages can be targeted?

Step 2 — Data (toy). ILLUSTRATIVE TOY DATASET. A toy survey of 200 adults in "Toyapur district" (invented). Each person answered: age group, main reason for hesitating, most trusted information source.

Toy summary (invented counts):

Main reason for hesitating	Count (of 200)	Most trusted source (top answer)
Fear of side effects	70	Family doctor
"No time / centre too far"	60	Employer / community leader
Rumours seen on social media	45	Friends / WhatsApp forwards
"COVID is over, no need"	25	News channels

Toy social-media sample (invented, 100 public posts): sentiment analysis (concept) finds 55% neutral, 30% negative (mostly side-effect rumours), 15% positive.

Step 3 — Method (concept). (i) **Segmentation**: group the 200 respondents by similar answers → hesitancy segments. (ii) **Sentiment/theme analysis** on the toy posts to find which rumours repeat.

How the segmentation would look on the toy data (invented cluster centres):

Segment (toy)	Typical profile (toy)	Share
1. Side-effect worriers	Older, trusts doctor, fear = side effects	~35%

Segment (toy)	Typical profile (toy)	Share
2. Time-pressed	Daily-wage worker, barrier = distance/time	~30%
3. Rumour-exposed	Young, heavy social-media use, repeats one viral rumour	~22%
4. "No need"	Mixed, believes COVID is over	~13%

Step 4 — Toy result. Three clear segments emerge (invented):

- **Segment 1 "Worried about side effects"** (~35%): mostly older adults; trust doctors.
- **Segment 2 "Time-pressed"** (~30%): daily-wage workers; centre distance is the barrier.
- **Segment 3 "Rumour-exposed"** (~22%): young social-media users repeating one viral side-effect rumour; the rest (~13%) say "no need."

Step 5 — Interpretation. One generic poster cannot work: Segment 1 needs doctor-led counselling in the local language; Segment 2 needs evening/weekend camps near workplaces (a logistics fix, not a persuasion fix!); Segment 3 needs targeted myth-busting where the rumour spreads. The officer now has a *plan per segment* instead of one blunt campaign.

Concrete message sketches (illustrative) show what "targeted" means:

- *Segment 1 (side-effect worriers):* a 2-minute talk by a local doctor — "Fever for a day after the shot is your body learning to fight; serious reactions are rare, and we watch you for 30 minutes after vaccination." Honest, calm, no exaggeration.
- *Segment 2 (time-pressed):* no persuasion needed — a notice: "Vaccination camp at the mandi gate, Sunday 8–11 am, 10 minutes, no wage loss." Remove the barrier; the decision follows.
- *Segment 3 (rumour-exposed):* a short video in local language directly addressing the viral rumour with a trusted local figure, shared in the same WhatsApp groups where the rumour travelled. Fight the rumour where it lives.

Notice the pattern: same vaccine, three completely different interventions — counselling, logistics, and counter-messaging. That is the payoff of segmentation.

Step 6 — Limitation. The survey likely missed the most hesitant people (they refuse surveys too) — the toy result may understate hesitancy. Social-media posts come from smartphone users only, missing the poor and elderly. Toy numbers cannot be quoted as real findings.

Step 7 — Ethical caution. Segments must never be used to **label, shame, or coerce** people ("these people are ignorant"). Communication must be respectful, honest about side effects, and voluntary — trust is the goal, not compliance numbers. Also: analysing social-media data needs privacy care — use only public, aggregated, de-identified trends.

★ ILLUSTRATIVE CASE STUDY B — Diabetes prevalence forecasting for a district

(ILLUSTRATIVE CASE STUDY — toy data, not real findings)

Step 1 — Problem. A district plans next year's diabetes screening camps and medicine budget. Question: *roughly how many adults may have diabetes next year*, so staff and strips can be arranged?

Step 2 — Data (toy). ILLUSTRATIVE TOY DATASET. Toy adult diabetes prevalence (%) in "Toyapur district," invented:

Year	2020	2021	2022	2023	2024	2025
Prevalence %	9.0	9.8	10.5	11.6	12.4	13.2

Step 3 — Method (concept). Fit a trend line (simple regression: prevalence vs. year) and extend it 2 years forward.

