UNIVERSITY OF OSLO

OSLO UNIVERSITY HOSPITAL

One step closer to

Better Electronic Health Records data

Chi Zhang, PhD

OCBE (Oslo Center for Biostatistics and Epidemiology)
@ University of Oslo & Oslo University Hospital

PHUSE SDE Basel, 2024.9.18

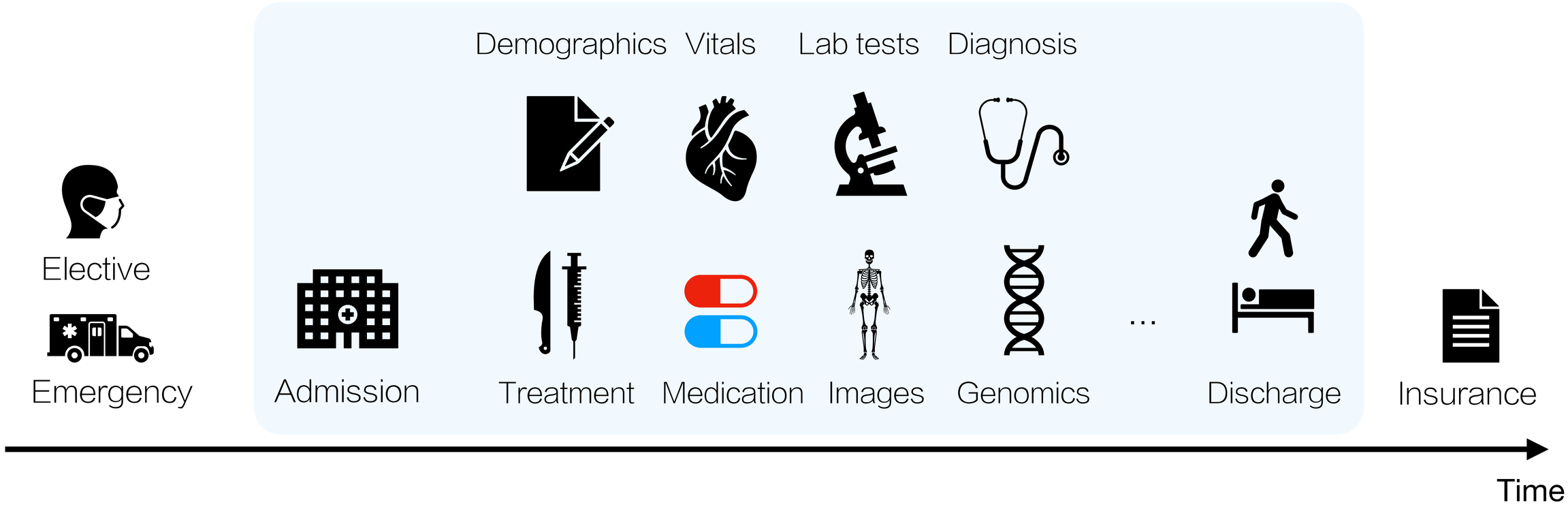
About me

- Statistician - RWD: **hospital EHR and large public health registries** (PhD on EHR)
- R developer, used **R packages and Shiny** for large-scale public health surveillance in Norway; **Quarto, webR** for teaching
- Co-lead of **CAMIS** (PHUSE DVOST project), contributor to **RWD guideline** for statistical programmer

Disclaimer: opinions in this talk are all mine and do not reflect that of my employer

Electronic Health Records

EHR data is generated through out the patient encounter with the healthcare system.



EHR and RWD

1960s
Development of IT systems in hospitals

...

2009
Wide adoption of modern EHR system begins

2016
Over 96% US hospitals have modern EHR.
MIMIC−III data available

2023
MIMIC−IV available

60 years history of Electronic Health Records (EHR) systems adoption in USA. Over 95% hospitals in Europe have EHR systems

2020s

Regulatory: FDA, EMA publish reports and guidelines on Real World Data and Evidence (RWE)

Research: causal inference, target trials, estimands with observational data

Industry: companies using RWD and AI for oncology, e.g Flatiron, Syneos

MIMIC databases are the largest publicly available EHR data from ICU (US data in Boston Area). Thousands of publications use these data

Google Trends Home Explore Trending Now

real world data
Search term

electronic health...
Search term

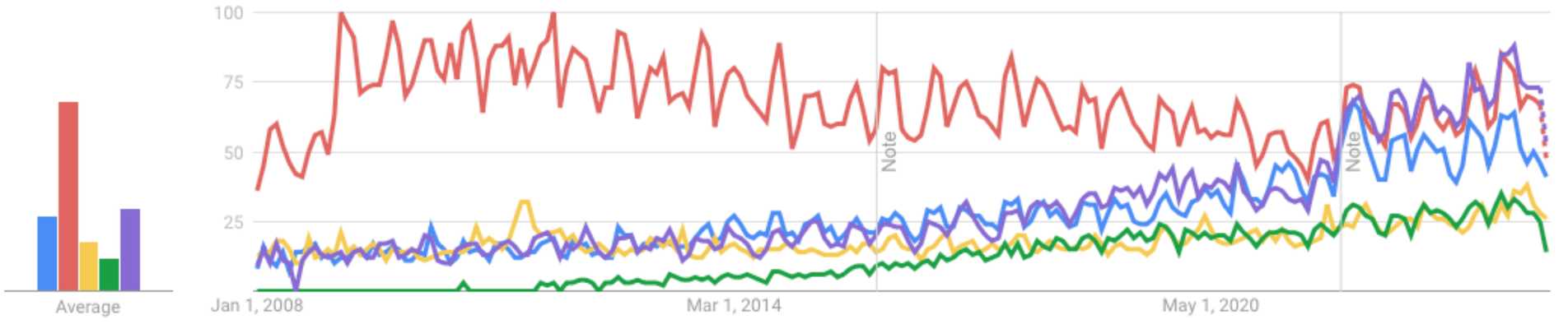
target trial
Search term

real world evide...
Search term

causal inference
Search term

Worldwide | 1/1/08 - 9/1/24 | All categories | Web Search

Interest over time



Why Better EHR?

