



การตรวจสอบความน่าเชื่อถือของข้อมูล

Data Verification & Quality Assessment

หลักสูตรการวิเคราะห์ข้อมูลสุขภาพ

1.0 INTRODUCTION

วัตถุประสงค์การเรียนรู้

1 ประเมินคุณภาพข้อมูล

ใช้กรอบ TRUST และ DQR (WHO) ในการประเมินคุณภาพข้อมูลสุขภาพ ครอบคลุม 4 มิติหลัก: completeness, consistency, timeliness, population data

2 Cross-Validation

ใช้เทคนิค triangulation, internal checks, benchmarking เพื่อลดความเสี่ยงจากข้อมูลผิดพลาด และยืนยันความน่าเชื่อถือ

3 จัดการปัญหาข้อมูล

ตรวจสอบและจัดการปัญหาที่พบบ่อย: missing data (MCAR/MAR/MNAR), outliers, logical consistency

ผลลัพธ์ที่คาดหวัง

ผู้เรียนสามารถออกแบบการตรวจสอบคุณภาพข้อมูล สรุปผลในรูปแบบ **Data Quality Report** และนำไปใช้ปรับปรุงระบบข้อมูลสุขภาพได้จริง

1.1 IMPORTANCE

ทำไม Data Quality สำคัญ

⚠️ ข้อมูลคุณภาพต่ำ

- ✗ ข้อมูลคลาดเคลื่อนจากความไม่ครบถ้วน
- ✗ การตัดสินใจเชิงนโยบายผิดพลาด
- ✗ สูญเสียความน่าเชื่อถือของระบบ
- ✗ ต้นทุนการแก้ไขสูงในอนาคต

✓ ข้อมูลคุณภาพสูง

- ✓ ข้อมูลที่ถูกต้องและน่าเชื่อถือ
- ✓ การตัดสินใจที่มีหลักฐานสนับสนุน
- ✓ ความมั่นใจของผู้ใช้ข้อมูล
- ✓ ประหยัดต้นทุนระยะยาว

"Sound decisions are based on sound data." — การตัดสินใจที่ดีต้องอิงจากข้อมูลที่มีคุณภาพ (WHO, 2022)

Figure 1: Impact of data quality on decision-making

1.2 FRAMEWORK

DQR Framework (WHO) ภาระม

1 Completeness & Timeliness

ความครบถ้วนและทันเวลาของการรายงานข้อมูล

2 Internal Consistency

ความสอดคล้องภายในของข้อมูล

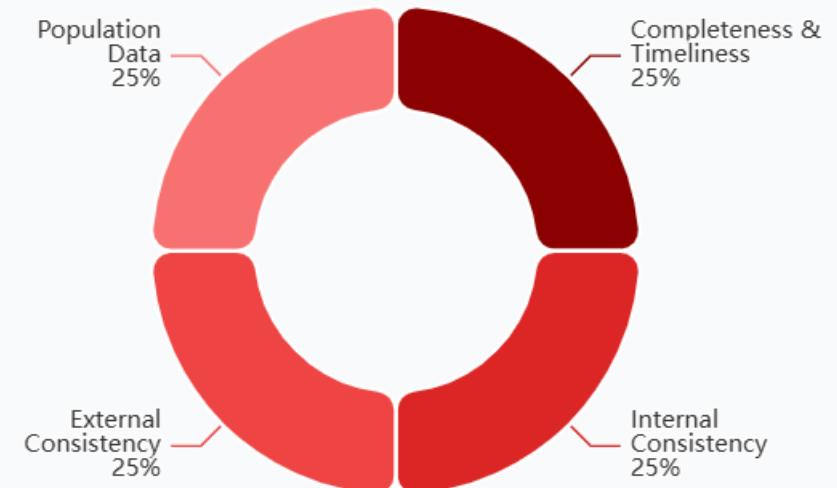
3 External Consistency

ความสอดคล้องกับแหล่งข้อมูลภายนอก

4 Population Data

ความเหมาะสมของข้อมูลประชากร (ตัวหาร)

