

Achieving quality primary care EMR data: a description of the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) data in Alberta

Stephanie Garies, PhD
Michael Cummings, PhD

Power of Population Data Science Webinar
July 29, 2020

Webinar Outline

- Introduction to EMR Data & CPCSSN
- Why Care About Data Quality?
- CPCSSN Data Flow & Processes in Alberta
- CPCSSN Data Quality in Alberta

EMR Data

- ~82% Canadian physicians using EMR in practice
(CMA Physician Workforce Survey, 2017)
- New(er) source of health data
- Advantages: large volume of detailed clinical & risk factor information; routinely collected; 'real-world' population
- Limitations: sample of all patients/providers; unstructured, 'messy' data

Canadian Primary Care Sentinel Surveillance Network (CPCSSN)



cpcssn.ca

CPCSSN in Alberta



CPCSSN in Alberta



EMR Data Quality

- Generated from processes of patient care and administrative functions
- Poor quality data can result in incorrect conclusions affecting patient care, research findings, policy decisions, clinical practice guidelines
- Are data appropriate for the intended secondary use?

EMR Data Quality

Potential sources of bias and data quality issues in Canadian primary care EMR data (adapted from Verheij et al.)



Policy / System Level

Provincial/territorial payment models
Incentives
Clinical practice guidelines

Delivery of Care

Patient access to healthcare
Patient healthcare utilization
Practice workload
Data sharing between providers



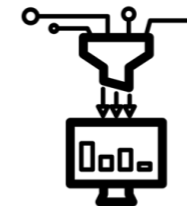
Recording in EMR

Clinical relevance
Extent of EMR use
Data entry practices
Measurement errors
EMR functionalities / interface
Patients' ability to modify own health record
Knowledge of EMR & coding system



Extraction from EMR

Different extraction methods for each EMR
Availability of data elements varies by EMR
Some data difficult/not possible to extract
Extractor personnel (e.g. vendor, data manager)
Types of providers permitting data extraction
Provincial/territorial patient consent models



Translation into Database

Integration of different classification & coding systems/terminologies
Processing methods
Common data model standardization
Cleaning & coding algorithms
Case definition algorithms
Complexity of data limits processing



Data cut for Researcher

Knowledge of EMR data
Linkages to other databases

Methods, Outcome & Interpretation

Different data users make different choices about cohort selection, methods, analysis, interpretation
Degree of content knowledge