This talk is not about modeling

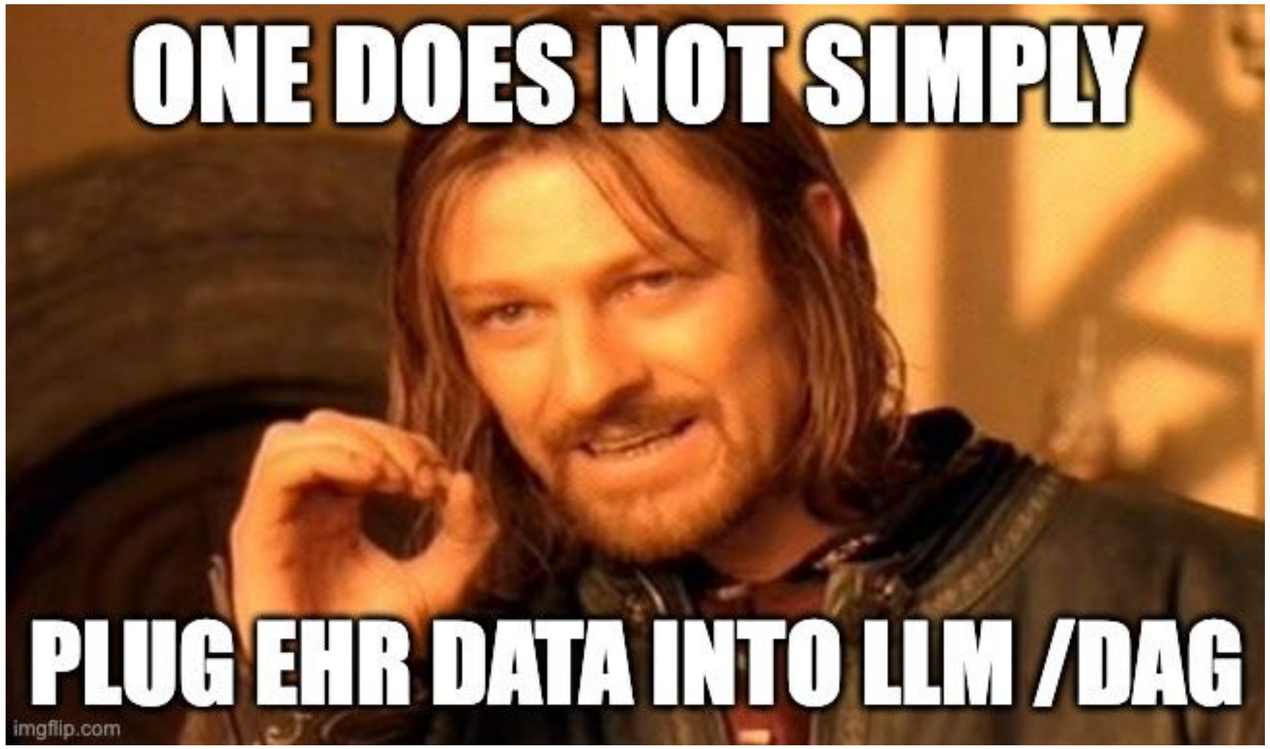
Machine learning with EHR has been going on for years

Prediction, classification (e.g. in-hospital mortality) -> **interpretable** ML

LLM

Statistical inference (causal or else)

Need **good data**!



Features of EHR

A few features of EHR data

Large scale	> 95% US and European hospitals; covers a variety of patients
Rich information	Demographics, diagnosis, lab tests, medication and procedures, ⋯ Exist in different formats (tabular data, text, waveforms, ⋯)
Temporal events	Detailed time stamps on measurements
Generated during clinical care	Originally for billing and legal purposes; different EHR systems co-exist

Not research quality data

Not RCT, observational in nature

Multi-modal, mixed type, temporal x static

Errors and missing

⋯

Case study:

Antibiotics stewardship at a Nordic hospital

EHR system in Norway: MetaVision

MetaVision from iMD-soft

Used by **27 countries, 400+ hospitals, 2M patients** per year (e.g. NHS in UK, Univeristy Hospital Geneva)

MetaVision in Norway

Hospitals under Helse Sør-Øst (**Southern and Eastern Norway** Regional Health Authority), cover over 57% of total population of Norway

2008 until now

40k users (physicians, nurses)

~1.000 requests to add features (e.g. new drugs, new settings) per year

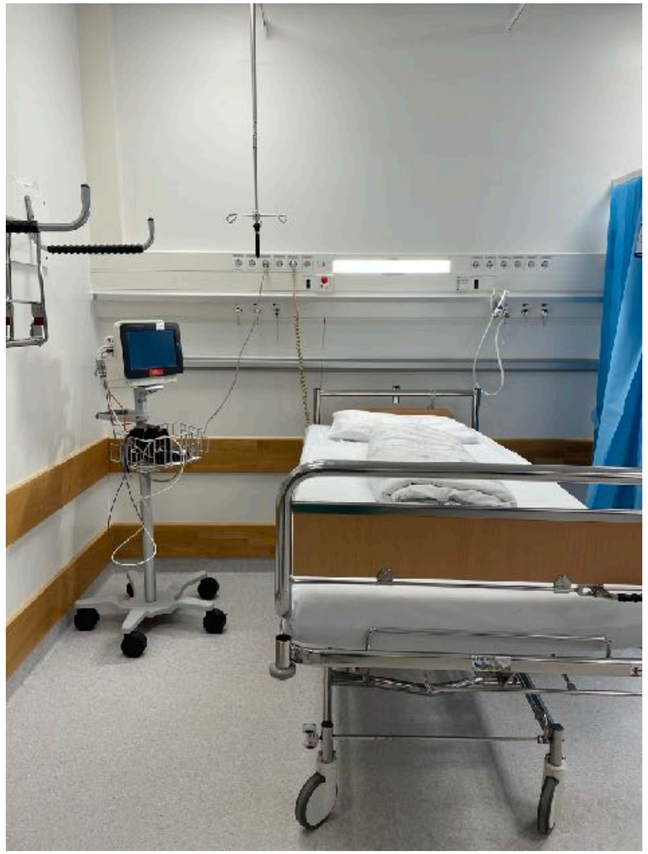
~80% granted (regular updates by Helse Sør-Øst)