Data Quality Review Framework



1.3 DIMENSION 1

Completeness & Timeliness

✓ Completeness

ระดับหน่วยบริการ

- % หน่วยบริการที่ส่งรายงานครบตามรอบ
- % ตัวแปรที่กรอกข้อมูลครบถ้วน
- จำนวนรายงานที่ขาดหายไป

ระดับพื้นที่

- % อำเภอ/จังหวัด ที่ส่งข้อมูลครบ
- แนวโน้มการส่งข้อมูลรายเดือน
- จุดบกพร่องที่ต้องติดตาม

⌚ Timeliness

นิยามและเกณฑ์

- รายงานรายเดือน: ส่งภายในวันที่ 10 ของเดือนถัดไป
- รายงานรายไตรมาส: ส่งภายใน 15 วันหลังสิ้นไตรมาส
- รายงานรายปี: ส่งภายใน 30 วันหลังสิ้นปี

ตัวชี้วัด

$$\% \text{ Timeliness} = (\text{รายงานที่ส่งทันเวลา} / \text{รายงานทั้งหมดที่ต้องส่ง}) \times 100$$

🕒 ตัวอย่าง Timeline การติดตาม

ม.ค. → ก.พ. (deadline 10 ก.พ.) → ติดตาม 15 ก.พ. → รายงานผล 20 ก.พ.

▣ Source: WHO. Data Quality Assurance: Module 2 – Desk Review. Geneva: WHO; 2022.

1.4 COMPLETENESS TYPES

Completeness ที่ต้องตรวจ

Reporting Completeness

นิยาม

ความครบถ้วนของการส่งรายงานจากหน่วยบริการ

ตัวอย่าง

หน่วยบริการ 100 แห่ง ส่งรายงาน 95 แห่ง = 95% completeness

ข้อจำกัด

รายงานครบแต่อาจมีว้าวุ่นขาดหาย

Data Element Completeness

นิยาม

ความครบถ้วนของข้อมูลในแต่ละตัวแปร

ตัวอย่าง

ตัวแปร "อายุ" มีข้อมูล 850/1000 ระเบียน = 85% completeness

ความสำคัญ

ตัวแปรสำคัญต้องมี completeness > 90%

ข้อควรระวัง

รายงานครบถ้วน ≠ ข้อมูลมีคุณภาพ ต้องตรวจสอบทั้ง Reporting Completeness และ Data Element Completeness พร้อมกัน

Table 1: Comparison of reporting completeness vs data element completeness

1.5 TIMELINESS

Timeliness: นิยามและตัวชี้วัด

📅 รายเดือน

Deadline
วันที่ 10 ของเดือนก็ดีไป

ตัวอย่าง
ข้อมูลม.c. → ส่งภายใน 10 ก.พ.

📅 รายไตรมาส

Deadline
15 วันหลังสิ้นไตรมาส

ตัวอย่าง
ไตรมาส 1 (มี.ค.) → ส่งภายใน 15 เม.ย.

📅 รายปี

Deadline
30 วันหลังสิ้นปี

ตัวอย่าง
ข้อมูล 2024 → ส่งภายใน 30 ม.ค. 2025

สูตรคำนวณ % Timeliness

$$\% \text{ Timeliness} = (\text{รายงานที่ส่งก่อน Deadline} / \text{รายงานทั้งหมดที่ต้องส่ง}) \times 100$$

ตัวอย่างที่ 1

ส่งครบ 95/100 รายงาน ทันเวลา 90 รายงาน = **90%**

ตัวอย่างที่ 2

ส่งครบ 88/100 รายงาน ทันเวลา 85 รายงาน = **85%**

1.6 DIMENSION 2

Internal Consistency

การตรวจสอบภายใน

1. แนวโน้มผิดปกติ (Trend Check)

ตรวจสอบความผิดปกติของแนวโน้มข้อมูล เช่น จำนวนผู้ป่วยเพิ่มขึ้นหรือลดลงผิดปกติ

2. ค่าผิดปกติ (Outliers)