Care Zone

Database Zone

Research Zone

EMR Data Quality Assessment

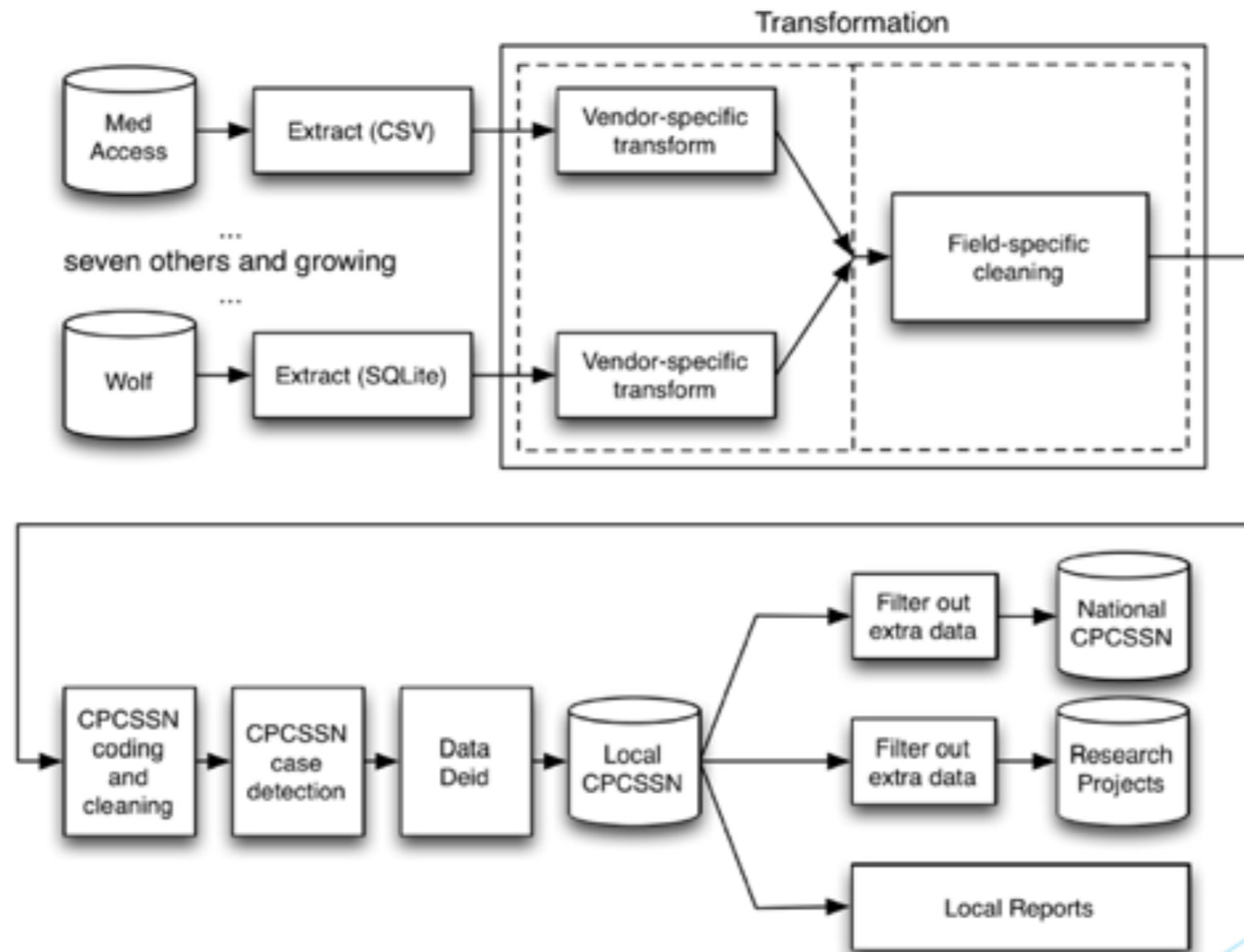
1. Data quality reporting framework

Kahn, Brown, Chun, et al. Transparent reporting of data quality in distributed data networks. EGEMS (Washington, DC). 2015;3(1):Article 7

2. Detailed documentation for CPCSSN data in Alberta

Garies, Cummings, Forst, et al. Achieving quality primary care data: a description of the Canadian Primary Care Sentinel Surveillance Network data capture, extraction, and processing in Alberta. IJPDS 2019;4(2)