https://www.imd-softde.de/

MetaVision at Oslo University Hospital

Oslo University Hospital, Ullevål

MV connected to a bed



Emergency room

MetaVision at Oslo University Hospital

Anesthesia, ultrasound to MV

Blood gas machine, connected to MV directly

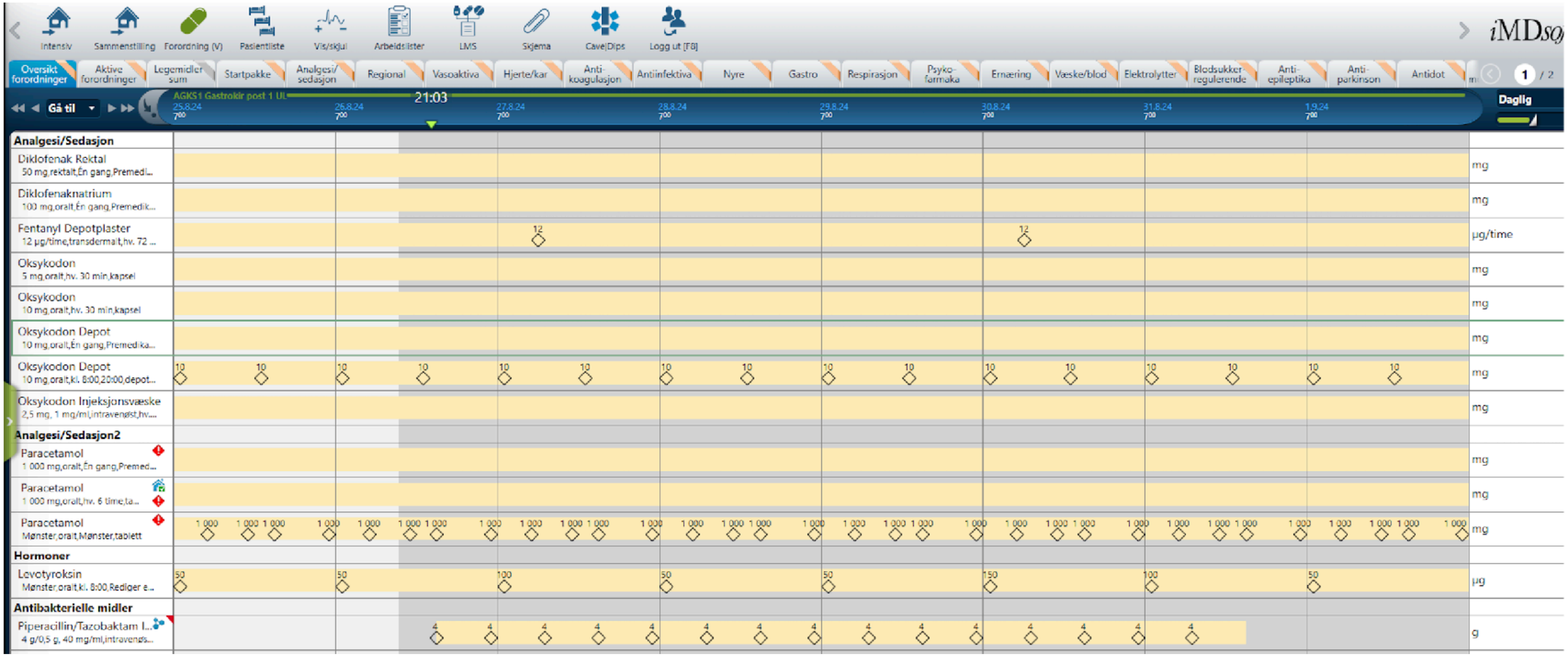
MetaVision at Oslo University Hospital

Operating room

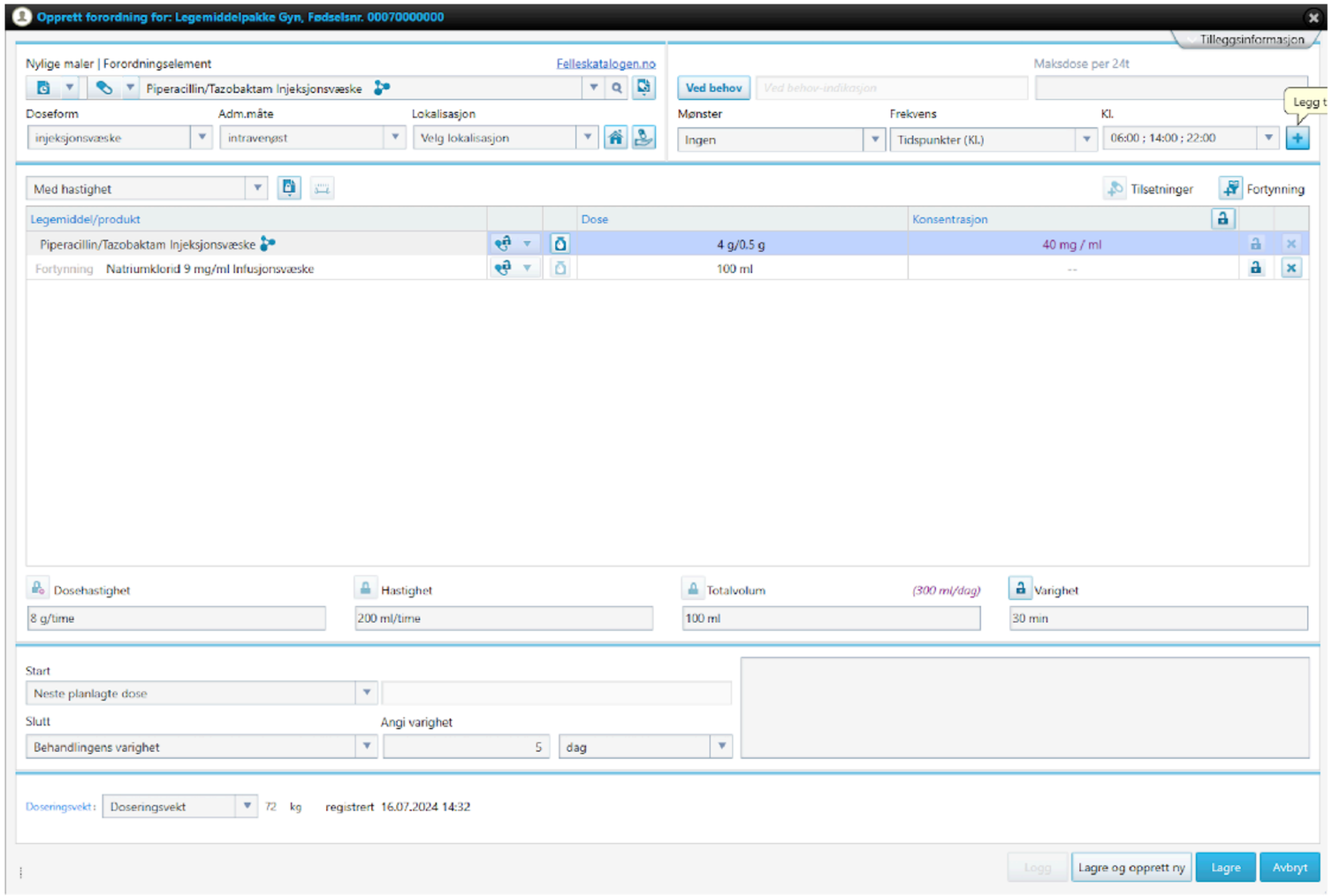
Every physician and nurse have access to MetaVision at a **hospital computer**

Used to **record procedures and treatment**, as well as **monitoring** the physiological status

Medication records in MV (pseudo patient data)



Medication records in MV (pseudo patient data)



Drop-down menus to select medication

Prescription and **use** are recorded separately

Pre-configured settings for

- Drug
- Dose
- Duration

From MV to data analysts

PI makes application to REK

Regional Committees for Medical and Health Research Ethics

Data **anonymization and extraction**, produce data in xlsx, csv

PI uploads data into **TSD (Service for Sensitive Data)**, used across Norway. TSD is disconnected to internet

Analysts work on data **inside TSD**, R packages and other softwares need to be installed as mirrors

Results (data, reports, scripts) **can only be exported by PI**

Case study: antibiotics stewardship

We acquired data from Akershus University Hospital in Greater Oslo Area (same EHR system)

One year (2018-2019), **hourly in-patient data** for 2 surgical wards (gastrosurgical, orthopaedics); 2 medical wards (neurology, gastroenterology), ~ 17k unique hospital stays

Initially data was extracted for a different project. This is only an example for antibiotics, but is same logic can be applied to *any other medication* at hospitals

Effective antibiotics stewardship is key to prevent **antimicrobial resistance** and improve **patient outcome** (shorter stay)

How is antibiotics (AB) **prescribed** and **used** at a hospital setting? Quality Assurance

- **plan vs adherence**: time, dose
- **purpose** for different kinds infection: preventative (prophylactic), hospital acquired, community acquired?