Step 4 — Toy result. The toy values rise roughly ~0.85 percentage points per year. Extending the line: **2026 ≈ 14.0%, 2027 ≈ 14.9%** (invented projection). If the adult population is 8,00,000 (toy figure), expected cases in 2027 ≈ 8,00,000 × 0.149 ≈ **1,19,200** people needing screening follow-up and possible treatment.

How the arithmetic works (so you can reproduce it): the yearly rises are 0.8, 0.7, 1.1, 0.8, 0.8 — average ≈ 0.84 points per year. Starting from 13.2% in 2025: 2026 ≈ 13.2 + 0.84 ≈ 14.0%; 2027 ≈ 14.0 + 0.84 ≈ 14.9%. A real regression does this fitting automatically and also reports an uncertainty range (e.g., 13.5%–16.3%), which honest forecasts always show.

Step 5 — Interpretation. The health team can now size the programme: screening strips, staff days, and medicine budgets for ~1.2 lakh people rather than guessing. It also justifies prevention spending — the trend is rising, so lifestyle programmes matter.

Step 6 — Limitation. The projection assumes "the future behaves like the past." A new sugar tax, a screening drive that finds hidden cases (raising recorded prevalence without real increase), or changed diagnostic cut-offs would all break the trend. With only 6 toy points, the line is shaky — real forecasts need more years and uncertainty ranges.

Step 7 — Ethical caution. A rising projection must not be used to **panic** the public or to deny anyone care ("too many patients, we cannot treat all"). It is a planning tool. Also, prevalence mixes *diagnosed* and *undiagnosed* — publishing "15% have diabetes" without explaining this can mislead.

★ ILLUSTRATIVE CASE STUDY C — Antibiotic resistance trend analysis

(ILLUSTRATIVE CASE STUDY — toy data, not real findings)

Step 1 — Problem. A hospital's antibiotic committee asks: *is resistance to our commonly used antibiotics rising, and which drug-bacteria pairs need urgent action?*

Step 2 — Data (toy). ILLUSTRATIVE TOY DATASET. Toy lab data: % of *E. coli* samples resistant to three antibiotics in "Toyapur Medical College lab" (invented):

Year	Drug P	Drug Q	Drug R
2021	20%	35%	10%
2022	24%	36%	11%
2023	31%	38%	10%
2024	39%	37%	12%
2025	48%	39%	11%

Step 3 — Method (concept). Time-trend reading: plot each drug's percentages over the years; read direction and speed; flag steep risers.

Step 4 — Toy result. Drug P: 20% → 48% in 5 years — steep, alarming rise. **Drug Q**: high but flat (~35–39%) — stable, still concerning. **Drug R**: low and flat (~10–12%) — currently effective. The committee would flag Drug P for urgent review (prescribing audit, possible guideline change).

Step 5 — Interpretation. In plain words: "Our first-choice drug P is failing in nearly half the samples now — doctors need updated guidance. Drug R still works; protect it by not overusing it." This is how surveillance data becomes a prescribing policy.

A toy committee action plan (illustrative) would read: (i) audit 50 recent Drug P prescriptions to check whether use matched guidelines; (ii) issue an advisory preferring Drug R for uncomplicated cases, with Drug P reserved and justified; (iii) repeat the trend analysis quarterly — if Drug P crosses 60%, escalate to a formal guideline revision; (iv) feed the findings into prescriber CME (continuing medical education) rather than individual warnings. Each action is system-level, measurable, and blame-free — exactly the ethics step 7 demands.

Step 6 — Limitation. Resistance % depends on *which samples get tested* — if only the sickest patients' samples reach the lab, percentages look worse than reality. Testing methods or breakpoints may have changed across years. Toy data, tiny sample — real analysis needs sample counts and lab-quality checks.

Step 7 — Ethical caution. Resistance data must not be used to **name-and-shame** individual doctors or hospitals publicly — that kills honest reporting. Share trends with the committee confidentially, act on systems (guidelines, stewardship), not on blame. Also, restricting an antibiotic has access implications for poor patients — balance is needed.

