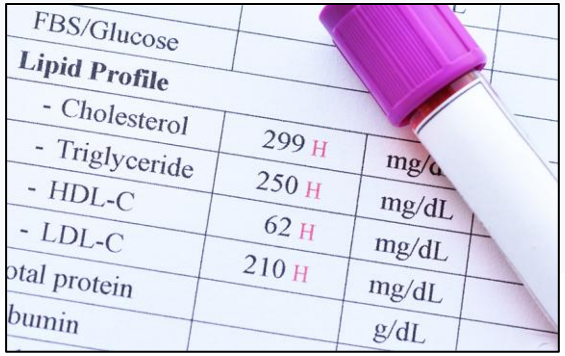
BIG DATA ANALYTICS IN DIAGNOSTIC SERVICES

DR. OOI THENG CHOON

Strategic Planning and Innovation, Premier Integrated Labs Sdn Bhd, Kuala Lumpur, Malaysia

Introduction & Context

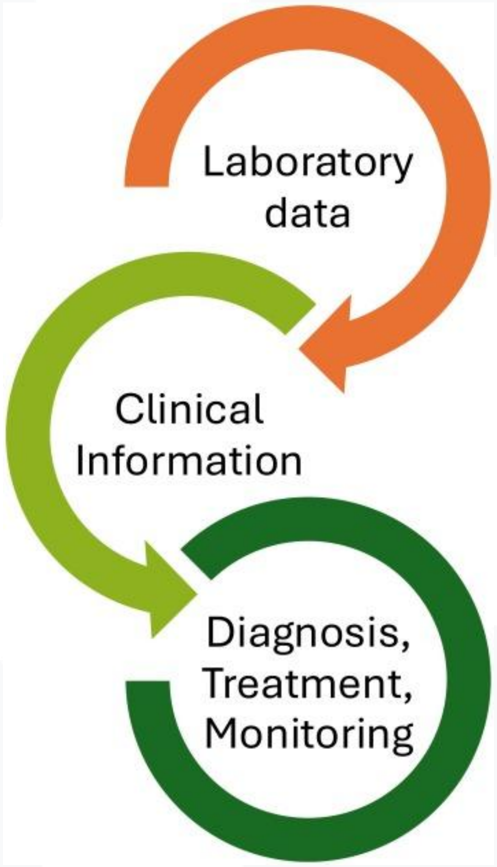
- Diagnostic laboratories generate **thousands to millions** of lab test results daily (depending on size).
- Each test result is not just a number/text/image — it's **data** that can be **aggregated**, **compared**, and **tracked**.
- **Increasing data complexity**: from simple chemistry tests → advanced molecular and omics testing (e.g. NGS).
- Its creates a rich but **underutilized data ecosystem**
- Big data analytics helps move from **raw test results → actionable insights** for clinicians, researchers, and policymakers.



Premier Integrated Labs

Why Diagnostic Lab Data is Valuable?

- **Clinically:** Most of the medical decisions depend on laboratory data. Its guides diagnosis, monitoring, and treatment.
- **Operationally:** Reflects lab performance, efficiency, test utilization.
- **Population level:** Can show disease trends, patient demographics, and health system needs.
- **Example:** HbA1c results → used not only for managing individual diabetes patients, but also to measure institutional/regional or national diabetes burden.



