1. Versioned OHDSI docker library
2. Argos Project

Seng Chan You

Mission, Vision, and Values of OHDSI

- Our Mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

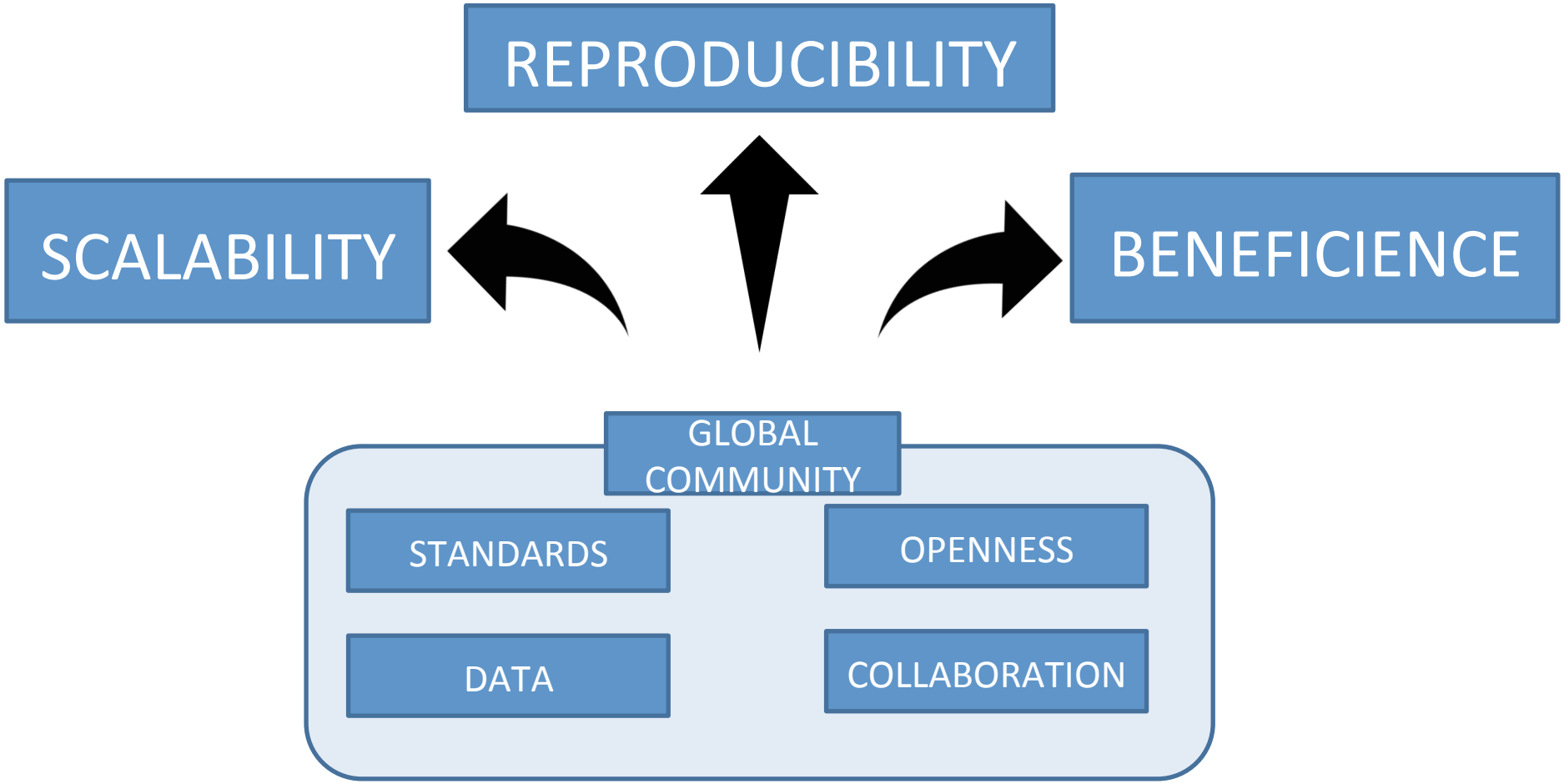
- Our Vision

A world in which observational research produces a comprehensive understanding of health and disease.

Mission, Vision, and Values of OHDSI

- Innovation: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.
- Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.
- Community: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.
- Collaboration: We work collectively to prioritize and address the real world needs of our community's participants.
- Openness: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.
- Beneficence: We seek to protect the rights of individuals and organizations within our community at all times.

OHDSI: Open Innovation based on the open community



Reproducibility in OHDSI research

Comparative Effectiveness Study of Febuxostat versus Allopurinol in Gout

Researchers

SCYou Seng Chan You 1 Sep '18

Dear all,

The new network study is launched to evaluate the efficacy and safety febuxostat in gout compared to allopurinol.

Comparative Effectiveness Study of Febuxostat versus Allopurinol in Gout

Objective: The goal of this protocols is conducting comparative effectiveness research to establish evidences for benefits and harms of febuxostat and allopurinol. The primary endpoint is the risk of sudden cardiac death. The secondary endpoints include acute myocardial infarction, stroke, heart failure, gout flare and drug hypersensitivity (TEN, SJS, and DRESS)

Rationale: Febuxostat is widely used urate-lowering agent because it is more effective than allopurinol to lower serum urate in patients with gout. Furthermore, febuxostat can be used without dosage adjustment in chronic kidney disease. The Cardiovascular Safety of Febuxostat and Allopurinol in Patients with Gout and Cardiovascular Morbidities (CARES) group was a prospective multicenter, double-blind randomized clinical trial, which assessed the cardiovascular risk of febuxostat compared with allopurinol in patients with gout and a history of CVD. This study concluded that febuxostat was associated with significantly higher overall and cardiovascular mortality compared to allopurinol, mostly driven by sudden cardiac death. Still, there is scarce evidence for risk of sudden cardiac death between febuxostat and allopurinol in real-world practice.

The whole protocol is released at github

GitHub 10

OHDSI/StudyProtocolSandbox

This repository is for developing study packages for OHDSI studies. Once completed, they can be moved to the StudyProtocols repository. - OHDSI/StudyProtocolSandbox

Reproducibility in OHDSI research

The whole protocol is released at github

GitHub 10

OHDSI/StudyProtocolSandbox

This repository is for developing study packages for OHDSI studies. Once completed, they can be moved to the StudyProtocols repository. - OHDSI/StudyProtocolSandbox

Ian Holmes
@ianholmes

팔로우

You can download our code from the URL supplied. Good luck downloading the only postdoc who can get it to run, though #overlyhonestmethods

오전 8:52 - 2013년 1월 8일

https://twitter.com/ianholmes/status/288689712636493824

Reproducibility in OHDSI research

2 MONTHS LATER

George_Argyriou 1 ✎ 27d

Hi Chan,

I have difficulties installing the package in my local machine.

I've tried to install using two different ways:

1. With **install_github("chandryou/FebuxostatVsAllopurinolCVD")** I get:

Error in read.dcf(path) :

Found continuation line starting ' DatabaseConnecto …' at begin of record.