Cross-case lessons

1. **Same skeleton, different questions.** Behaviour (A), trend (B), change-over-time (C) — all run on the identical seven-step workflow. Learn the workflow once, reuse everywhere.
2. **Toy results teach method, not facts.** Every number above is a classroom prop. The skill you keep is *how to read* such tables, not the numbers themselves.
3. **Interpretation is a human job.** The model outputs numbers; the health officer, pharmacist, or doctor converts them into decisions. AI proposes, humans dispose.
4. **Limitations are part of the answer.** Each case's limitation (who was missed, what assumption was made) is as important as its result.
5. **Ethics is step 7, not an afterthought.** No labelling/shaming, no panic, no blame — design the communication and the action with dignity from the start.
6. **Pharmacists sit at the centre.** Pharmacists counsel the hesitant (A), dispense diabetes medicines whose demand was forecasted (B), and guard antibiotics through stewardship (C).
7. **Start with the decision, not the data.** In all three cases, step 1 names a concrete decision (campaign plan, camp sizing, guideline review). Analyses that start from "we have data, let's find something" usually end as reports nobody reads.

8 **Small, clean, and honest beats big and flashy.** A 200-person toy survey analysed carefully teaches more than a million-row dataset analysed carelessly. In real work, a small representative sample with documented limits is more trustworthy than a giant biased one.

11. Analogies

1 **Seven-step workflow = a doctor's prescription pad format.** Every prescription has the same fields (patient, drug, dose, duration) but different content per patient. Every case study has the same seven steps but different content per question.

2 **Segmentation = sorting a mixed grain bag.** You don't throw the mixture away; you sort it into bowls so each can be cooked properly. Segments let each group's concern be "cooked" with the right message.

3 **Trend projection = extending a road on a map.** You draw where the road is heading — but a new flyover (a policy change) can redirect it.

4 **Resistance trend = watching a thief learn locks.** Drug P's rising bars are like a thief slowly mastering one lock — change the lock (guideline) before it opens every door.

5 **Toy data = flight simulator.** Pilots train on simulators, not real storms. You train on toy data, not real patients.

12. Pharmacy Relevance

- **Case A:** Pharmacists are among the most trusted health voices in communities — ideal counsellors for the "worried about side effects" segment, and reporters of the rumours they hear at the counter.
- **Case B:** Forecasted diabetes numbers become the pharmacist's indent: how much metformin, glimepiride, and testing strips to stock; how many adherence-counselling sessions to plan.
- **Case C:** Antimicrobial stewardship is a pharmacist-led activity — auditing prescriptions, enforcing guidelines, and feeding resistance data back to prescribers. The pharmacist turns trend tables into changed prescribing habits.
- Across all three: pharmacists **generate** the underlying RWD (dispensing, vaccination, culture-sensitivity reports) and **implement** the AI-informed actions.

Pharmacist's action checklist per case (illustrative):

- *Case A (hesitancy):* Keep a small notebook of rumours patients mention → share monthly with the health officer; offer 2-minute side-effect counselling using simple language; never argue — listen first, then inform.
- *Case B (diabetes forecast):* Convert the district projection into your own quarterly indent; plan strip and medicine buffer stock before screening-camp months; schedule adherence calls for high-risk patients the forecast identifies.
- *Case C (AMR trends):* Refuse over-the-counter antibiotic sales without prescription and record attempts (data!); join the hospital antibiotic committee's prescription audit; counsel every antibiotic patient to complete the full course.

13. India Context

- **Case A in India:** Vaccine hesitancy work in India must handle linguistic diversity (messages in the local language), varied trusted sources (doctors, ASHA workers, religious leaders differ by community), and the digital divide (social-media analysis misses non-smartphone users — a large group). CoWIN's scale showed that logistics (slots, centres) matter as much as persuasion — matching the "time-pressed" segment lesson.
- **Case B in India:** India carries a large diabetes burden; district-level projections help state programme managers plan screening under national NCD programmes. But projections built on urban hospital data will mislead for rural districts — representativeness is the Indian challenge.
- **Case C in India:** Antibiotic resistance is a recognised national concern; India has national AMR surveillance networks linking laboratories. Pharmacists' role in curbing over-the-counter antibiotic sales is central to any resistance trend improving. (No statistics quoted — refer to official surveillance reports for figures.)
- **Common thread:** All three cases depend on data systems India is still building (IDSP reporting, ABDM records). The pharmacist's routine records — prescriptions, dispensing logs, vaccination registers — are the grassroots feed for all of them.
- **PvPI linkage:** The Pharmacovigilance Programme of India collects ADR reports from across the country — including from pharmacists. Case C-style trend thinking applies here too: a rising count of a specific suspected reaction is a *signal*, handled with the same read-direction-and-speed discipline, and the same no-blame ethics.
- **ASHA and field workers:** For Cases A and B, ASHA workers and pharmacists together form the last-mile network — one brings community trust and doorstep reach, the other brings medicines, counselling, and records. AI outputs (segment lists, risk lists) are most powerful when handed to this combined network, not left in a district office file.
- **The digital-divide warning for all three cases:** Any Indian analysis built on smartphone, internet, or private-hospital data systematically under-represents the poor, the rural, and the elderly — the very groups public health must prioritise. A mandatory line in every Indian case write-up should be: "Who is missing from this data, and how does that change the conclusion?" If you cannot answer, say so openly.