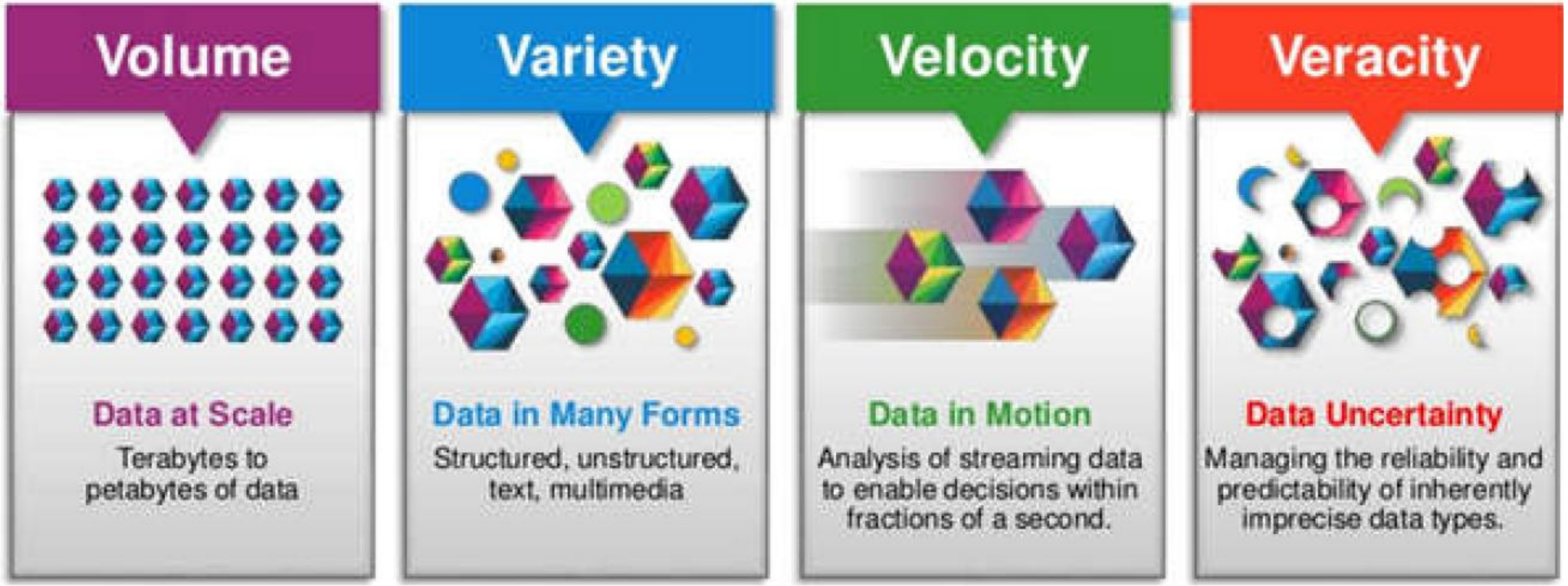
Source: https://doi.org/10.1016/j.cca.2025.120269

Current Context on Data Generated

- Most data stays siloed in the Laboratory Information System (LIS).
- Used mainly for **patient reports** and billing, not for broader analysis.
- This leads to wasted opportunities:
  - Rarely used for public health surveillance (e.g., **emerging infections**).