ตรวจหาค่าที่อยู่นอกช่วงปกติโดยใช้ IQR หรือ Z-score

3. ความสัมพันธ์ของตัวแปร

ตรวจสอบความสอดคล้องระหว่างตัวแปรที่ควรสัมพันธ์กัน

ตัวอย่างการตรวจจับ Outliers



หมายเหตุ: จุดสีแดงแสดงค่าที่น่าสงสัย ต้องตรวจสอบเพิ่มเติม

ตัวอย่างความสัมพันธ์ที่ต้องตรวจ

อายุ VS น้ำหนัก

หากอายุ 1 ปี น้ำหนัก 20 กก. → ตรวจสอบ

วันที่รักษา VS วันที่จำหน่าย

วันจำหน่ายก่อนวันรักษา → ผิด

อายุครรภ์ VS น้ำหนักหารา

อายุครรภ์ 40 สัปดาห์ น้ำหนัก 1,500 กรัม → ตรวจสอบ

1.7 DIMENSION 3

External Consistency

การเทียบเคียงภายนอก

एक Benchmarking

เปรียบเทียบกับมาตรฐานสากลหรือแหล่งข้อมูลที่เชื่อถือได้

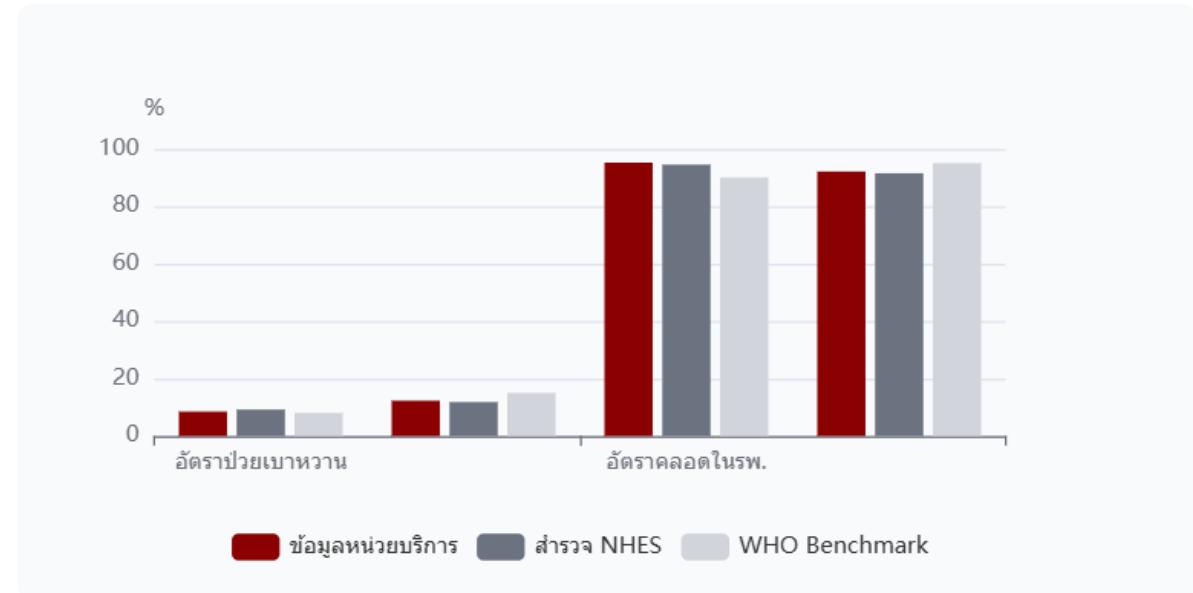
- WHO Global Health Observatory
- World Bank Open Data
- ข้อมูลจากประเทศเพื่อนบ้าน