First glimpse: prescription

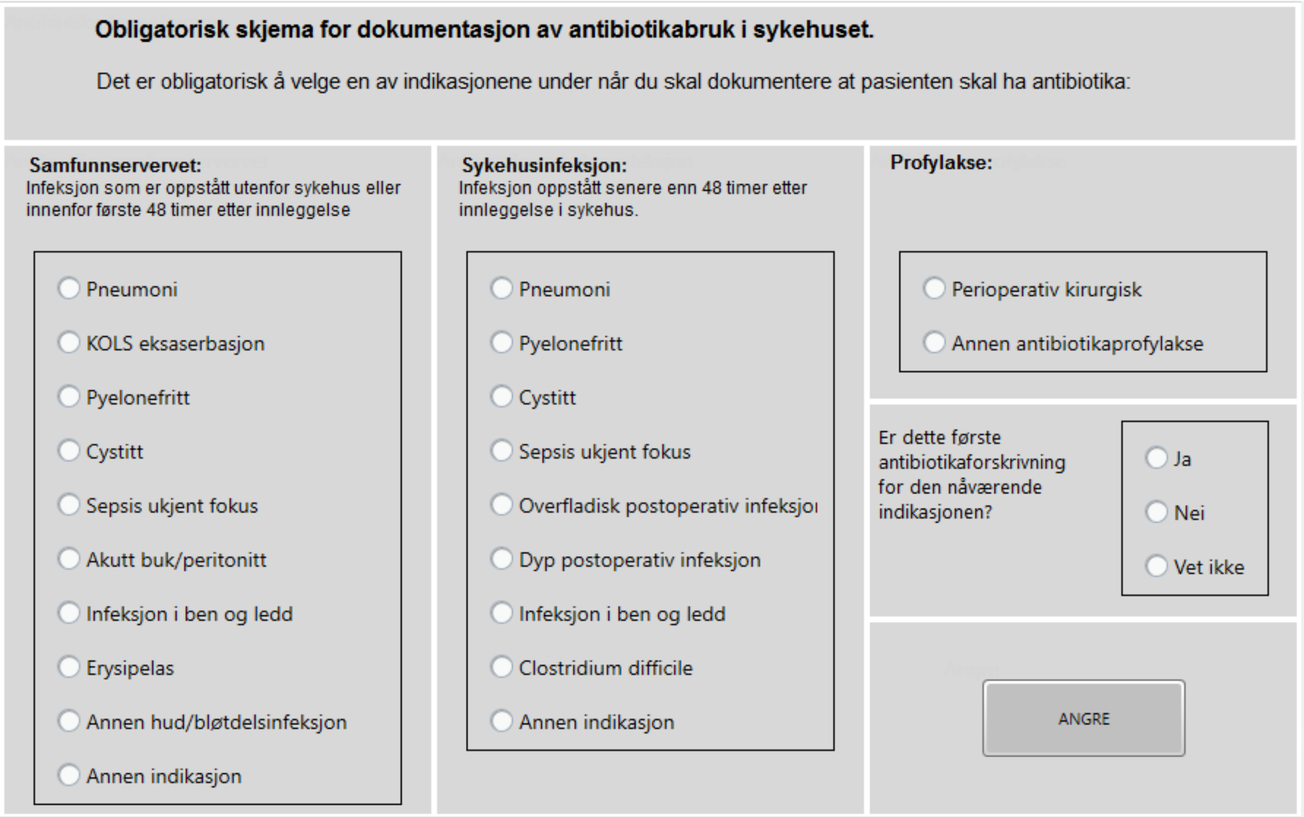
Every time an AB is prescribed (in MetaVision), a form pops up to ask the physician:

- *what the AB is for,*
- *is it the **first time** this AB is prescribed for "this indication"*

Community acquired pneumonia,
Surgical **prophylactic**,
Hospital acquired sepsis, …

Ideal: mutually exclusive, informative selections

Execution: ?



First glimpse: prescription

Analysed 7327 patients from four wards who had AB prescriptions

25 different AB

Did not fill in the pop-up form but *got away*: 144, as many elective as emergency

<u>What is the AB for?</u>
19 have **multiple purposes** for the same drug, 3/4 from emergency

<u>First time for this AB?</u>
mixture of ‘Yes’, ‘No’, **>1000 missing (1 in 7)**

Not un-usable, but makes you wonder …

B	C	D	E	F	G	H
KontaktID	AB_SkjemaTidRelativ(Timer)	AntibiotikaSkjema	AB_ForsteGang	Profylakse	Samfunnservervet	Sykehuservervet
11439	-30	Ampicillin	Ja		Akutt buk/peritonitt	
11439	-30	Gentamicin	Ja		Akutt buk/peritonitt	
11439	-30	Metronidazol	Ja		Akutt buk/peritonitt	
11439	-24	Ampicillin			Akutt buk/peritonitt	
16406	2607	Cefalotin	Ja	Perioperativ kirurgisk		
16406	2609	Cefalotin	Nei	Perioperativ kirurgisk		
16406	2612	Cefalotin	Nei	Perioperativ kirurgisk		
13549	829	Klindamycin	Ja	Perioperativ kirurgisk		
13549	830	Klindamycin	Nei	Perioperativ kirurgisk		
12577	2605	Cefalotin		Perioperativ kirurgisk		
12577	2606	Cefalotin	Nei	Perioperativ kirurgisk		
10245	6858	Metronidazol		Perioperativ kirurgisk		
11341	5194	Cefalotin		Perioperativ kirurgisk		
11341	5196	Cefalotin	Nei	Perioperativ kirurgisk		
338	4940	Pivmecillinam	Nei		Cystitt	
17279	2561	Kloksacillin	Nei		Infeksjon i ben og ledd	
17279	2563	Kloksacillin	Ja	Perioperativ kirurgisk		
13722	2257	Piperacillin/Tazobaktam			Akutt buk/peritonitt	
13722	2337	Ciprofloksacin			Akutt buk/peritonitt	
13722	2337	Metronidazol			Annen infeksjon	
13722	2337	Trimetoprim/Sulfametoksazol	Nei		Annen infeksjon	
13722	2361	Metronidazol	Nei		Annen infeksjon	
4144	5428	Cefalotin		Perioperativ kirurgisk		
4144	5433	Cefalotin	Ja	Perioperativ kirurgisk		
1808	5462	Fenoksymetylpenicillin	Ja		Pneumoni	
1808	5463	Benzylpenicillin	Ja			Pneumoni
13285	5364	Klindamycin	Ja	Perioperativ kirurgisk		

Compare with actual usage records

A different MetaVision form for the “processes”:

AB type, injection/pill, doses, time

(Same table also has all the catheters)

3.4% of all records have drug **usage before prescription**

1/3 differ by one hour
1/5 has the prescription time after the last dose - **very delayed**

~15% AB have **mismatches** in prescription and usage recording: prescribed, never used; vice versa.