  - Underutilized for **non-communicable disease** monitoring.
  - Hardly used to support **public health decision making** and **strategic planning** at national level.

Big Data Analytics in Diagnostics Services

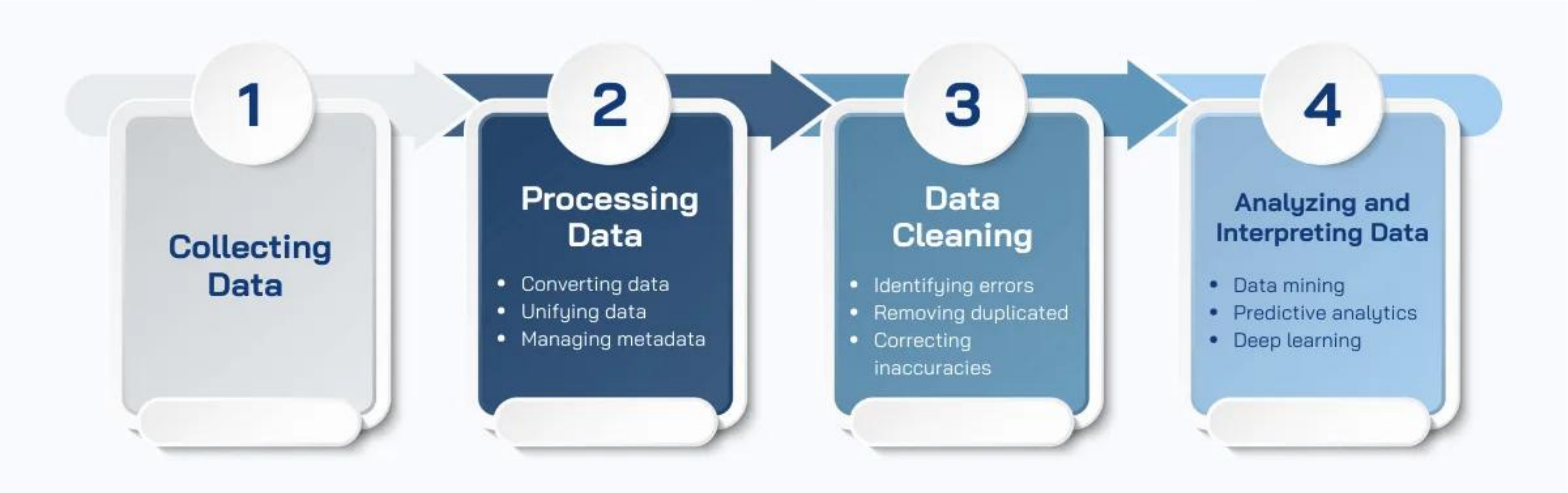
- "**BIG DATA**" - not just data size, but complexity and speed of data.
- The "**4 V**" in BIG DATA