एก Survey Comparison

เทียบกับข้อมูลจากการสำรวจพิเศษ

- สำรวจสุขภาพประชาชน (NHES)
- สำรวจภาวะเศรษฐกิจและสังคม
- การสำรวจเฉพาะเรื่อง

ตัวอย่างการเปรียบเทียบ



✓ สอดคล้อง: ข้อมูลใกล้เคียงกัน < 10% ความแตกต่าง

⚠ ตรวจสอบ: ความแตกต่าง 10-20% ต้องหาสาเหตุ

✗ ผิดปกติ: ความแตกต่าง > 20% ต้องสืบสวน

หลักการ: ความแตกต่างที่พบไม่ใช่เรื่องผิด แต่ต้องมีคำอธิบายที่สมเหตุสมผล เช่น ความแตกต่างของนิยาม วิธีการเก็บข้อมูล หรือช่วงเวลา

1.8 DIMENSION 4

Population Data (Denominator Check)

ความสำคัญของตัวหาร

Rate =

Numerator
Denominator
 $\times 100,000$ (หรือ %)

ปัญหาที่พบบ่อย

- ใช้ข้อมูลประชากรที่ไม่ตรงปี
- ไม่ปรับปรุงตามการย้ายถิ่น
- ข้อมูลประชากรไม่ครอบคลุมพื้นที่

แหล่งข้อมูลประชากร

1. สำนักงานสถิติแห่งชาติ

ข้อมูลประชากรจากการสำรวจสำมะโนประชากร

2. กรมการปกครอง

ทะเบียนราษฎร์ ข้อมูลการย้ายถิ่น

3. ข้อมูลประมาณการ (Projection)

ใช้เมื่อไม่มีข้อมูลจริง ต้องระบุวิธีการคำนวณ

4. ข้อมูลเฉพาะกลุ่มเป้าหมาย

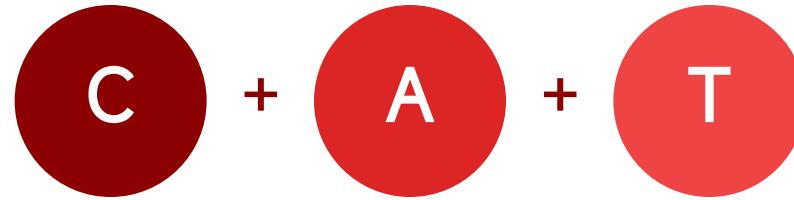
เช่น หญิงตั้งครรภ์ ผู้สูงอายุ เด็ก 0-5 ปี

i ตัวอย่างการตรวจสอบ

อัตราการตายของมาตราดา = $(\text{จำนวนมาตราตาย} / \text{จำนวนหญิงตั้งครรภ์ทั้งหมด}) \times 100,000$ ต้องใช้ข้อมูลหญิงตั้งครรภ์ที่ถูกต้อง ไม่ใช่ประชากรหญิงทั้งหมด

2.1 CDC FRAMEWORK

CAT: Completeness, Accuracy, Timeliness



Completeness

นิยาม

ข้อมูลครบถ้วนทุกเคสและทุกตัวแปร

การวัด

- % เคสที่รายงานครบ
- % ตัวแปรที่กรอกข้อมูล

เกณฑ์

> 90% สำหรับตัวแปรสำคัญ

Accuracy

นิยาม

ข้อมูลถูกต้องตามความจริง

การวัด

- % ความถูกต้องของการเข้ารหัส
- การตรวจสอบกับแหล่งที่มา

เกณฑ์

> 95% ความถูกต้อง

Timeliness

นิยาม

ข้อมูลพร้อมใช้ตามเวลาที่กำหนด

การวัด

- % รายงานที่ส่งทันเวลา
- ระยะเวลาจากเก็บข้อมูลถึงรายงาน

เกณฑ์

> 80% ส่งทันเวลา

หมายเหตุ: CAT เป็นกรอบพื้นฐานที่ใช้กันอย่างแพร่หลายในงาน public health surveillance ของ CDC และสามารถนำไปประยุกต์ใช้กับระบบข้อมูลสุขภาพของไทยได้

Source: CDC. Key characteristics of data quality in public health surveillance. Birth Defects Surveillance Manual; 2020.