2. With **install.packages("https://github.com/chandryou/FebuxostatVsAllopurinolCVD.git** 1 **")** I get:

Warning in install.packages :

package 'https://github.com/chandryou/FebuxostatVsAllopurinolCVD.git' is not available (for R version 3.5.1)

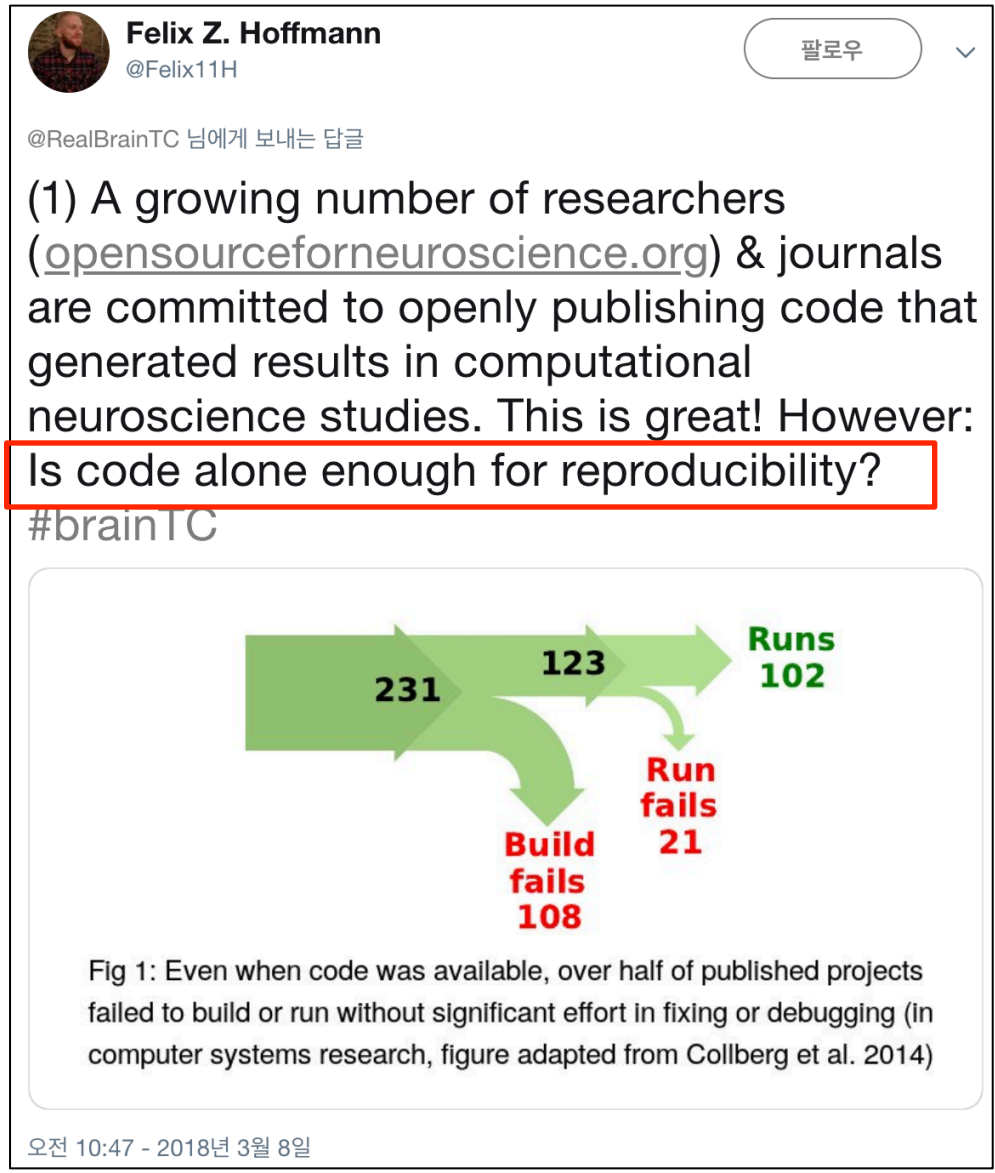
Can you help?

Reproducibility

Table 2: Build error summary.

error	count	percentage
incomplete documentation	10	7.9%
none	1	0.8%
distribution is missing files	21	16.7%
unavailable environment	14	11.1%
missing third party package	22	17.5%
other errors	22	17.5%
prerequisite failed to build	23	18.3%
runtime error	12	9.5%
internal compiler error	1	0.8%
Total	126	

Collberg et al., Measuring Reproducibility in Computer Systems Research



https://twitter.com/Felix11H/status/971819728040808448

Broadsea

Broadsea – The OHDSI Open Source Standard Software Stack Packaged as Docker Container Images for Cross-Platform Installation

Lee D. Evans[1], Marc A. Suchard, MD, PhD[2], Jon D. Duke, MD, MS[3]
[1]LTS Computing LLC, West Chester, PA; [2]Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA; [3]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN;

Abstract

We packaged the OHDSI open source standard software stack into Docker containers with the aim of simplifying cross-platform OHDSI software installation on a range of Operating Systems, DBMSs and infrastructure. We believe this simpler software deployment option will help encourage OHDSI Community members to download and install the full OHDSI software stack for research on their own CDM databases and more easily participate in OHDSI Network studies.

The software container image build process and the OHDSI container images are collectively known as Broadsea. The OHDSI Broadsea Docker containers may be configured to connect to an OMOP Common Data Model Version 5 database.

OHDSI symposium, 2016

LTS Broadsea API

LTS Broadsea API – OHDSI methods as a service

Lee D. Evans
[1]LTS Computing LLC, West Chester, PA

Abstract

The LTS Broadsea API provides access to OHDSI methods as a service in the cloud. For example, run Achilles on any OMOP CDM dataset in any OHDSI supported database within Amazon AWS and Google Cloud by making a simple secure web service call referencing the database connection details. The Broadsea API is currently under development.

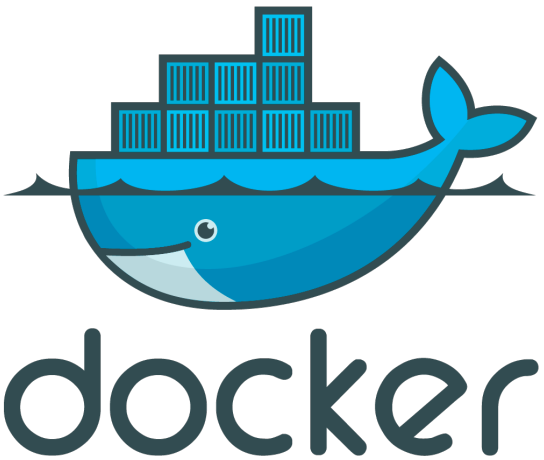
Introduction

Many organizations are taking advantage of managed database services like AWS RDS, Redshift, Google BigQuery and Hadoop to host their observational databases and convert them into the OMOP Common Data Model.

However, running the OHDSI methods on that data is not so simple. It requires additional skilled resources to deploy, administer and upgrade the required web servers, R servers, Proxy servers and middleware (tomcat, rstudio, docker, etc).

The aim of the Broadsea API is to provide simple, immediate access to the OHDSI methods as a cloud service via a simple REST API service that can be called from any web application/service.

Docker, What is it? and What for?



What is Docker?

- Docker is an **open platform** for developers and sysadmins **to build, ship, and run distributed applications**. Consisting of Docker Engine, a portable, lightweight runtime and packaging tool, and Docker Hub, a cloud service for sharing applications and automating workflows, Docker enables apps to be quickly assembled from components and eliminates the friction between development, QA, and production environments.