Big Data Analytics in Diagnostics Services

- **Volume:** Millions of lab tests/year in large laboratory settings.
- **Velocity:** Results are produced continuously in real-time.
- **Variety:** Structured (numerical results) to unstructured (microscope images, NGS sequences) data.
- **Veracity:** data quality, consistency, missing values.

General Analytic Workflow

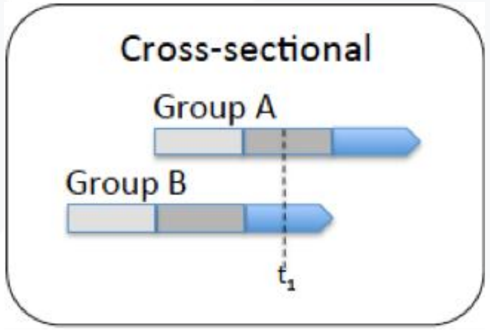


Premier Integrated Labs

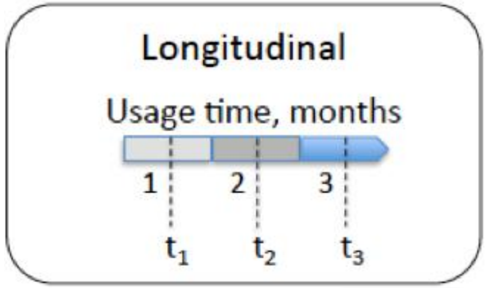
Considerations When Conducting Big Data Analytics

1. Choosing the Right Population / Cohort:

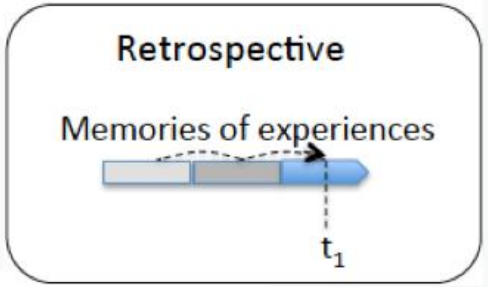
- **Cross-sectional analysis:** compare lab results across groups at **one time point (Prevalence study)**



- **Longitudinal analysis:** track test results and outcomes **over months** or **years** to see **progression** and patterns
    - Trend of HbA1c levels in a diabetes population over 5 years
    - Annual increase in antimicrobial resistance rates



- **Retrospective analysis:** Uses previously collected laboratory and clinical data to identify **historical trends**, **evaluate interventions**, or generate hypotheses.
    - Analyzing historical thalassemia genotyping results to map carrier prevalence by region.



Premier Integrated Labs

Additional Considerations...

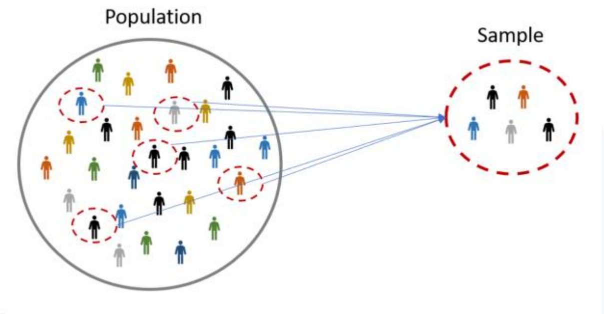
2. **Data Quality and Standardization:**
   - Address **errors**, **missing values** & **inconsistent coding** (e.g., multiple terms for same diagnosis)
   - Harmonize test codes, units, and reference ranges (important when combining multi-center data).
   - Need for **harmonized coding systems** (LOINC, ICD, Snomed)

LOINC®

SNOMED CT
The global language of healthcare

The World Health Organization's
International Classification of Diseases

3. **Choosing the Right Population:**
   - Define clear **inclusion & exclusion criteria** & ensure **representativeness** (age, sex, ethnicity, geography).
   - Need to consider **risk of bias**.
   - Example: Hospital-based lab data may inflate disease prevalence estimates compared to community screening data.