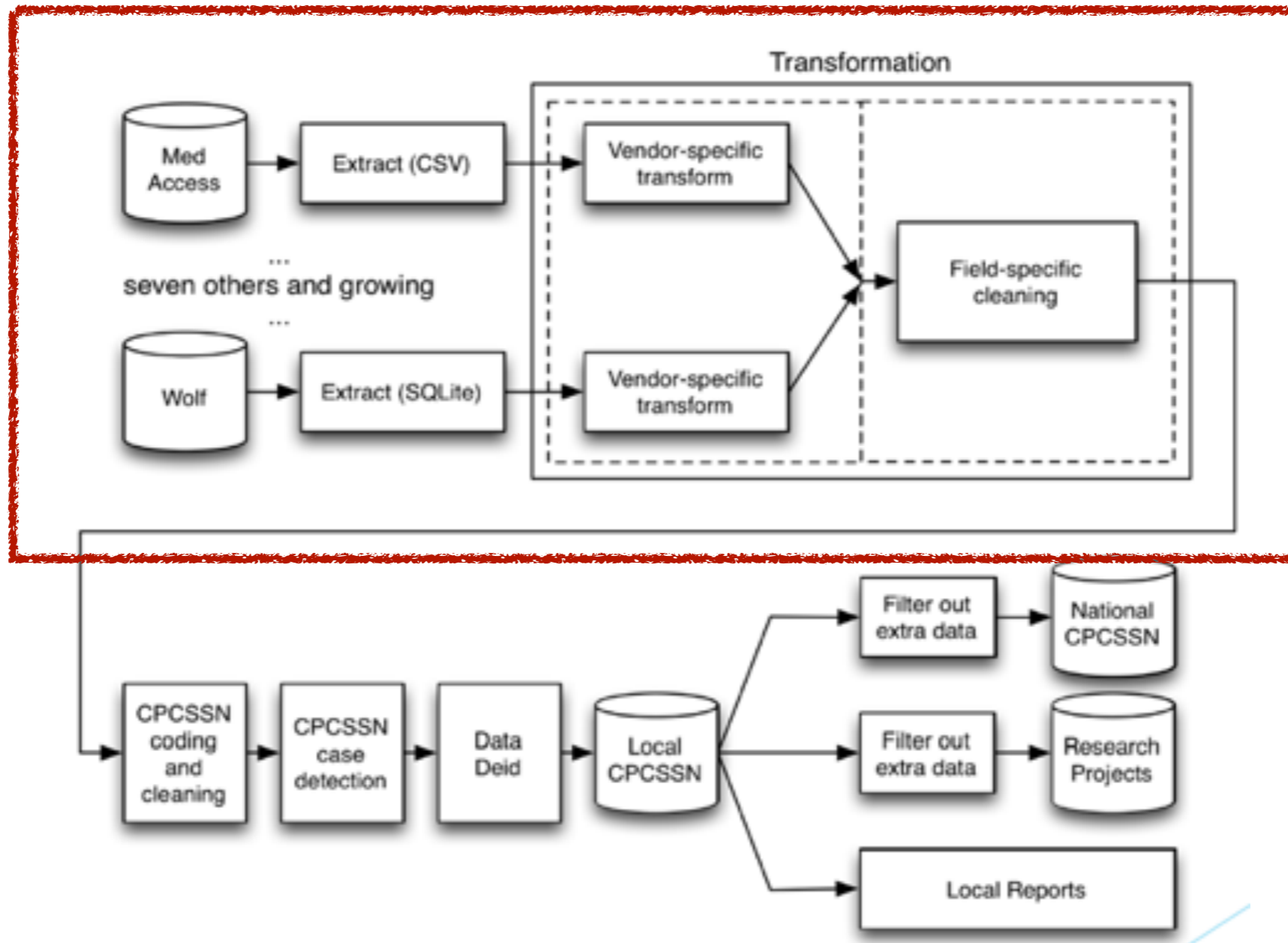
CPCSSN Data Flow in AB



Data Capture

- In Alberta (as of December 31, 2019):
 - ▶ 436,927 patients
 - ▶ 354 providers in 55 clinics
 - ▶ 5 EMR systems
 - ▶ NAPCReN & SAPCReN as ‘data stewards’
- CPCSSN data includes: patient demographics, visits (diagnoses, symptoms), physician billing claims, physical measurements (BP, ht, wt, etc.), prescribed medications, lab results, medical procedures, referrals, risk factors, vaccinations, allergies