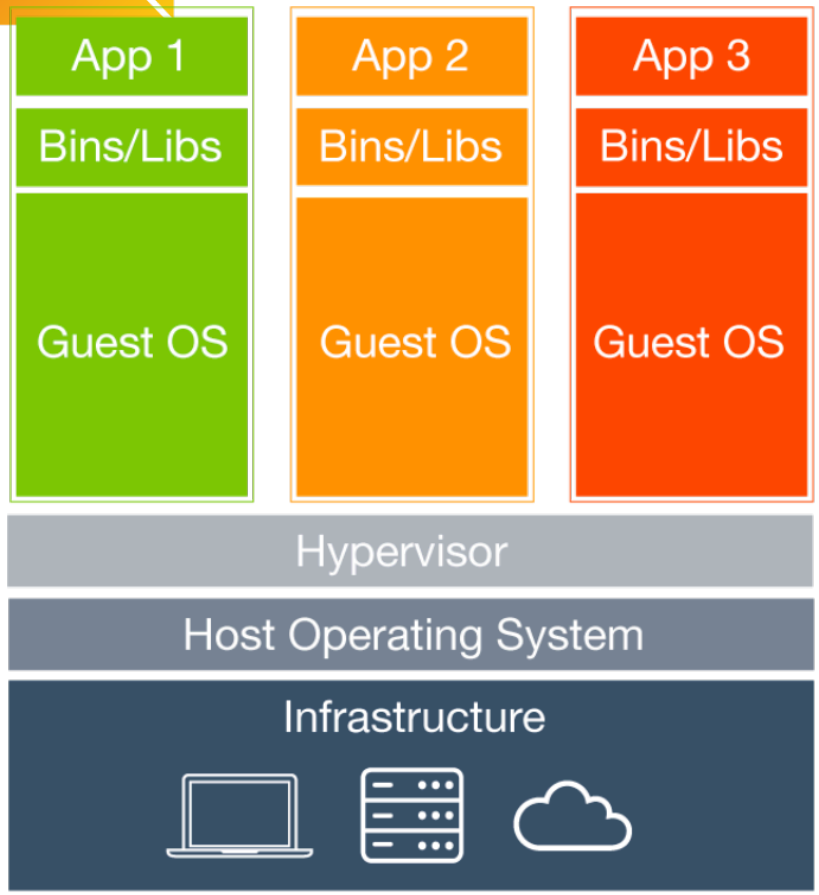
Text from http://docker.com as of June 24, 2014.

Ok, seriously, what *is* Docker?

- Docker is a **very lightweight abstraction** using recent Linux kernel features which lets us to run code in **cheap** (to launch) and **easy** (to build) units: containers
- We can share containers across OSs, across time.

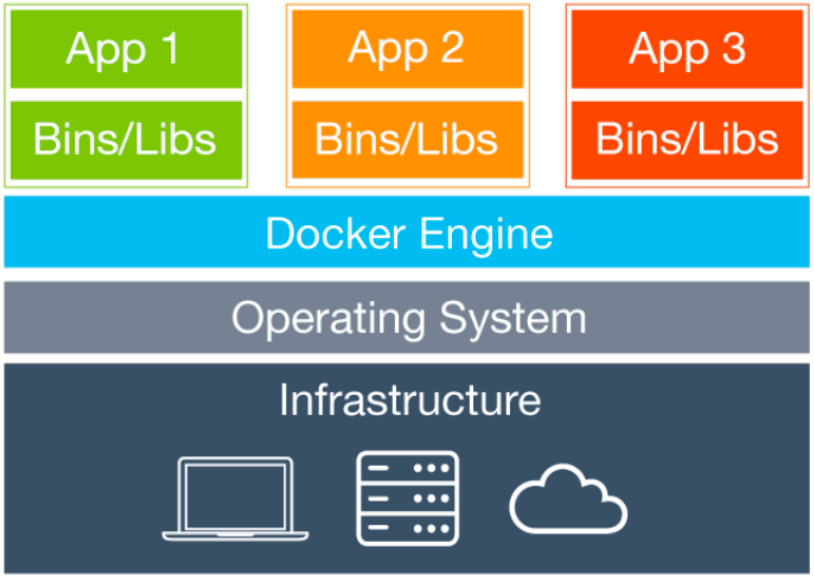
Text from http://dirk.eddelbuettel.com/papers/r_and_docker_sep2014.pdf as of Mar 11, 2019.

Virtual Machine vs Docker



Virtual machines

Virtual machine runs one complete OS on top of another OS



Containers

A Docker container is like a virtual machine that shares guest OSs, which makes them very lightweight

https://matsen.github.io/2018/04/19/docker.html

Docker and **Reproducibility**

- Capturing the computational environment
  - A **substantial challenge** in reproducing analysis is **installing and configuring the web of dependencies of specific versions** of various analytic tools.
  - Popular VM applications include VirtalBox and VMWare. One challenge of working with VMs is that the files that contain the environment are not small, typically one gigabyte or more, which can be awkward to share.

From http://ropensci.github.io/reproducibility-guide/sections/introduction/

Docker Advantage for **Reproducibility**

- Small footprint
- Easier deployment
- Easier sharing and publication
- Open source platform
- Standard scripting of image setup with Dockerfile
- Rocker images as baseline

From http://ropensci.github.io/reproducibility-guide/sections/introduction/

http://www.cascadia-analytics.com/docs/user2018-docker-r-tutorial-slides.pdf

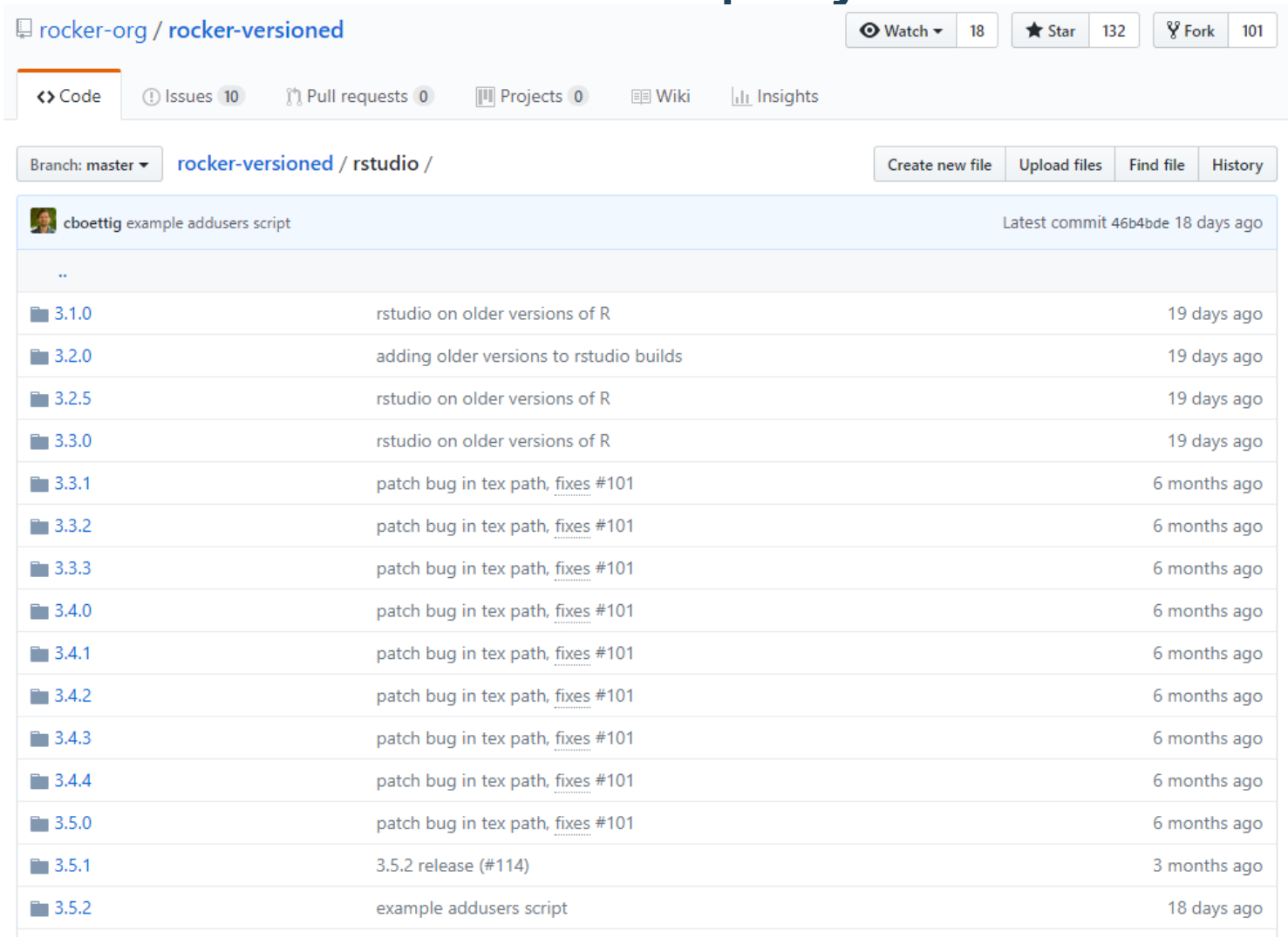
Docker Advantage for **Reproducibility**