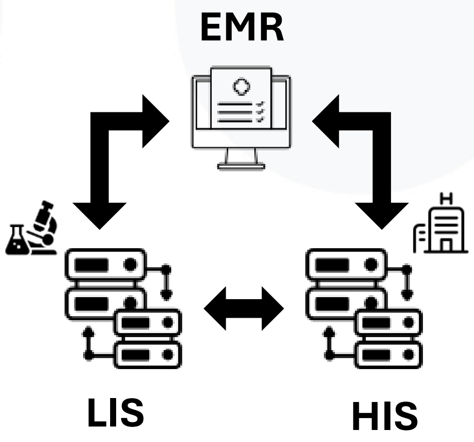


Additional Considerations…

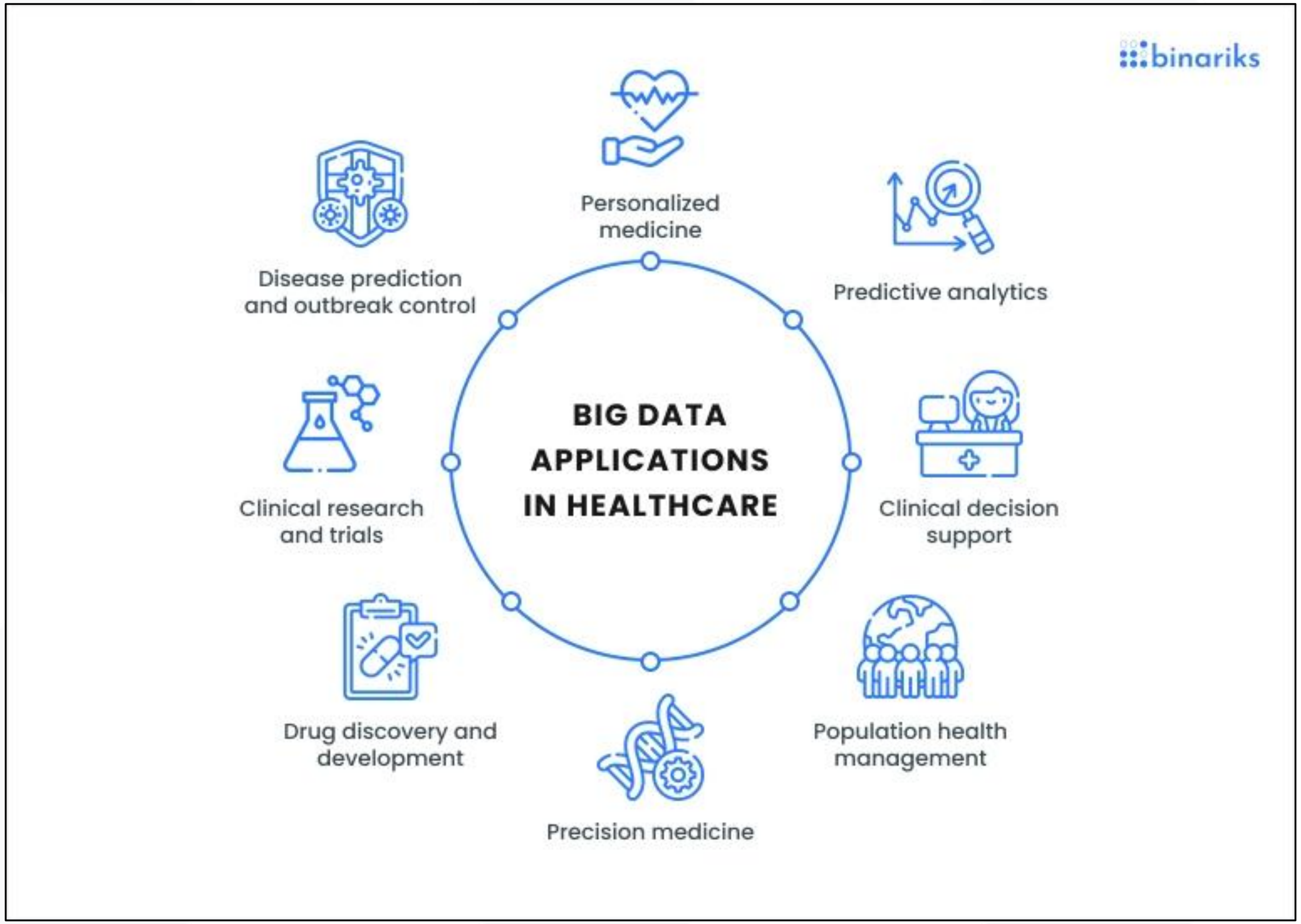
4. **Choosing the Right Visualization Tools:**
   - Use tools that match **complexity** & **volume** of the data (Python/R/Power BI/Tableau).
   - Creation of **interactive dashboards** allow **stratification** of population by sociodemographic factor, geography or other potential risk factors. Examples:
     - Prevalence of HPV-positive women stratified by age group, ethnicity & geography to enable tailored interventions (e.g., vaccination programs) for specific high-risk groups.
5. **Ethics, Privacy, and Governance:**
   - Data **privacy** (de-identification) & **security** (controlled access).
   - Obtain necessary **ethical** approvals.
   - Ensure compliance with local **regulations** (e.g., PDPA act and etc.)
6. **Sustainability and Scalability:**
   - Plan for continuous **data inflow**, system **updates**, and evolving **standards**.
   - Build solutions that can be scaled to new diseases or expanded to national registries.
7. **Also**, lets your system "**TALK**"!