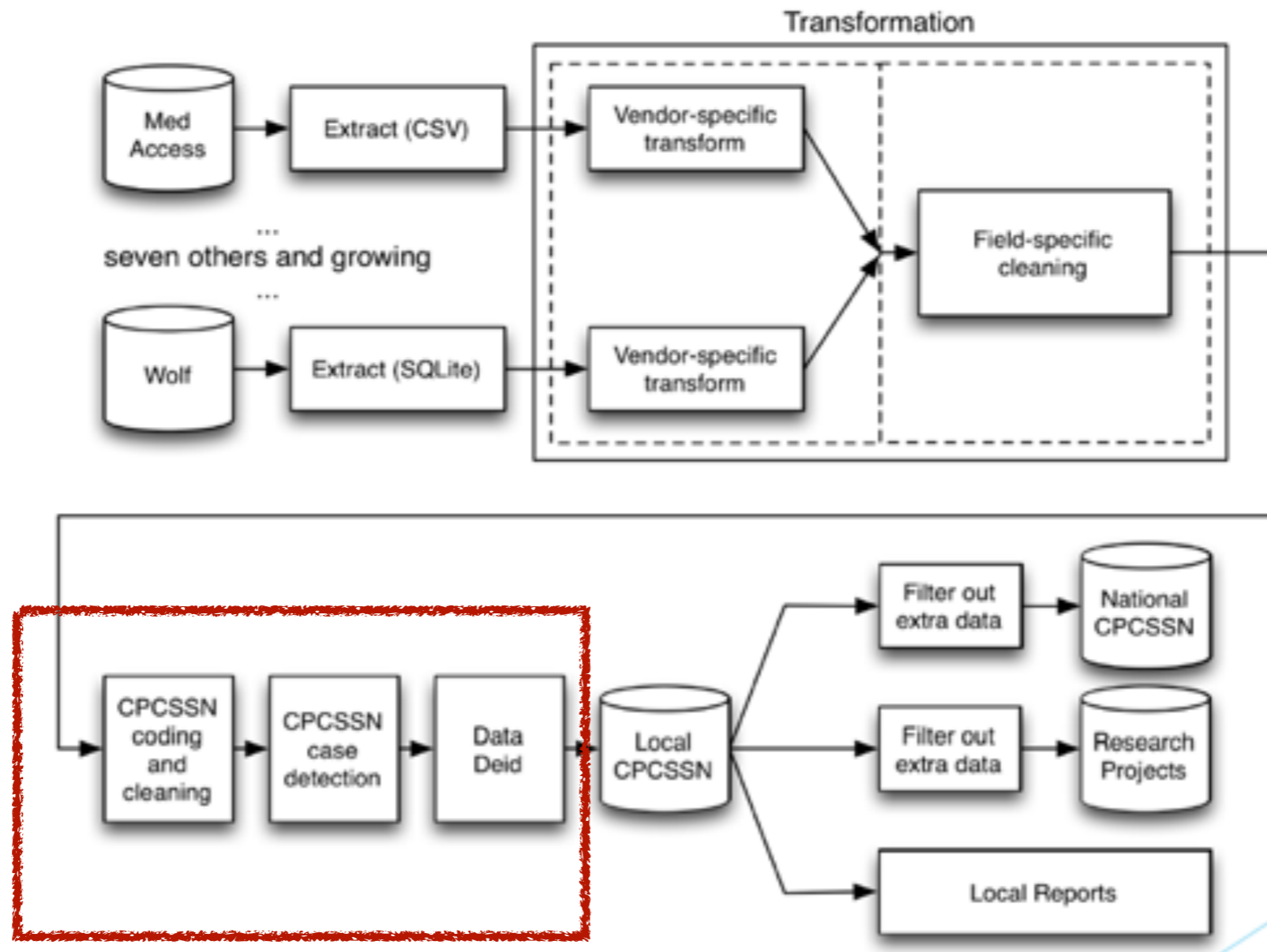
Extraction & Transformation



Extraction & Transformation

- Backend or front-end by vendor or CPCSSN data manager
- Select patients with assigned CPCSSN provider
- Null columns with identifiable information
- Create CPCSSN Mapping file
- Transformation: EMR structure to CPCSSN schema