A	B	C	D	E	F	G
Kontak	PID	Value	ParameterName	TidTilProsess	Forbruk	Benevnelse
10009	1	1.0000000	#NAME?	7978	1889.000000000(	Min
10009	1	1.0000000	#NAME?	7978	67.000000000000	Min
10009	1	1.0000000	- PVK (1,3 mm/18 G)	7978	88.000000000000	Min
10009	1	1.0000000	- PVK (1,3 mm/18 G)	7942	2040.000000000(	Min
10009	1	1.0000000	- PVK (1,1 mm/20 G)	7976	2004.000000000(	Min
10009	1	1.0000000	- PVK (1,1 mm/20 G)	7976	2004.000000000(	Min
11439	3	0.0187500	Metronidazol Injeksjonsvæske	-29	1.5000000000000	Gram
11439	3	0.0250000	Metronidazol Injeksjonsvæske	-17	1.0000000000000	Gram
11439	3	0.0250000	Metronidazol Injeksjonsvæske	30	1.0000000000000	Gram
11439	3	0.0083333	Gentamicin Injeksjonsvæske	-6	0.5000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	17	2.0000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	23	2.0000000000000	Gram
11439	3	0.0666666	Ampicillin Injeksjonsvæske	30	2.0000000000000	Gram

Reflection

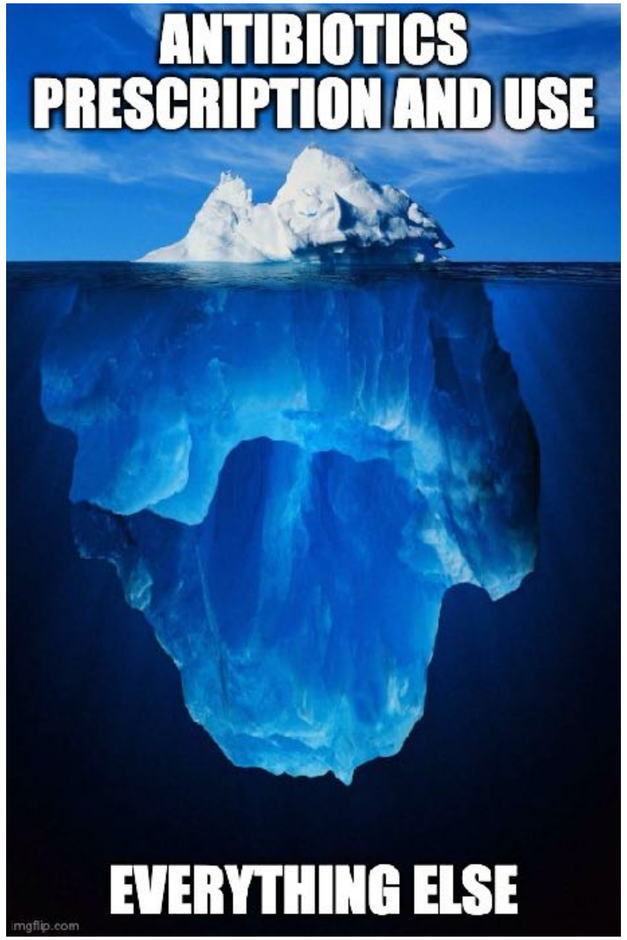
Norway has fairly **good antibiotics stewardship** at hospitals.

Physicians and infection control team have a decent idea of what is going on at the hospital.

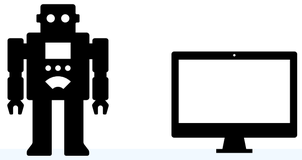
But **analysts (statisticians, data scientists, ML engineers)** don’t!

Can we trust and use the data?

What are the possible causes for **errors, mismatches and missingness**?



Can we do better?

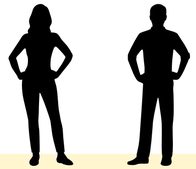


Clearer question prompts
- ‘What do you mean by first AB for this indication?’

Improve prescription workflow
- mutually exclusive selection for drug purpose, make it impossible to have confusing records

Re-think the process
- if one can avoid filling in the form - is it still necessary to have the form?

Reduce machine errors
- outliers caused by removing the equipment (spikes in data time series)
- daylight saving time related



Fatigue, busy schedule

‘*I’m saving lives, I don’t have time to fill-in all the forms*’

‘*Just want to get rid of the prompt screen*’

Forms filled in by **different teams**, different departments have their own ways of doing things