Integration of Laboratory Data with Other Systems

- On its own, LIS provides results but often **lacks clinical context** (e.g., diagnosis, treatment, patient history).
- **LIS (Lab Information System):** Lab test results, QC, workflows.
- **HIS (Hospital Information System):** Admissions, diagnoses, procedures.
- **EMR (Electronic Medical Record):** Longitudinal patient history, prescriptions, outcomes.
- Linking LIS with HIS/EMR provides **richer**, **contextualized data**
- Integration allows:
  - ➢ Enriching lab data with **clinical context**
  - ➢ Complete **patient journey** (from screening → diagnosis → treatment → outcome).
  - ➢ Enabling **predictive**, **preventive** & **prescriptive** healthcare.
- Example: HbA1c trends from LIS + EMR comorbidities (hypertension, renal disease) → enables risk stratification and resource allocation



Applications of Big Data Analytics

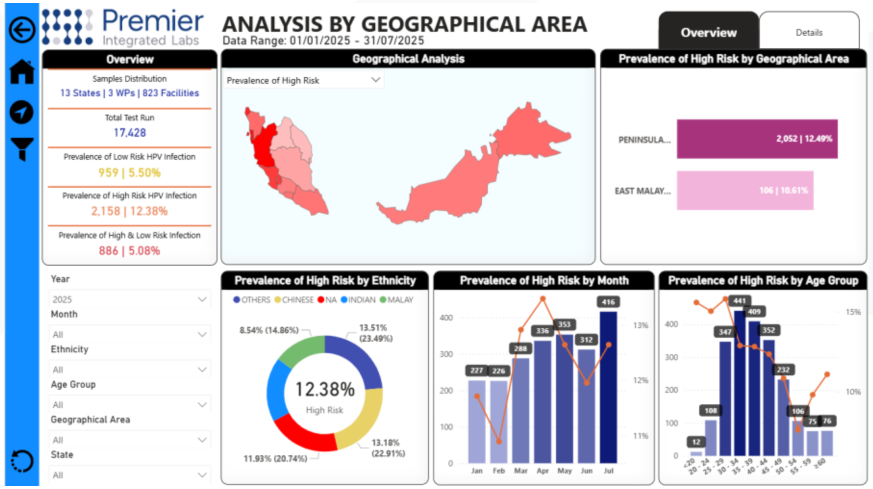


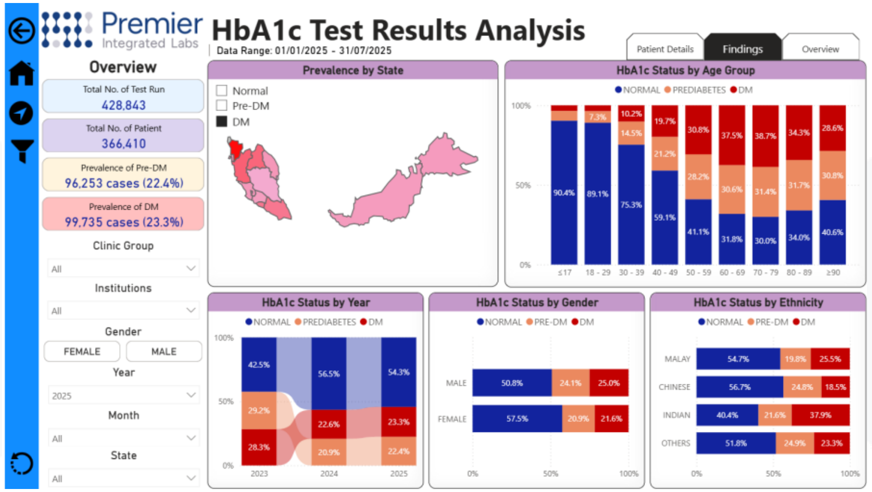
Source: https://binariks.com/blog/big-data-applications-in-healthcare/

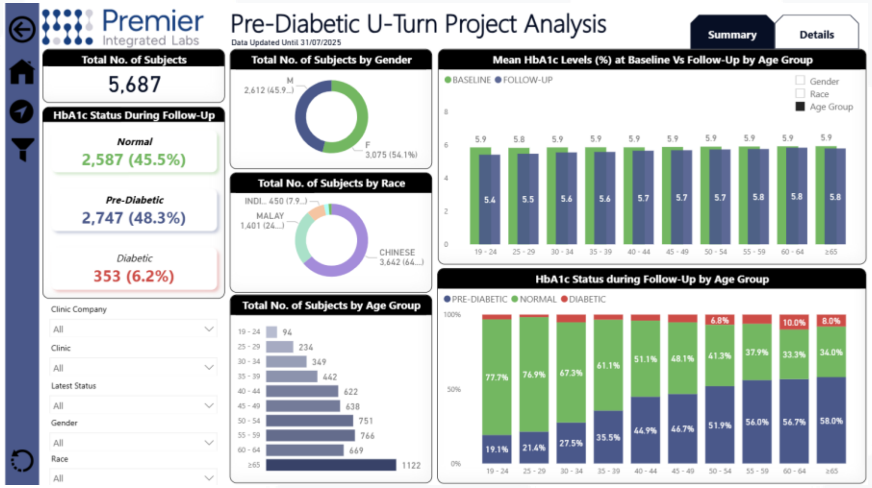
Big data analytics enables labs to move:

- From individual **patient level** → to **population** insights.
- From **descriptive reporting** → to **actionable knowledge**.

Examples of Applications...