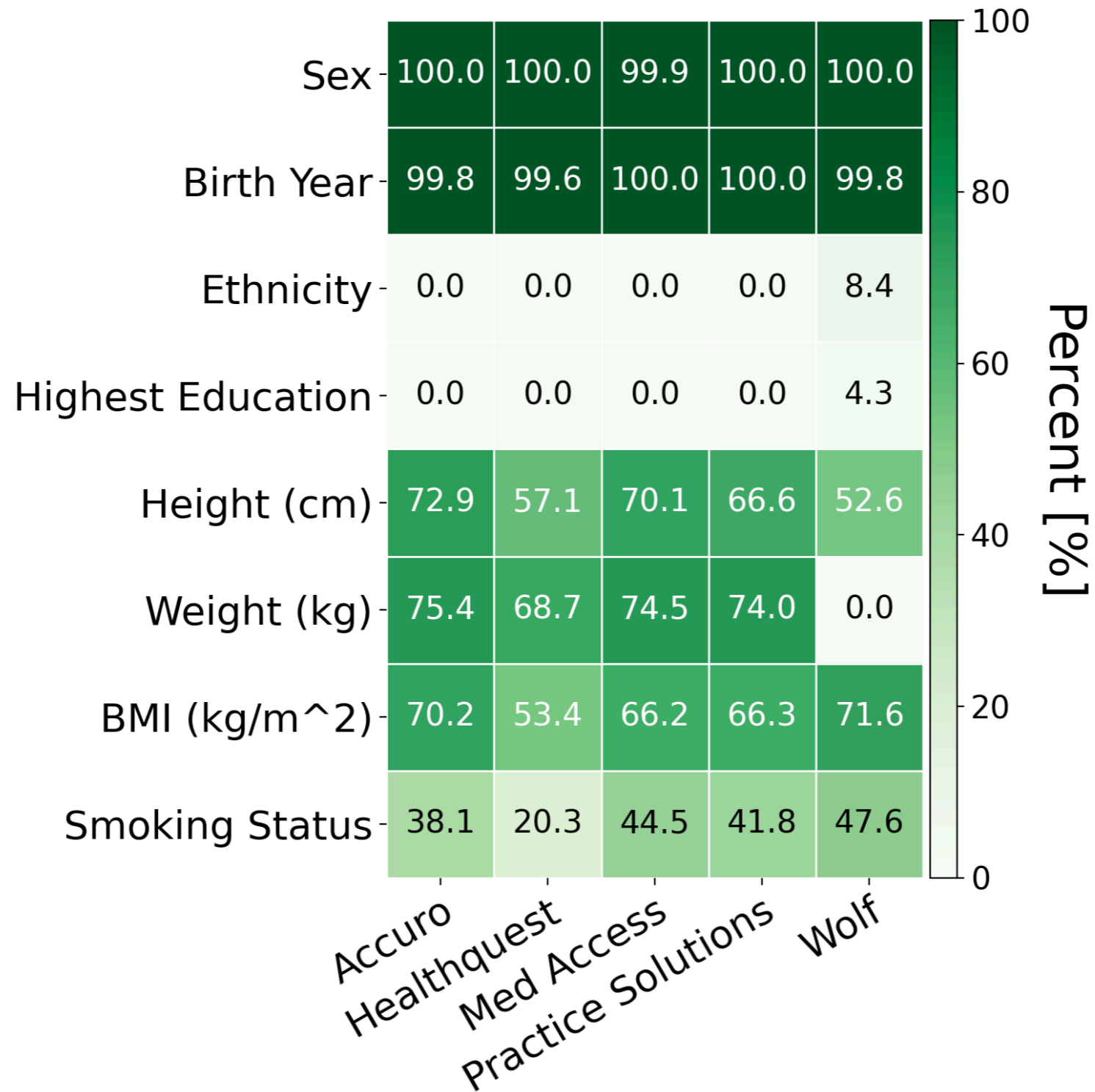
Data Cleaning and Coding



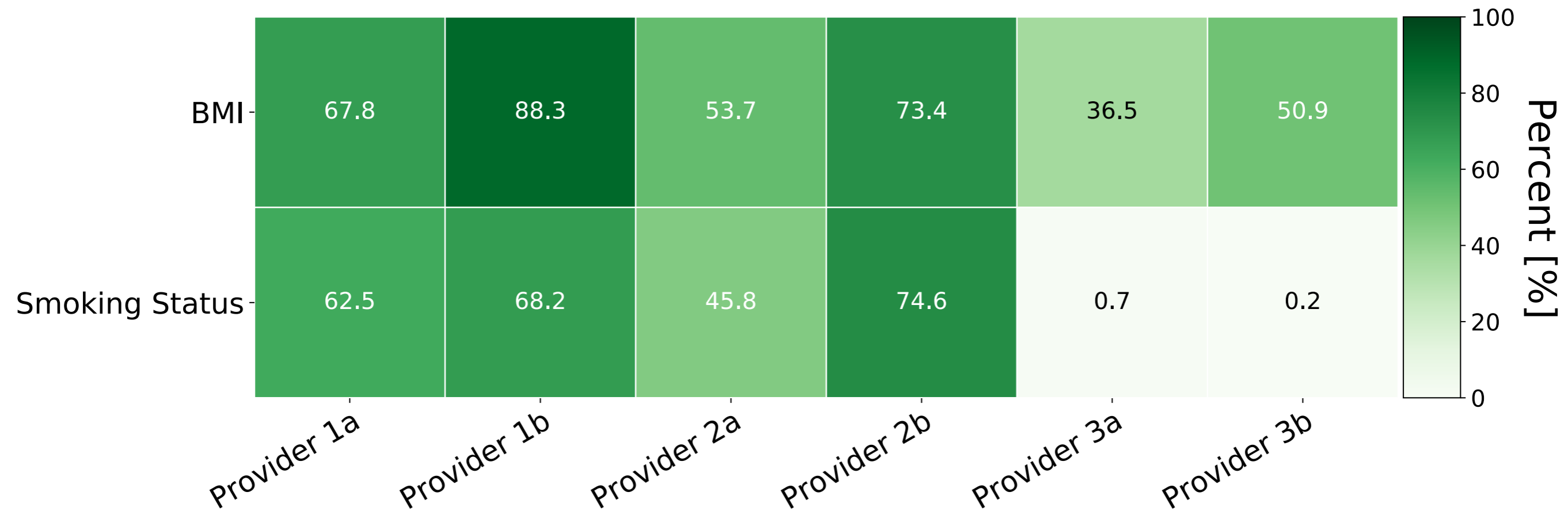
Data Cleaning and Coding

- **General data cleaning:** deleting empty, duplicate, or orphan records, standardizing dates & units, bounds checking
- **Text classification:** ATC, ICD-9, LOINC, risk factors
- **New variables:** BMI, deprivation scores
- **Disease case identification:** CPCSSN definitions for 17 health conditions (and growing)
- **Deidentification:** Remove direct identifiers (names, PHNs, phone numbers, etc.) from text

Completeness: EMR



Completeness: Clinic & Provider



Cleaning (Medication)