- Small footprint
- Easier deployment
- Easier sharing and publication
- Open source platform
- **Standard scripting of image setup with Dockerfile**
- **Rocker images as baseline**

From http://ropensci.github.io/reproducibility-guide/sections/introduction/

http://www.cascadia-analytics.com/docs/user2018-docker-r-tutorial-slides.pdf

Versioned docker library for Rstudio: ROCKER project



rocker-org / rocker-versioned

Watch 18 | Star 132 | Fork 101

Code | Issues 10 | Pull requests 0 | Projects 0 | Wiki | Insights

Branch: master ▾ rocker-versioned / rstudio /

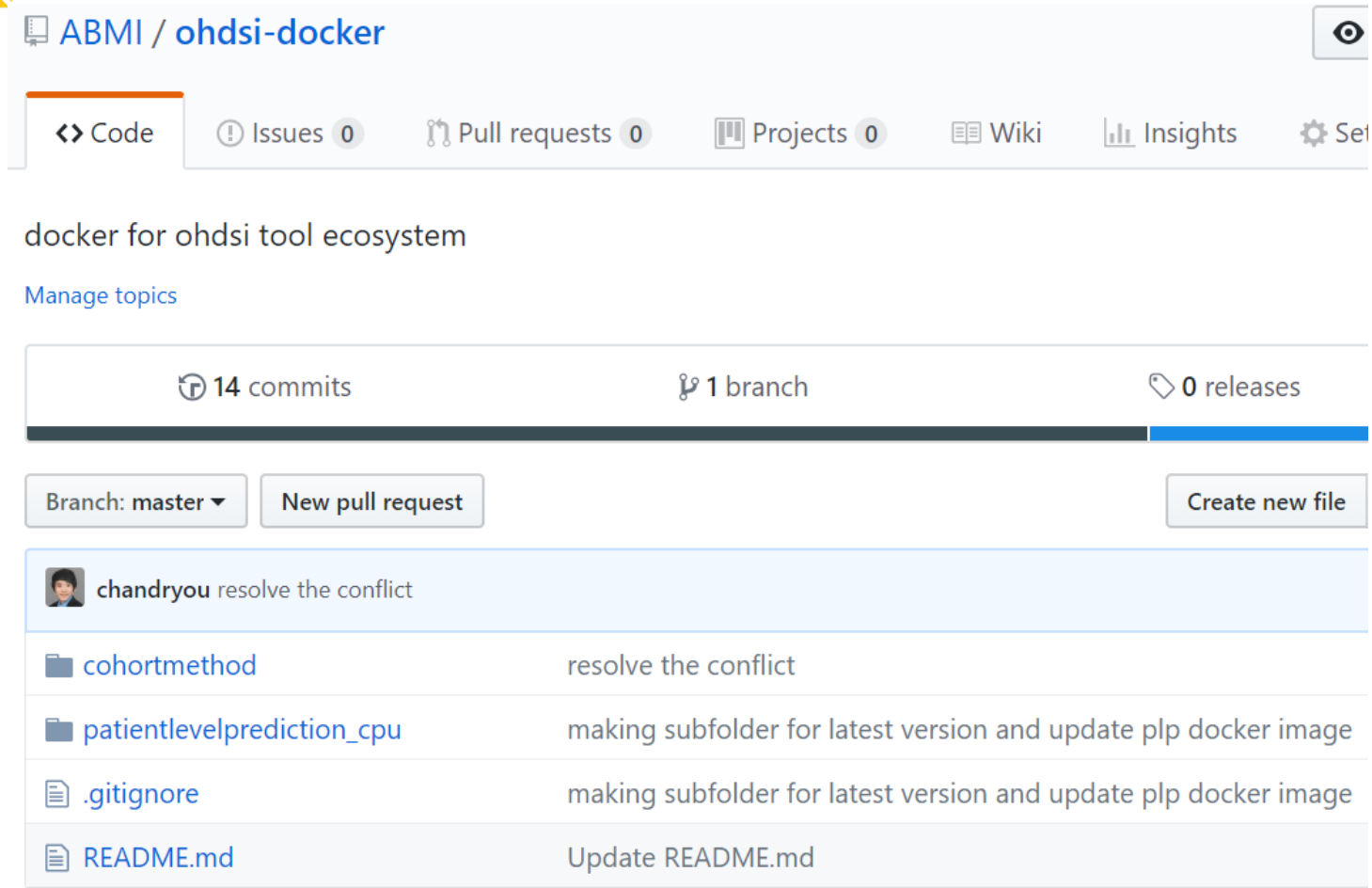
Create new file | Upload files | Find file | History

cboettig example addusers script — Latest commit 46b4bde 18 days ago

..		
3.1.0	rstudio on older versions of R	19 days ago
3.2.0	adding older versions to rstudio builds	19 days ago
3.2.5	rstudio on older versions of R	19 days ago
3.3.0	rstudio on older versions of R	19 days ago
3.3.1	patch bug in tex path, fixes #101	6 months ago
3.3.2	patch bug in tex path, fixes #101	6 months ago
3.3.3	patch bug in tex path, fixes #101	6 months ago
3.4.0	patch bug in tex path, fixes #101	6 months ago
3.4.1	patch bug in tex path, fixes #101	6 months ago
3.4.2	patch bug in tex path, fixes #101	6 months ago
3.4.3	patch bug in tex path, fixes #101	6 months ago
3.4.4	patch bug in tex path, fixes #101	6 months ago
3.5.0	patch bug in tex path, fixes #101	6 months ago
3.5.1	3.5.2 release (#114)	3 months ago
3.5.2	example addusers script	18 days ago

https://github.com/rocker-org/rocker-versioned/tree/master/rstudio

Versioned docker library for OHDSI Tools



https://github.com/ABMI/ohdsi-docker

How to build docker image for individual research

- Make a Dockerfile for the individual research based on versioned OHDSI docker image

Branch: master ▾ FebuxostatVsAllopurinolCVD / docker / Dockerfile

chandryou add dockerfile

1 contributor

4 lines (2 sloc) | 111 Bytes

```
1  FROM chandryou/cohortmethod:3.0.2
2
3  RUN R -e "devtools::install_github('chandryou/FebuxostatVsAllopurinolCVD')"
```

https://github.com/chandryou/FebuxostatVsAllopurinolCVD/blob/master/docker/

Run Docker

3. Alternatively, you can pull docker image for FebuxostatVsAllopurinolCVD. In the 'shell', use following code to pull docker image for FebuxostatVsAllopurinolCVD