2.2 ACCURACY

Accuracy/Validity ในเชิงปฏิบัติ

ตัวอย่างข้อมูลที่ถูกต้อง VS ผิด

✓ ข้อมูลถูกต้อง

โรค:
เบาหวานชนิดที่ 2 (E11)

อายุ:
55 ปี

วันที่:
15 ม.ค. 2024

✗ ข้อมูลผิด

โรค:
เบาหวานชนิดที่ 1 (E10) ← ผิดนิยาม

อายุ:
155 ปี ← เกินช่วง

วันที่:
15 ม.ค. 2029 ← อนาคต

Checklist ตรวจสอบความถูกต้อง

1. นิยามเคส (Case Definition)

- ใช้นิยามตามมาตรฐาน
- ผู้ป่วยตรงตามเกณฑ์ในนิยาม
- ไม่มีการคัดกรองที่ผิดพลาด

2. การเข้ารหัส (Coding)

- ICD-10 ถูกต้อง
- รหัสตัวแปรถูกต้อง
- ไม่มีการบันทึกซ้ำ

3. การตรวจสอบกับแหล่งต้นทาง

- สุมตรวจ 5-10% ของข้อมูล
- เปรียบเทียบกับ medical record
- ตรวจสอบ tracer indicators

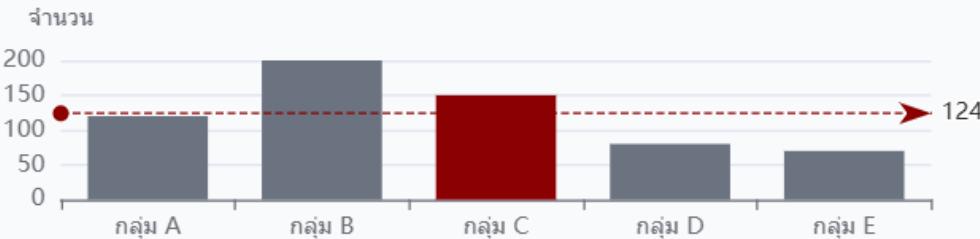
2.3 REPRESENTATIVENESS

Representativeness และมิติคุณภาพอื่น

ความเป็นตัวแทน

นิยาม

ข้อมูลต้องสะท้อนประชากรเป้าหมายจริง ไม่บิดเบี้ยวจากการเลือกตัวอย่างหรือวิธีการเก็บข้อมูล



ตัวอย่าง Bias ที่พบ

หน่วยบริการในเมืองส่งข้อมูลดี แต่ชนบทส่งไม่ครบ → อัตราป่วยสูงกว่าความจริง

มิติคุณภาพข้อมูลอื่น

Relevance (ความเกี่ยวข้อง)

ข้อมูลตอบโจทย์วัตถุประสงค์การใช้งาน

Accessibility (การเข้าถึง)

ผู้ใช้สามารถเข้าถึงข้อมูลได้ง่าย

Granularity (ระดับรายละเอียด)

ข้อมูลมีรายละเอียดเพียงพอต่อการวิเคราะห์

Interoperability (ความสามารถเชื่อมโยง)

ข้อมูลสามารถเชื่อมโยงกับระบบอื่นได้

การประเมิน: ต้องพิจารณาความเหมาะสมต่อวัตถุประสงค์การใช้งาน (fitness for purpose) ไม่ใช้แค่คุณภาพในทางเทคนิคอย่างเดียว