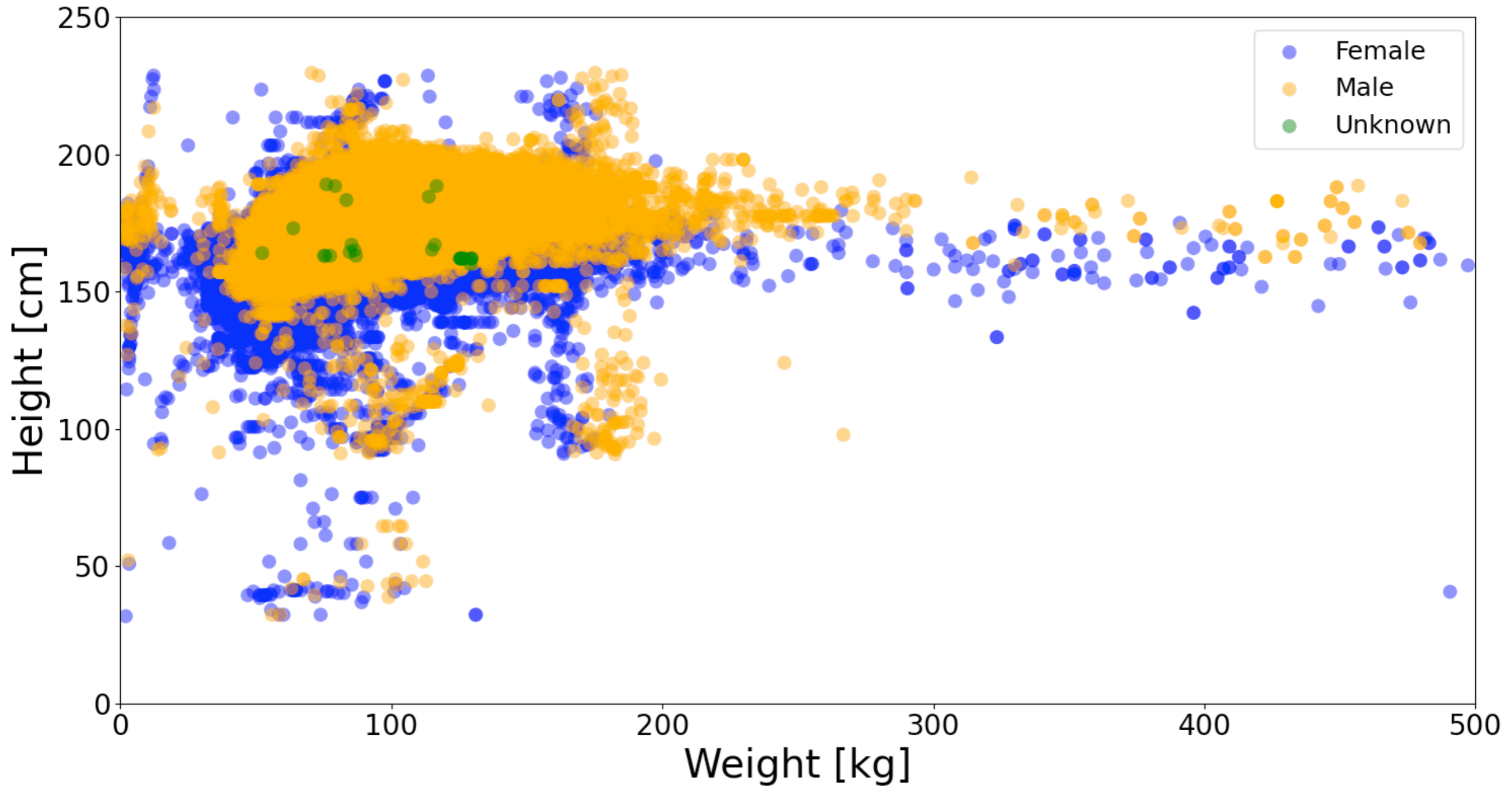
EMR Text	Cleaned Text	ATC Code
(Polytrim) drops 1 drop qhourly today then reduce to QID tomorrow	Combinations of Different Antibiotics	S01AA30
PERCOCET (Tabs) Sig 1 tab(s) Oral PRN if migraine Quantity 25 tab(s)	Oxycodone and Paracetamol	N02AJ17
TOUJEO SOLOSTAR 300 UNIT/ML (300/ML)	Insulin Glargine	A10AE04

Cleaning (Labs)