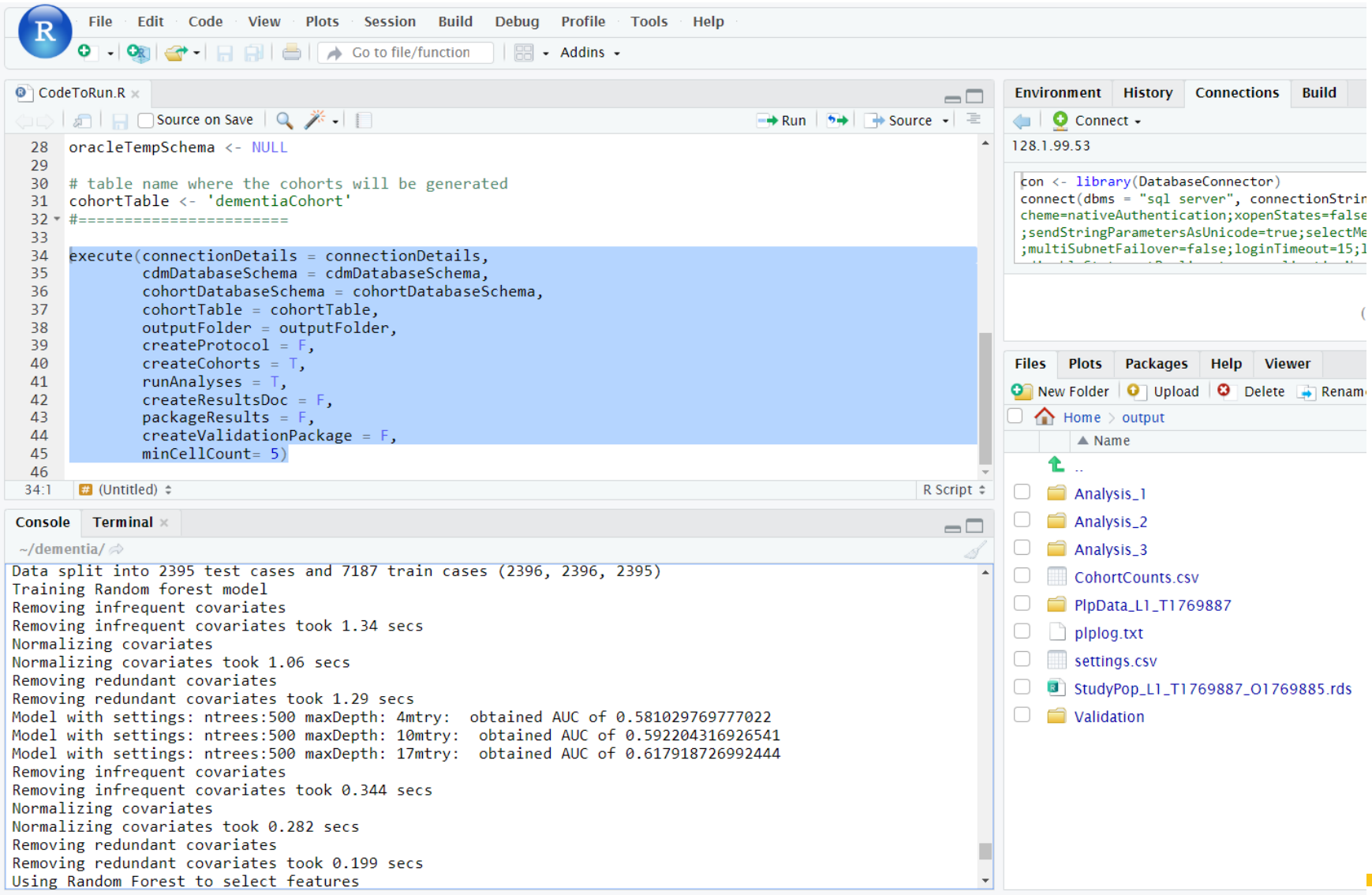
```
$docker run --name plp -e USER=user -e PASSWORD=password1 -d -p 8787:8787 chandryou/febuxostatvsallopurinolcvd
```

Then in the 'browser' activate Rstudio with following address

http://localhost:8787

The ID and PW are user and password1 as set above.

Run entangled system of PLP package (R + **Python** + **Deep learning**) in Docker



Suggestion for better reproducibility in OHDSI

- Build and maintain versioned Docker image library for OHDSI tools
  - I hope this versioned docker image library to be merged with BroadSea project
- Release docker image for each OHDSI research

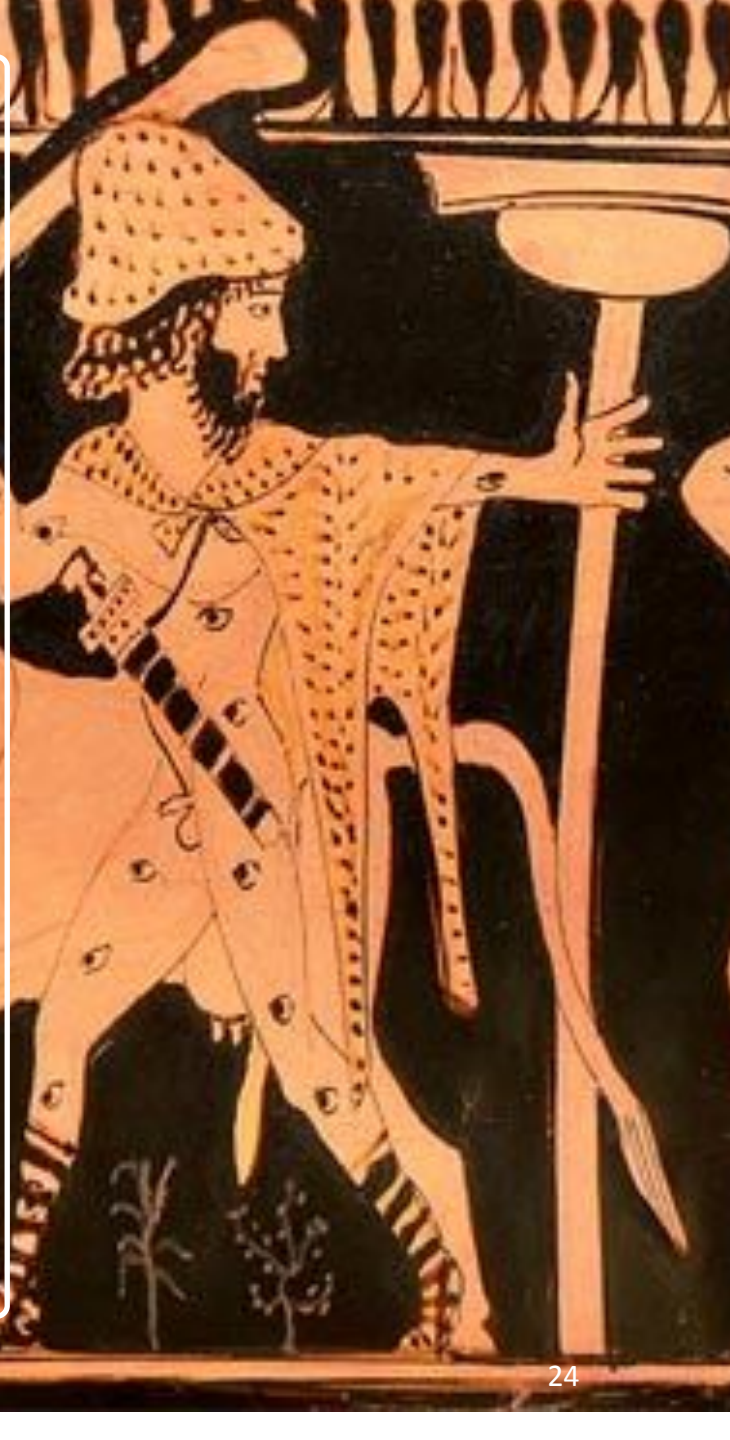
ARGOS

- **A R**igorous **G**lobal **O**bservation **S**ystem for burden of diseases
- Monitoring system of disease burden across OHDSI community

  Temporal trends in

  - Incidence
  - Mortality
  - Cost
  - DALY (Disability-Adjusted Life Year)

https://github.com/ABMI/Argos



Measuring the Global Burden of Disease

- To improve population health, it is crucial to estimate the burden of disease and understand how it changes over time.
- For obtaining comprehensive and consistent information for global burden of disease, the World Bank and the WHO launched the Global burden of Disease (GBD) study in 1991

World Bank. World Development Report 1993 — investing in health: world development indicators. Oxford, United Kingdom: Oxford University Press, 1993.