3.1 DQA METHODS

Desk Review vs Data Verification

Desk Review

ลักษณะ

ตรวจสอบชุดข้อมูลรายงานที่รวบรวมแล้ว

สถานที่

ที่ทำการ ไม่ต้องเป็นหน่วยบริการ

เครื่องมือ

Excel, DHIS2, ซอฟต์แวร์วิเคราะห์

ข้อดี

- รวดเร็ว ต้นทุนต่ำ
- ตรวจชื่อ穆ยบ้อนหลังได้
- เปลี่ยนเทียบหลักแหล่งได้

ข้อจำกัด

ไม่สามารถตรวจสอบความถูกต้องกับเอกสารต้นทาง

Data Verification

ลักษณะ

ตรวจสอบเอกสารต้นทางที่หน่วยบริการ

สถานที่

ต้องไปตรวจที่หน่วยบริการจริง

เครื่องมือ

Checklist, แบบฟอร์มตรวจสอบ

ข้อดี

- ตรวจสอบความถูกต้องได้จริง
- พบปัญหาระบบงาน
- ประเมินความสามารถระบบ

ข้อจำกัด

ใช้เวลาและต้นทุนสูง ต้องเดินทาง

2-Layer Audit Approach

Desk Review



Data Verification



Quality Report

3.2 FACILITY VERIFICATION

Data Verification ที่ระดับหน่วยบริการ

ขั้นตอนการตรวจ

1 เตรียมตัว

เลือกหน่วยบริการ กำหนด tracer indicators

2 ตรวจสอบสารต้นทาง

OPD card, IPD record, Lab result

3 เปรียบเทียบกับรายงาน

ตรวจสอบว่าข้อมูลตรงกันหรือไม่

4 บันทึกผลและสรุป

คำนวณ % accuracy และปัญหาที่พบ

Tracer Indicators ตัวอย่าง

Indicator	Source	Check
จำนวนผู้ป่วยนอก	OPD card	ตรงกับรายงาน
จำนวนผู้ป่วยใน	IPD record	ตรงกับรายงาน
จำนวนคลอด	Birth record	ตรงกับรายงาน
จำนวนตาย	Death certificate	ตรงกับรายงาน
การฉีดวัคซีน	Vaccine log	ตรงกับรายงาน

หมายเหตุ: เลือก tracer indicators ที่สะท้อนคุณภาพข้อมูลโดยรวม ไม่ใช่แค่ตัวชี้วัดที่ง่ายตรวจ

■ Source: WHO. Country-level Holistic Data Quality Assurance. Geneva: WHO; 2022.

3.3 DESK REVIEW

Desk Review: โครงสร้างการประเมิน

4 ขั้นตอนหลัก

1 Completeness & Timeliness

- ตรวจสอบ % รายงานที่ส่งครบ
- ตรวจสอบ % รายงานที่ส่งทันเวลา
- วิเคราะห์แนวโน้มการส่งข้อมูล

2 Internal Consistency

- ตรวจสอบ outliers และค่าผิดปกติ
- ตรวจสอบความสัมพันธ์ตัวแปร
- วิเคราะห์ trend over time

3 External Comparison

- เทียบกับ survey data
- เทียบกับ WHO benchmark
- เทียบกับข้อมูลปีก่อน

4 Population Data

- ตรวจสอบความเหมาะสมตัวหาร
- ตรวจสอบแหล่งข้อมูลประชากร
- คำนวนอัตราต่างๆ

Flowchart การดำเนินงาน

เริ่มต้น Desk Review



รวบรวมข้อมูล

รายงาน, ข้อมูลประชากร, benchmark



วิเคราะห์ 4 มิติ



สรุปผลและจัดลำดับปัญหา



จัดทำ Data Quality Report

Missing Data: MCAR, MAR, MNAR

MCAR

Missing

Completely at Random

นิยาม

ข้อมูลหายแบบสุ่ม ไม่เกี่ยวกับตัวแปรใดๆ

ตัวอย่าง

เครื่องวัดน้ำหนักเสีย ข้อมูลหายแบบสุ่ม

การจัดการ

Listwise deletion, Mean imputation ได้

MAR

Missing

at Random

นิยาม

ข้อมูลหายเกี่ยวกับตัวแปรที่สังเกตได้

ตัวอย่าง

ผู้สูงอายุไม่กรอกข้อมูลรายได้

การจัดการ

Multiple imputation, MLE

MNAR

Missing

Not at Random

นิยาม

ข้อมูลหายเกี่ยวกับค่าที่หายไป

ตัวอย่าง

คนเดิมหนักไม่กรอกข้อมูลการดื่ม

การจัดการ

Sensitivity analysis, ต้องใช้ domain knowledge

สรุปการเปรียบเทียบ

ลักษณะ	MCAR	MAR	MNAR
ความสัมพันธ์	ไม่มี	กับตัวแปรที่สังเกตได้	กับค่าที่หายไป
ความยากง่าย	ง่าย	ปานกลาง	ยาก
วิธีจัดการ	Deletion, Simple imputation	Multiple imputation	Special methods
ความเสี่ยง bias	ต่ำ	ปานกลาง	สูง