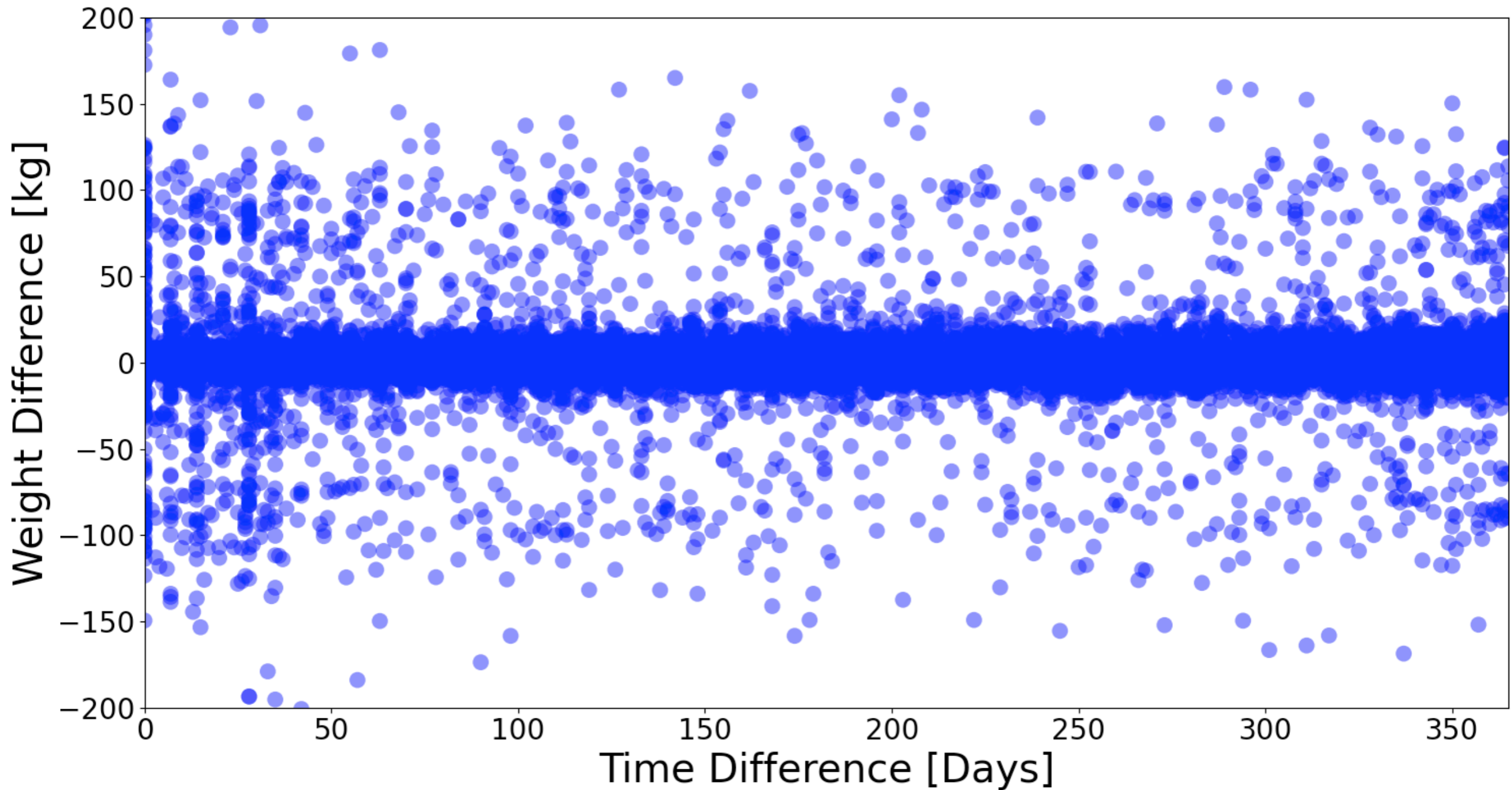
- CPCSSN standard text: Hemoglobin A1c (HbA1c)
- CPCSSN standard units: %

EMR Text	EMR Test Result	EMR Units	Cleaned Test Result
HEMOGLOBIN: HbA1C	0.065 0.0650	fraction	6.5
lipids and endocrine Hbg A1C	10.6 Recommended Targets: Males and non-pregnant females > 12 y old: <= 7.0% Pregnant adults: <= 6.0% Reference: 2003 CDA Guidelines. CJD 2003:27:S18	%	10.6
H??moglobine glyquée	<0.061 Ideal normal non-diabetic	None	<6.1