The result from GBD 2016

Figure 1. Age-Specific Global Contributions of Cancer Types to Total Cancer Incidence, Both Sexes, 2016

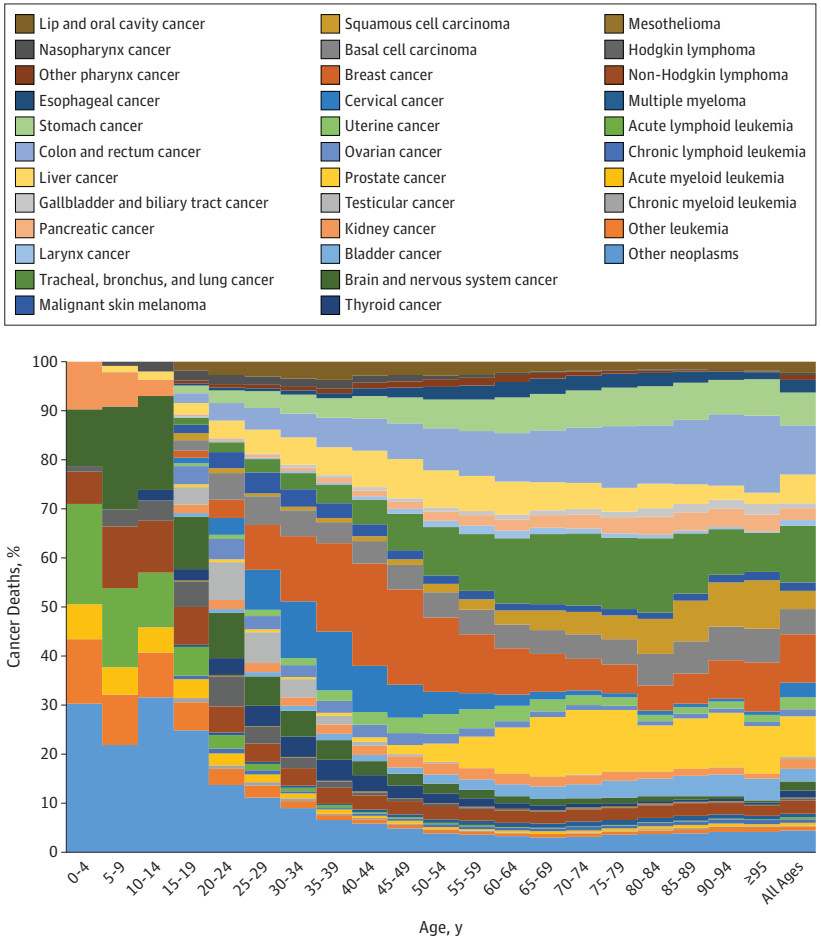
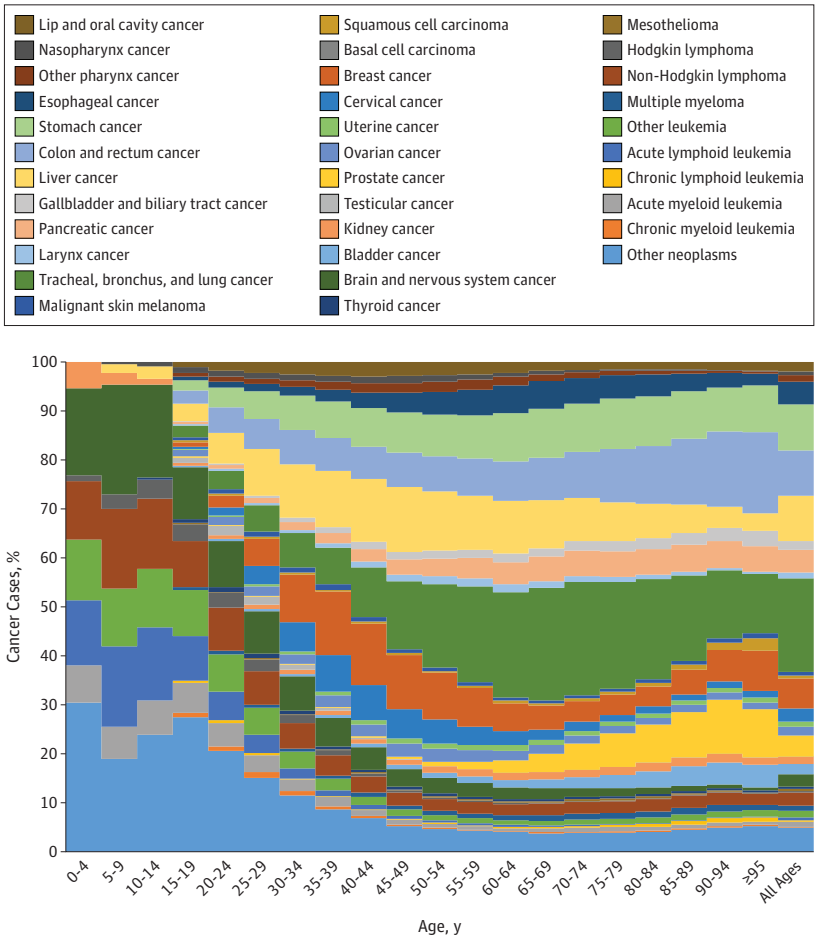


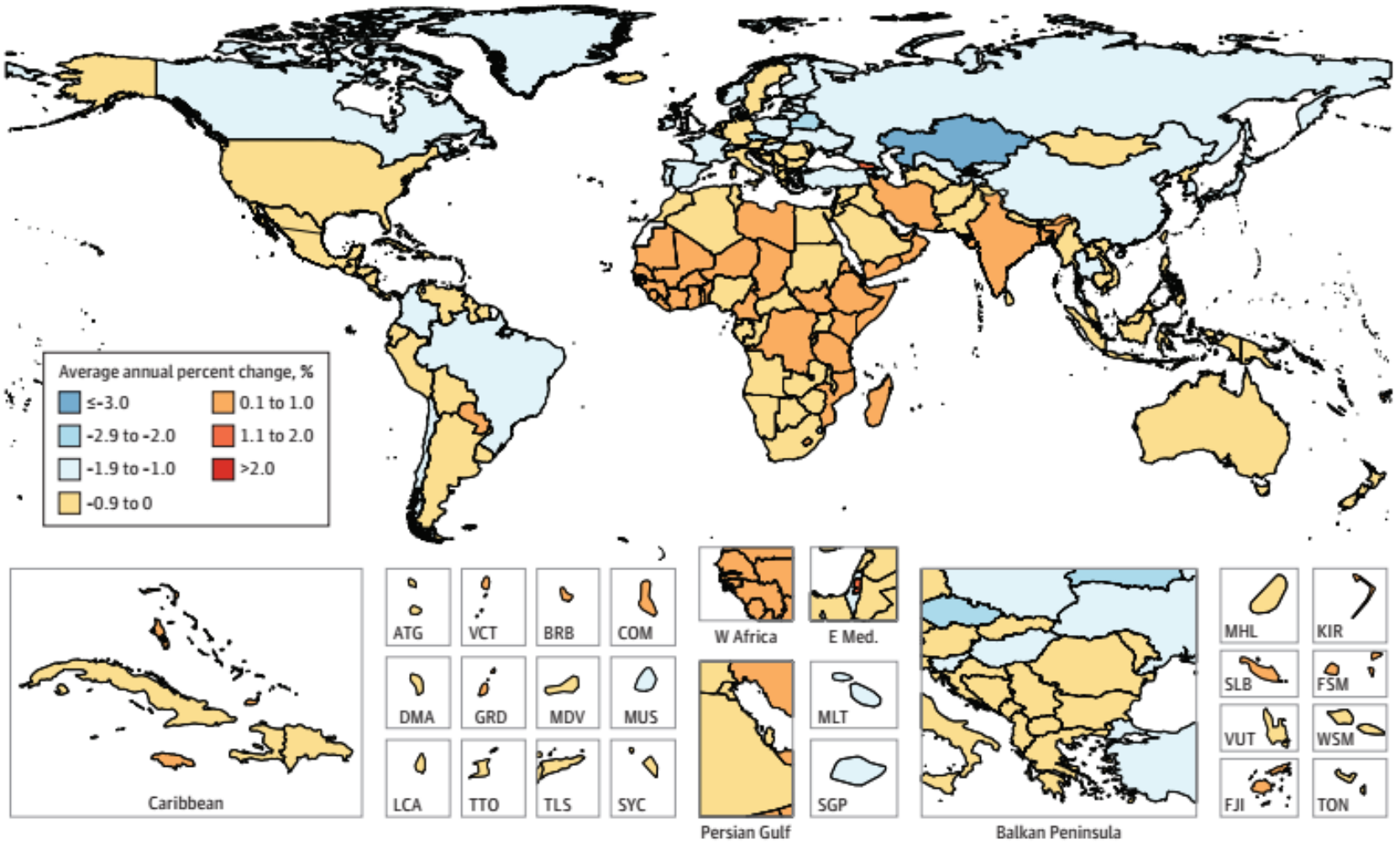
Figure 2. Age-Specific Global Contributions of Cancer Types to Total Cancer Mortality, Both Sexes, 2016