14. Key Points

- Every case study follows: problem → data (toy) → method (concept) → toy result → interpretation → limitation → ethical caution.
- Case A (hesitancy): segmentation finds *why*-groups; different segments need different messages; never label or shame.
- Case B (diabetes forecast): trend-line extension gives an "if nothing changes" estimate for planning staff and stock; assumptions can break.
- Case C (AMR trends): read direction and speed of resistance %; steep risers need guideline action; don't blame individuals.
- Toy data teaches method, never facts — never quote toy numbers as real findings.

- Cross-case lessons: same workflow for different questions; interpretation is human; limitations belong in the answer; ethics from the start; pharmacists at the centre.

15. Common Confusions

1. **"The case study proves vaccines are hesitated for these reasons."** No — it is illustrative. Real hesitancy reasons come from real, peer-reviewed studies, not this chapter's toy survey.
2. **"The diabetes projection is a prediction of what will happen."** No — it is an "if the past trend continues" estimate. Policies, screening drives, and diagnostic changes can all bend it.
3. **"Rising resistance % means doctors are at fault."** No — resistance is driven by many factors (overuse, incomplete courses, agricultural use, poor sanitation). Blame fixes nothing; stewardship systems do.
4. **"Segmentation stereotypes people."** Only if misused. Done ethically, it groups *reasons*, not worth — and each group gets respectful, tailored help.
5. **"Toy data is useless because it is fake."** Wrong — toy data is a training simulator. It teaches you to read real tables later without risking real patients now.
6. **"If the model is accurate, the project is ethical."** Accuracy and ethics are separate axes. A highly accurate hesitancy-segmentation model used to shame people is accurate *and* unethical. Always score your project on both axes.

16. Common Mistakes

1. **Quoting toy numbers in an exam or presentation as real findings** — always say "illustrative toy example."
2. **Skipping the limitation step** — presenting a forecast without its assumptions is misleading.
3. **Treating a segment label as a judgement** ("rumour-believers are ignorant") instead of a communication guide.
4. **Reading one year's resistance % instead of the trend** — a single point tells you nothing about direction.
5. **Forgetting who is missing** — survey non-responders, non-smartphone users, untested patients.
6. **Acting on the model output without human review** — e.g., auto-restricting an antibiotic on a toy trend.

17. Mnemonics

- **Seven steps — "Please Do Make Real Interpretations, Limits Explicit"**: **P**roblem, **D**ata, **M**ethod, **R**esult, **I**nterpretation, **L**imits, **E**thics.
- **Three cases — "HDT" (Hesitancy, Diabetes, Trends)**: behaviour, trend, change-over-time.
- **Cross-case ethics — "No S.P.B."**: No **S**haming, No **P**anic, No **B**lame.

18. Exam Summary

This chapter presents three ILLUSTRATIVE case studies (toy data, not real findings), each following the seven-step workflow: problem → data → method → result → interpretation → limitation → ethical caution. Case A models COVID-19 vaccine hesitancy: a toy survey plus toy social-media sentiment feed a segmentation analysis yielding three hesitancy segments (side-effect worriers, time-pressed workers, rumour-exposed youth), each needing a different communication strategy; limitations include missing the truly hesitant and smartphone-only social data; the ethical caution is never to label, shame, or coerce. Case B projects diabetes prevalence for a toy district: six years of toy prevalence data fit a trend line projecting ~14–15% in two years, guiding screening and stock planning; the projection assumes the future mirrors the past and must not be used to alarm the public. Case C analyses toy antibiotic-resistance time trends: Drug P's resistance rises steeply (20%→48%), Drug Q stays high but flat, Drug R stays low — prompting guideline review for Drug P; limitations include changing testing patterns; the ethical caution is confidential, system-level action without naming-and-shaming. Cross-case lessons: one workflow fits many questions, toy results teach method not facts, interpretation is human, limitations belong in every answer, ethics comes first, and pharmacists sit at the centre of all three.