4.2 OUTLIERS

ตรวจ Outliers และความผิดตระกูล

การตรวจสอบ Outliers

1. Range Check

ตรวจสอบว่าค่าอยู่ในช่วงที่เป็นไปได้

ตัวอย่าง:

- อายุ: 0-120 ปี
- น้ำหนักทารก: 500-5,000 กรัม
- ความดันโลหิต: 60-250 mmHg

2. IQR Method

Outliers $< Q1 - 1.5 \times \text{IQR}$ หรือ $> Q3 + 1.5 \times \text{IQR}$

3. Z-Score Method

$|Z\text{-score}| > 3 \rightarrow \text{Outlier}$

Logical Consistency Checks

อายุครรภ์

- อายุครรภ์: 20-44 สัปดาห์
- อายุครรภ์ < 37 สัปดาห์ = คลอดก่อนกำหนด
- อายุครรภ์ > 42 สัปดาห์ = ครรภ์เกินกำหนด

วันที่

- วันรักษา \leq วันจำหน่าย
- วันเกิด \leq วันรักษา
- ไม่มีวันที่ในอนาคต

ค่า Lab

- ซีโนโกลบิน: 5-20 g/dL
- น้ำตาลในเลือด: 50-500 mg/dL
- อุณหภูมิ: 35-42 °C

ตัวอย่าง Boxplot แสดง Outliers



4.3 CROSS-VALIDATION

Cross-Validation กับแหล่งอื่น

เทคนิค Triangulation

แหล่งข้อมูลที่ 1

ข้อมูลจากหน่วยบริการ (Routine data)

+

แหล่งข้อมูลที่ 2

ข้อมูลจากการสำรวจ (Survey data)