Fitzmaurice et al., *JAMA Oncology* 4, no. 11 (November 1, 2018)

The result from GBD 2016

Figure 4. Average Annual Percent Change in Age-Standarized Mortality Rate in Both Sexes for All Cancers From 2006 to 2016



Fitzmaurice et al., *JAMA Oncology* 4, no. 11 (November 1, 2018)

Comparison between GBD 2006 and GBD 2016

Figure 7. Cancers Ranked by Absolute Years of Life Lost (YLLs) Between 2006 and 2016[a]

Rank	Cancer 2006	Cancer 2016	Rank	Change in Absolute YLLs, % (UI)	Change in Age-Standardized YLL Rate, % (UI)
1	Tracheal, bronchus, and lung cancer	Tracheal, bronchus, and lung cancer	1	13.5 (9.9 to 16.8)	-11.9 (-14.6 to -9.3)
2	Stomach cancer	Liver cancer	2	15.1 (11.2 to 19.6)	-8.4 (-11.5 to -4.9)
3	Liver cancer	Stomach cancer	3	-4.0 (-6.5 to -1.5)	-24.7 (-26.7 to -22.8)
4	Colon and rectum cancer	Colon and rectum cancer	4	17.0 (11.1 to 21.7)	-8.9 (-13.4 to -5.3)
5	Breast cancer	Breast cancer	5	13.8 (5.6 to 21.9)	-9.5 (-15.9 to -3.5)
6	Leukemia	Leukemia	6	-2.4 (-6.6 to 1.9)	-15.2 (-18.7 to -11.6)
7	Esophageal cancer	Esophageal cancer	7	0.7 (-2.3 to 4.2)	-22.0 (-24.3 to -19.3)
8	Cervical cancer	Pancreatic cancer	8	26.7 (22.6 to 30.4)	-2.2 (-5.2 to 0.7)
9	Brain and nervous system cancer	Brain and nervous system cancer	9	13.5 (9.1 to 20.5)	-3.9 (-7.6 to 2.1)
10	Pancreatic cancer	Cervical cancer	10	4.9 (-1.4 to 13.1)	-15.8 (-20.9 to -9.3)
11	Non-Hodgkin lymphoma	Non-Hodgkin lymphoma	11	22.3 (15.5 to 26.8)	1.2 (-4.4 to 4.8)
12	Other leukemia	Prostate cancer	12	26.5 (19.3 to 32.2)	-4.1 (-9.4 to 0.4)
13	Prostate cancer	Lip and oral cavity cancer	13	26.2 (20.6 to 31.4)	-0.4 (-4.6 to 3.7)
14	Lip and oral cavity cancer	Ovarian cancer	14	20.8 (13.8 to 27.0)	-5.1 (-10.4 to -0.2)
15	Ovarian cancer	Other leukemia	15	-15.1 (-20.1 to -9.6)	-25.5 (-29.7 to -20.9)
16	Gallbladder and biliary tract cancer	Gallbladder and biliary tract cancer	16	14.7 (9.6 to 19.7)	-11.3 (-15.1 to -7.5)

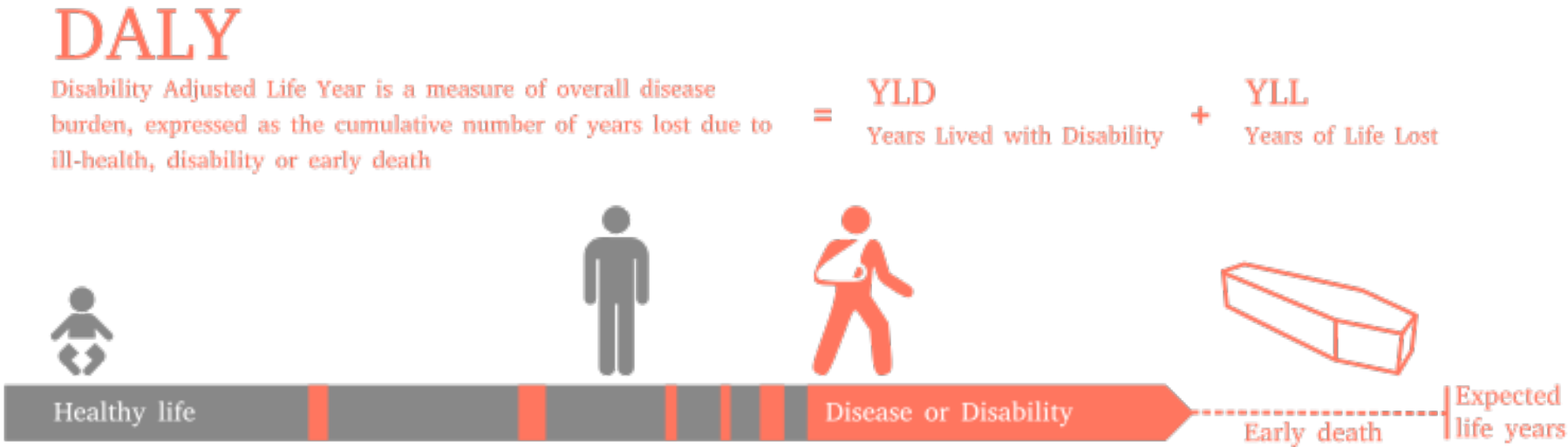
Fitzmaurice et al., *JAMA Oncology* 4, no. 11 (November 1, 2018)

Lesson from GBD study

- Large disparities exist between countries
  - In cancer incidence
  - In mortality of cancer patients
  - In cancer associated disability
- Large disparities exist across a decade
- **Differences** in **data collection practices** and **coding systems**, as well as quality of data sources, **remain major challenges**, as do underreporting of cancers requiring advanced diagnostics in low-resource settings

Disability-Adjusted Life Loss (DALY)

- A measure of overall disease burden, expressed as the number of years lost due to ill-health, disability or early death.
- One of fundamental estimates for cost-effectiveness research



https://en.wikipedia.org/wiki/Disability-adjusted_life_year

Objectives of ARGOS project

- Development of ARGOS package based on OHDSI tool ecosystem, which provides semi-automatic process to monitor burden of user-defined conditions in OMOP-CDM database by assessing
  - Temporal trend in incidence
  - Incidence according to age, gender, and birth year
  - Temporal trend in outcome of care
  - Disability-Adjusted Life Loss (DALY)
- In Korea
  - Conversion of HIRA DB for whole cancer patients from 2007~2017
- Under development
  - Today's result is based on NHIS-NSC (2003-2013, 1M sample)
  - Today's result is not fully evaluated.