Not error, just how it is

An R package for EHR visualization

ggplot2 extension for EHR data

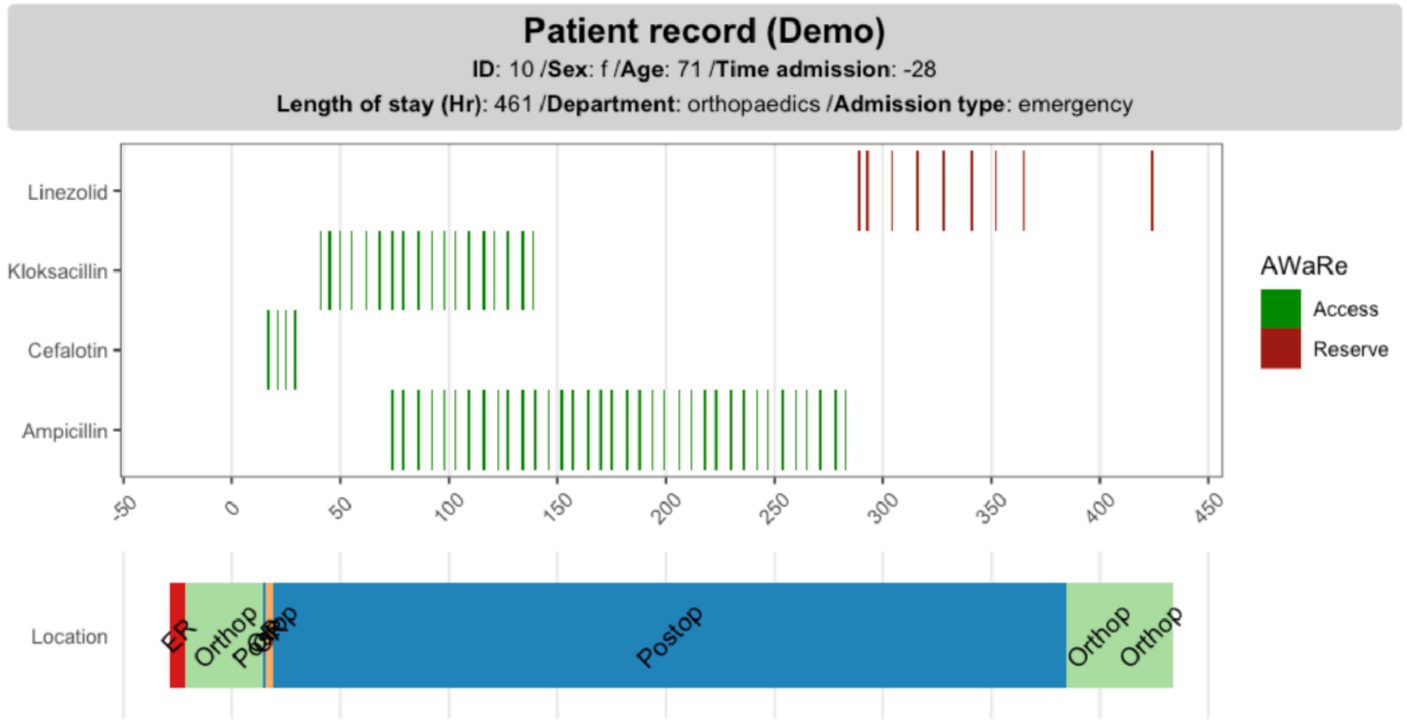
gg E-H-R

A set of R tools to visualize **mixed type of information** related to a patient

- **static** data: demographics, admission information

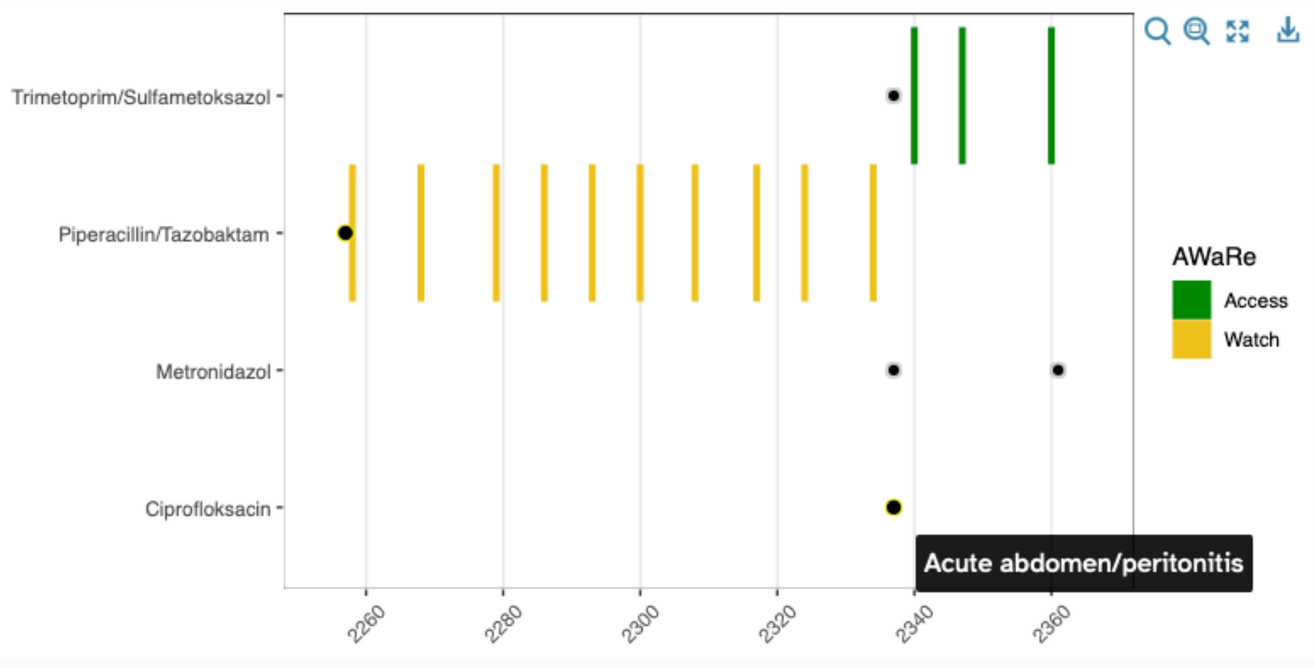
- **temporal**: treatments (drug, procedure), locations

Can add **more layers**, such as **locations**, physiological time series (**temperature**, **NEWS score**) and **catheters**, etc



ggplot2 extension for EHR data

```
p1i <- plot_drug_prescribe_interactive(data_prescribe = dpatient_presc,
                                       plot_obj = p1p)
p1i
```