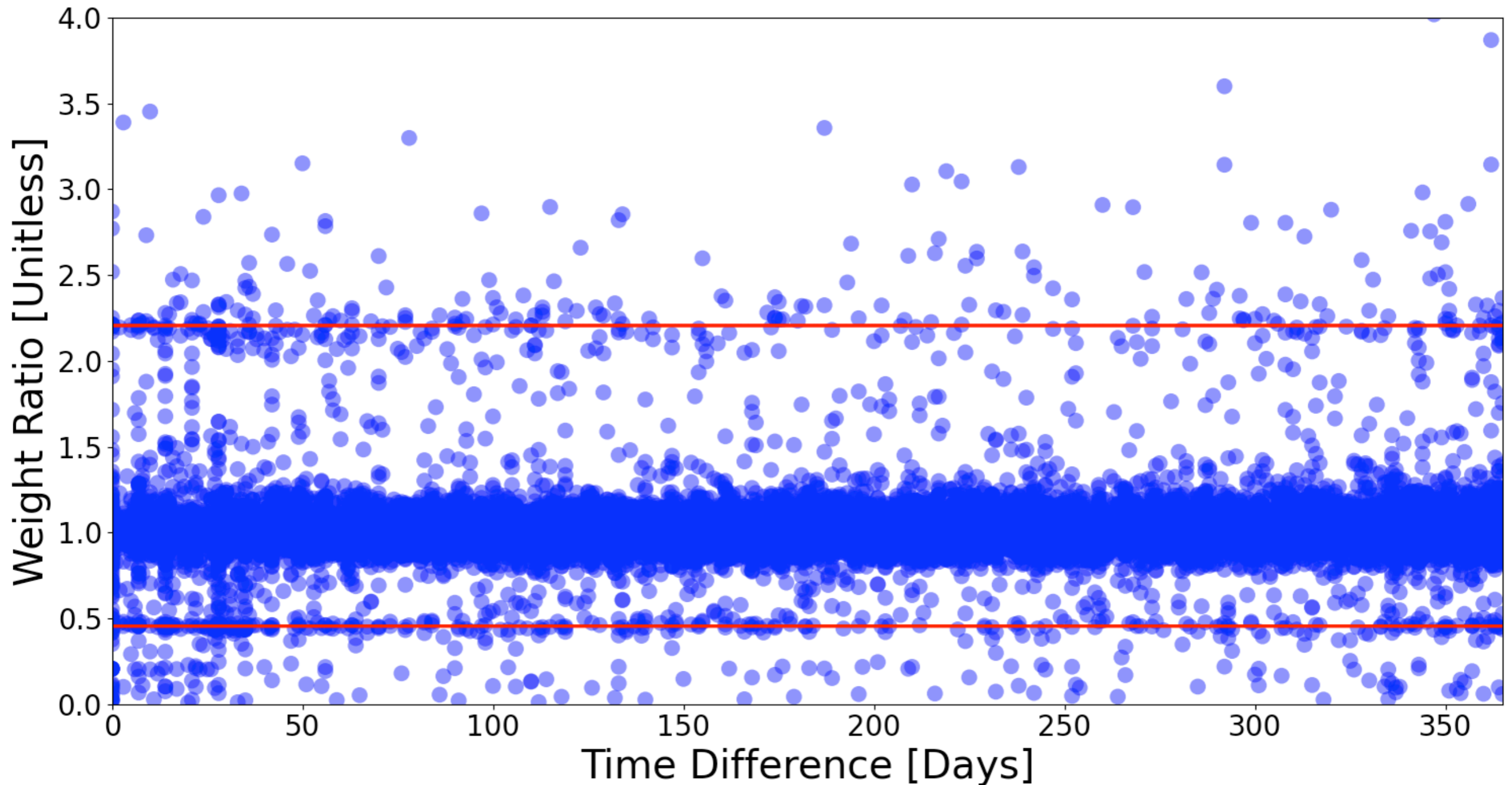
Plausibility: Height & Weight



Plausibility: Weight Change



Plausibility: Weight Ratio



Data Quality Summary

- **Documentation**
 - Processes for extraction, transformation, standardization
- **Completeness**
 - Different levels of granularity
- **Cleaning**
 - Single-step cleaning (e.g., Medication)
 - Multi-step cleaning (e.g., Labs)
- **Plausibility**
 - Outliers & anomalies

Thank you!



Questions or comments?



More information

- CPCSSN in Canada: cpcssn.ca
- CPCSSN in Alberta: cpcssn.alberta.ca
- Regional Primary Care Research Networks: napcren.ca, sapcren.ca



Contact

- Stephanie Garies: sgaries@ucalgary.ca
- Michael Cummings: michaelc@cpcssn.org