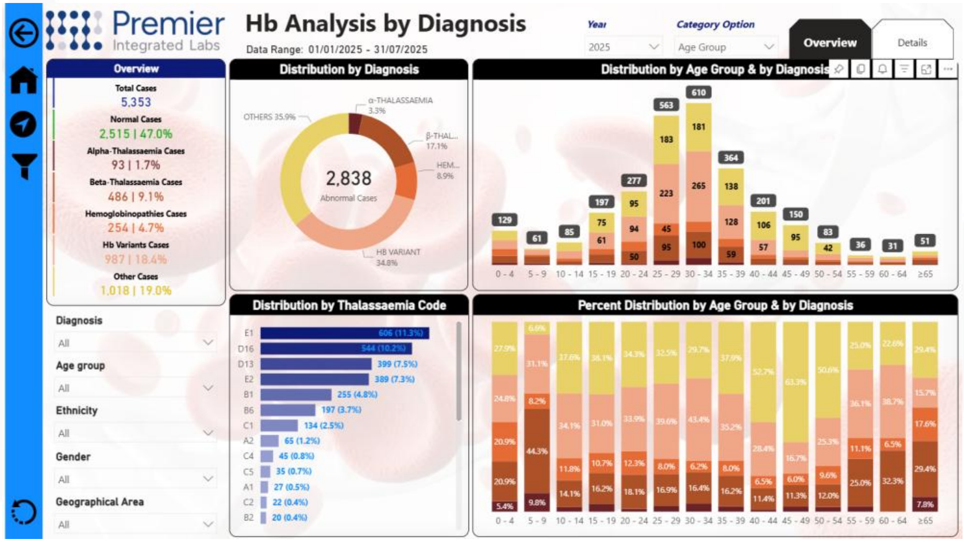


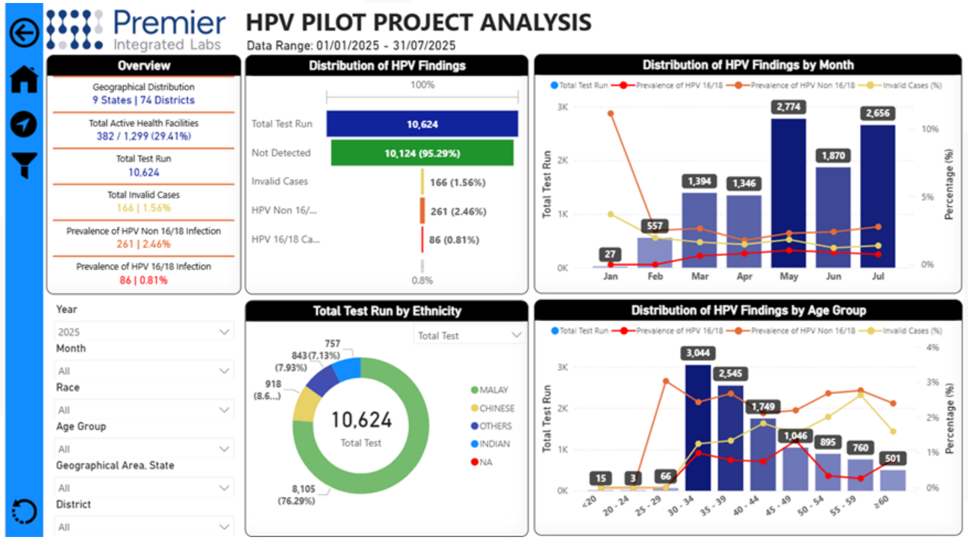




Integration of Laboratory Data with National Database/Registries

- Future Integration with National Health Databases / Public Health Systems
- Examples:
- **Thalassemia registry:** Hemoglobin electrophoresis + DNA testing results aggregated across labs → used for carrier screening programs and national prevention policies.
- **HPV screening program:** HPV DNA test results integrated with Pap smear/LBC screening → supports HPV vaccinations programs and cervical cancer preventive strategies.





AI and ML in Big Data Analytics

- AI thrives on big datasets — **more data** = **better training**, **more reliable** predictions
- Pattern recognition: ML can **detect hidden relationships** (e.g., biomarker clusters predicting cancer risk)
- Predictive analytics: Using historical lab data + patient demographics to build models that **forecast future disease trend**
    - ➢ Supports public health planning and resource allocation
- **Forecasting** infectious **disease outbreaks** from lab positivity trends
- **Predicting progression of NCDs** like diabetes or kidney disease
    - ➢ integrate multiple lab parameters to group patients by risk level towards certain NCD
    - ➢ Enables personalized monitoring and early intervention strategies
- Early warning: AI + lab big data could **flag unusual signals** (antimicrobial resistance patterns & etc.)

Other Challenges and Limitations

- **Technical challenges:** need for skilled personnels, computational power, data storage infrastructure.
  - Lack of data scientists, bioinformaticians, or machine learning experts in most lab
  - Requires advanced coding and computational skills to retrieve large data set from DB server, set-up automated ETL workflow and building of AI model

- **Cost & sustainability:** initial setup and ongoing maintenance cost can be high.
  - Implementing big data infrastructure (computational hardware, servers, cloud storage, analytics platforms, cybersecurity) requires significant investment
  - Smaller labs may struggle; large health systems may face budget prioritization issues

THANK YOU!

www.premierintegratedlabs.com.my