Drug prescription and use plotted together

Drugs (e.g. antibiotics) colored based on **WHO AWaRe** categories

Can visualize prescription time and purpose, **interactively**

Visually **identify if errors/strange things** exist in data recording

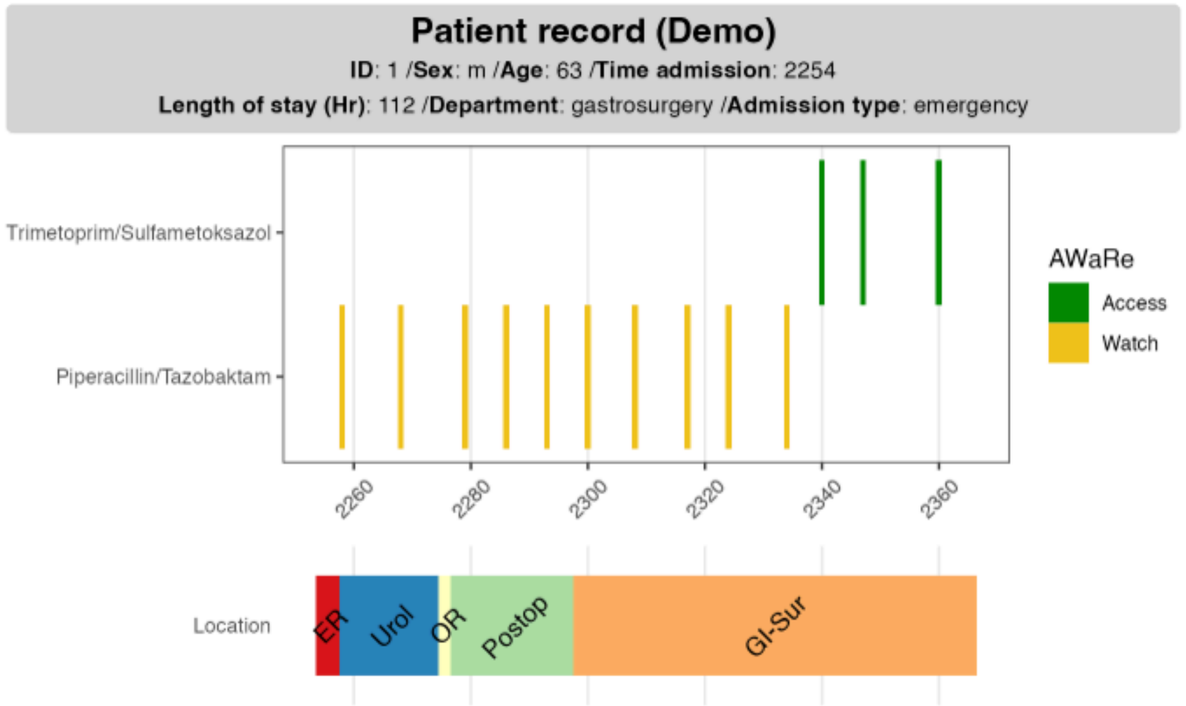
Can also identify if the use of drug has **escalated**

AWaRe: Access, Watch, Reserve - unrestricted use, use with care, use as last resort

ggplot2 extension for EHR data

Static information card provides patient background

Mixed type information all in one graph, easier communication with physicians



Next step: aggregated **visualization for a cohort** -> assist causal hypothesis generation

Better EDA -> better research questions

Why not shiny or quarto?

- limited resources, maintenance
- **not yet production level**
- Rpkg can be used by everyone

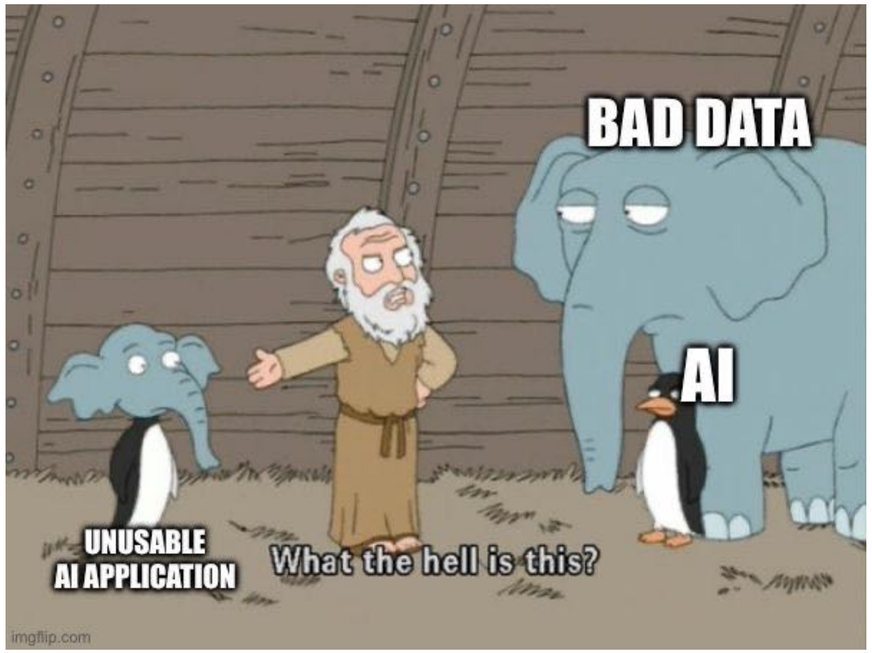
Under active development, openly accessible on GitHub

I'll give a talk at R/Pharma Oct 2024 on ggehr!

One step closer to better EHR

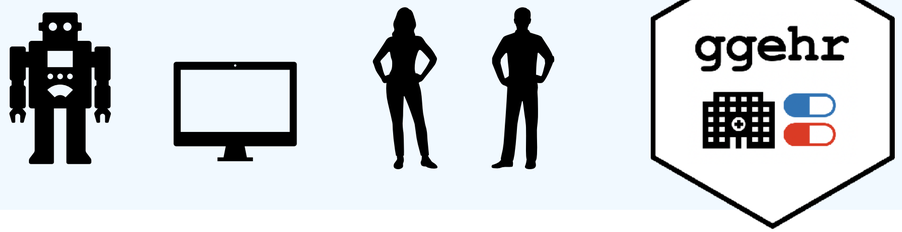
RWD data such as EHR are getting more and more important

Quality of data >> model (LLM, statistics)



Errors and confusions happen. Do your EDA!

- **Better vendor engagement**, improve the systems
- **Work together**: domain experts and data scientists



Looking for new opportunities as data scientist / R dev / statistician. Let's chat!

Website: https://andreaczhang.github.io/
LinkedIn: Chi Zhang
Email: chizapoth@gmail.com