19. One-Minute Revision

Three ILLUSTRATIVE cases, one workflow: **P**roblem → **D**ata → **M**ethod → **R**esult → **I**nterpretation → **L**imits → **E**thics ("Please Do Make Real Interpretations, Limits Explicit"). A: hesitancy segments → targeted messages, never shame. B: diabetes trend projection → plan staff/stock, "if nothing changes." C: AMR time trends → steep risers need guideline action, no blame. Toy numbers are classroom props — never quote as facts. No S.P.B.: no Shaming, no Panic, no Blame.

20. MCQs (with Answer Key)

Q1. The seven-step case-study workflow ends with: a) Toy result b) Interpretation c) Ethical caution d) Data collection **Answer: c**

Q2. In Case A, segmentation of the toy survey is used to: a) Diagnose hesitant individuals b) Group people by similar reasons for hesitating c) Count total vaccine doses d) Test vaccine efficacy **Answer: b**

Q3. The "time-pressed" hesitancy segment mainly needs: a) More fear-based posters b) Evening/weekend camps near workplaces — a logistics fix c) A different vaccine brand d) Compulsory vaccination orders **Answer: b**

Q4. A key limitation of the toy hesitancy survey is: a) It used too much mathematics b) The most hesitant people likely did not respond, and social data misses non-smartphone users c) It was conducted in English only d) It had no ethical caution **Answer: b**

Q5. In Case B, the diabetes projection is best described as: a) A guaranteed prediction of the future b) An "if the past trend continues" estimate for planning c) A diagnosis of individuals d) Proof that diabetes is caused by one factor **Answer: b**

Q6. Which assumption could break the Case B projection? a) A new screening drive finding hidden cases b) Using a calculator c) Publishing the result d) Having six years of data **Answer: a**

Q7. In Case C, Drug P's resistance moving 20% → 48% over five toy years means: a) Drug P should be praised b) A steep rise needing urgent guideline review c) Nothing — single numbers don't matter d) The lab made an error **Answer: b**

Q8. Why must resistance trends not be used to name-and-shame doctors publicly? a) Because resistance is never real b) Because blame kills honest reporting; action should be system-level and confidential c) Because pharmacists prefer it d) Because the data is always wrong **Answer: b**

Q9. A cross-case lesson is: a) Each question needs a completely different workflow b) Toy results teach the method, not facts c) Ethics can be added at the end if time permits d) AI outputs need no human review **Answer: b**

Q10. The pharmacist's role across all three cases is: a) Only dispensing medicines b) Generating underlying RWD and implementing AI-informed actions c) Replacing the health officer d) Publishing the toy results **Answer: b**

21. Viva / Application Questions

1 Recite the seven-step case-study workflow and explain why ethics is step 7, not an appendix.

2 In Case A, why is the "time-pressed" segment a logistics problem rather than a persuasion problem?

3 Give one targeted message strategy for each of the three toy hesitancy segments.

4 Why can toy survey numbers never be quoted as real findings?

5 Walk through how the Case B projection (14–15%) was calculated from the toy table.

6 Name two real-world events that could break a diabetes trend projection.

7 In Case C, which drug needs urgent action and why — direction or single value?

8 Explain "no S.P.B." with one example from each case.

9 Design a fourth illustrative case study (e.g., medicine demand forecasting) using the seven steps — outline only.

10 As a pharmacist, what would you do differently at your counter after studying Case A? Case C?

22. References

- World Health Organization. Guidance on vaccine hesitancy, antimicrobial resistance surveillance, and NCD monitoring — as per the latest guidance, confirm the current edition.
- Government of India. National programmes for immunization, NCD control, and AMR surveillance — official programme documents, as per the latest guidance, confirm the current edition.
- Standard epidemiology, biostatistics, and health-informatics textbooks on segmentation, trend analysis, and surveillance methods — as per the latest editions, confirm the current edition.
- All datasets, district names, and numerical results in this chapter are illustrative toy examples created for teaching; they are not real findings and must not be cited as evidence.