+

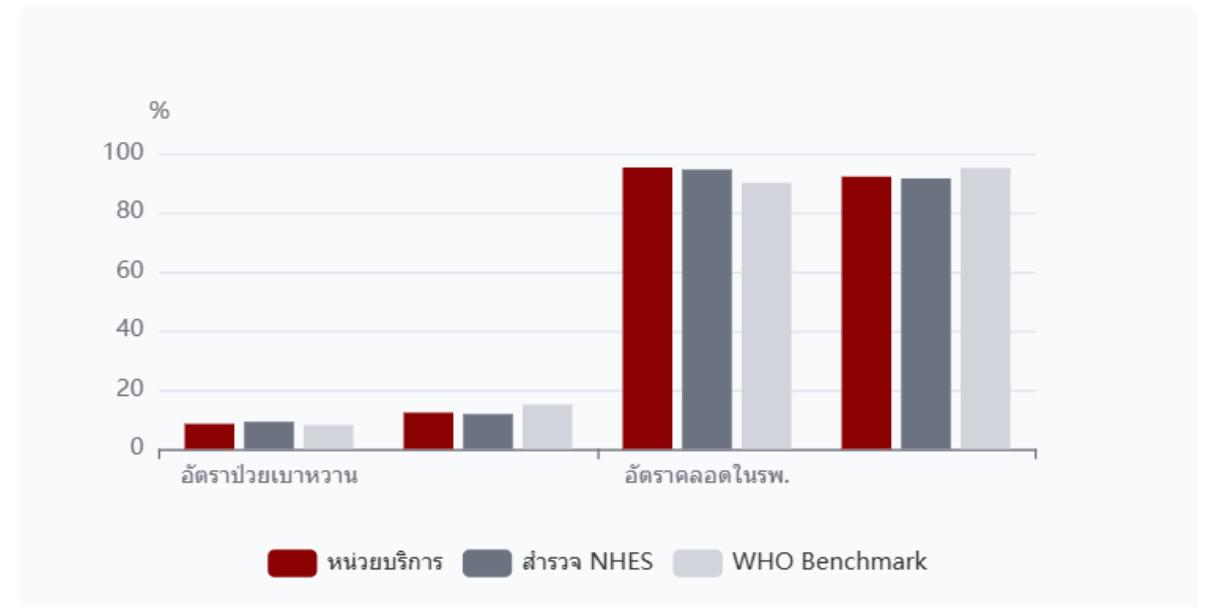
แหล่งข้อมูลที่ 3

ข้อมูลมาตรฐานสากล (WHO benchmark)

↓

ยืนยันความน่าเชื่อถือ

ตัวอย่างการเปรียบเทียบ



✓ ผลลัพธ์: ข้อมูลทั้ง 3 แหล่งสอดคล้องกัน → น่าเชื่อถือ

⚠ หากแตกต่าง: ต้องหาสาเหตุและอธิบายให้ได้

หลักการสำคัญ: ความแตกต่างที่พบไม่ใช่เรื่องผิด แต่ต้องมีคำอธิบายที่สมเหตุสมผล เช่น ความแตกต่างของนิยาม ช่วงเวลาเก็บข้อมูล หรือวิธีการเก็บข้อมูล

5.1 IMPROVEMENT

Data Quality Improvement Plan

จากผล DQR สู่แผนปรับปรุง

1. วิเคราะห์ผล DQR

ระบุปัญหาหลักและจัดลำดับความสำคัญ

2. กำหนดกิจกรรม

ออกแบบกิจกรรมแก้ไขตามสาเหตุ

3. กำหนดผู้รับผิดชอบ

มอบหมายหน่วยงาน/บุคคลรับผิดชอบ

4. กำหนดระยะเวลา

ตั้งเป้าหมายและ timeline ที่ชัดเจน

5. ติดตามและประเมิน

ตรวจสอบความก้าวหน้าและปรับแผน

ตัวอย่าง Action Plan

ปัญหา	กิจกรรม	ระยะเวลา
Completeness ต่ำ	อบรมการกรอกข้อมูล	1-2 เดือน
Timeliness ต่ำ	ปรับระบบการส่งรายงาน	2-3 เดือน
Outliers มาก	ตรวจสอบและแก้ไขข้อมูล	1 เดือน
ขาด benchmark	พัฒนา validation rules	3-6 เดือน

Timeline ตัวอย่าง



5.2 DASHBOARD

Data Quality Dashboard/Report

KPIs ที่ติดตาม

Completeness **92%**

เป้าหมาย: >90%

Timeliness **88%**

เป้าหมาย: >80%

Accuracy **85%**

เป้าหมาย: >95%

Validity **96%**

เป้าหมาย: >95%

Dashboard Mockup

Completeness by Month



Timeliness Trend



Outliers Detection



Data Quality Score



องค์ประกอบของ Data Quality Report

- สรุปผลโดยรวม

- วิเคราะห์แต่ละมิติ

- ปัญหาที่พบและสาเหตุ

- แนวทางการแก้ไข



สรุปและเชื่อมไปการใช้งานจริง

กรอบการทำงาน

- ✓ ใช้กรอบ DQR + CAT
- ✓ ตรวจ 4 มิติหลัก
- ✓ Desk Review + Data Verification
- ✓ Cross-validation หลายแหล่ง

ผลลัพธ์

- 📋 Data Quality Report
- ⚡ Dashboard ติดตาม KPIs
- ☰ Improvement Plan
- ⟳ ติดตามอย่างต่อเนื่อง

ต่อไป: ฝึกปฏิบัติ

ตรวจสอบภาพข้อมูลหลัง merge ด้วยแบบฝึกหัด M1-Workshop-Data-Integration