https://github.com/ABMI/Argos

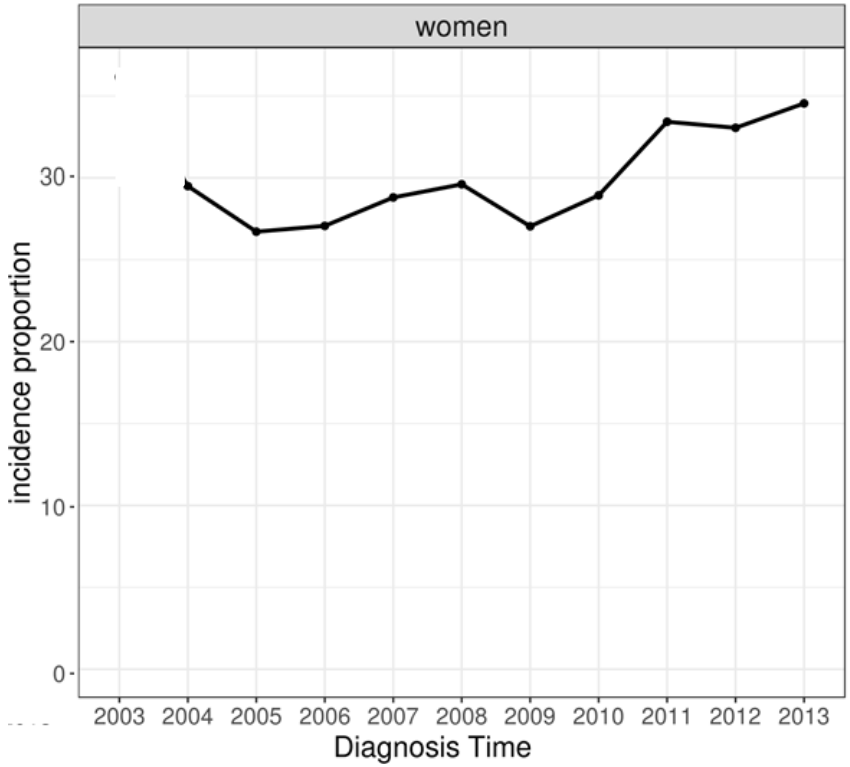
Validation of temporal trend in Incidence of cancers

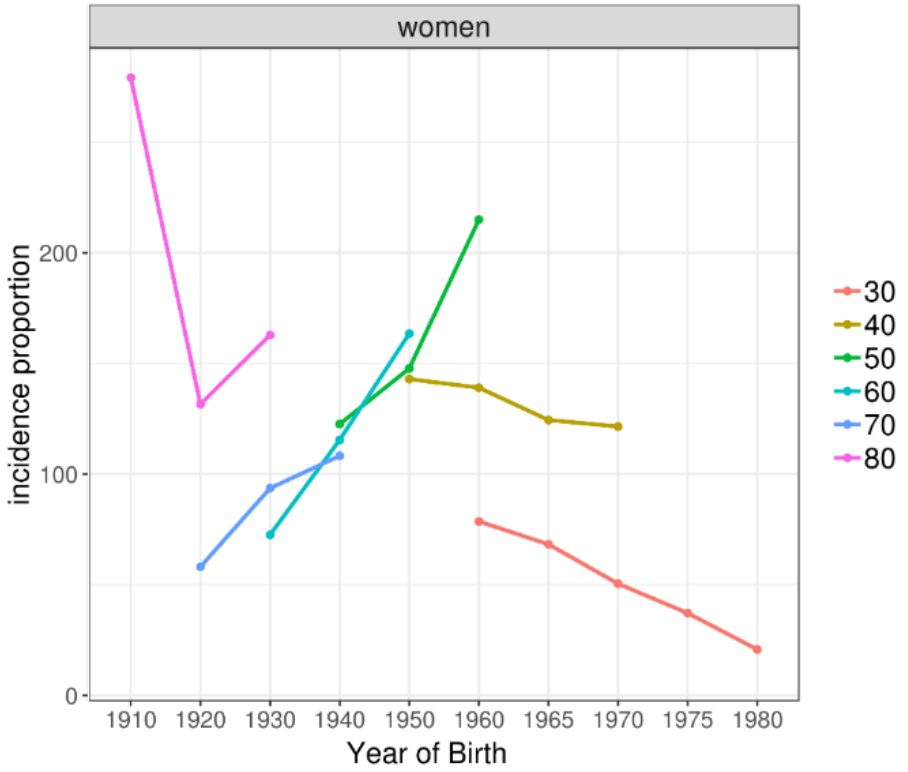
		Incidences (cases per 100,000 persons)									
Cancer site		2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Colorectal	Argos	32.6	32.0	36.4	37.2	39.3	39.0	41.7	47.4	51.4	53.0
	Statistics Korea	30.5	33.6	37.6	40.7	43.8	47.0	51.4	53.4	57.1	58.5
Lung	Argos	32.2	31.8	30.8	33.3	32.7	33.4	30.9	33.1	42.8	40.5
	Statistics Korea	31.5	33.8	35.4	36.3	37.6	40.4	42.7	44.3	44.6	46.4
Stomach	Argos	44.2	42.9	44.3	43.3	42.3	48.1	46.5	43.7	54.1	49.9
	Statistics Korea	49.5	48.8	54.2	54.1	54.6	57.5	60.5	61.6	63.8	61.8
Liver	Argos	34.9	29.9	32.0	28.4	26.1	30.1	28.0	26.3	29.3	32.4
	Statistics Korea	29.1	29.9	31.1	30.6	31.3	32	32.2	32.4	32.7	32.0
Breast	Argos	16.1	16.4	15.8	18.1	20.0	21.7	20.2	25.1	26.6	27.8
	Statistics Korea	17.7	19.0	21.1	22.4	24.5	26.0	27.5	29.4	32.3	33.2
Thyroid	Argos	16.9	18.7	23.9	31.9	40.0	51.6	66.8	62.6	81.0	87.6
	Statistics Korea	15.6	21.5	26.3	33.0	43.3	55.3	65.5	73.7	82.4	88.6

Table. Comparison of estimated cancer incidence from Argos with the findings of relevant published report

Temporal trends of incidence, incidence according to birth year:

Is breast cancer really increasing in Korea?

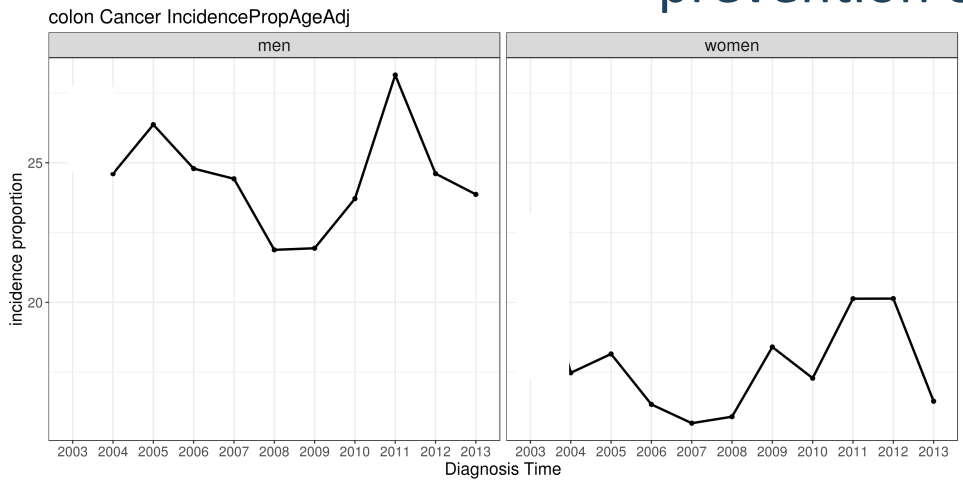


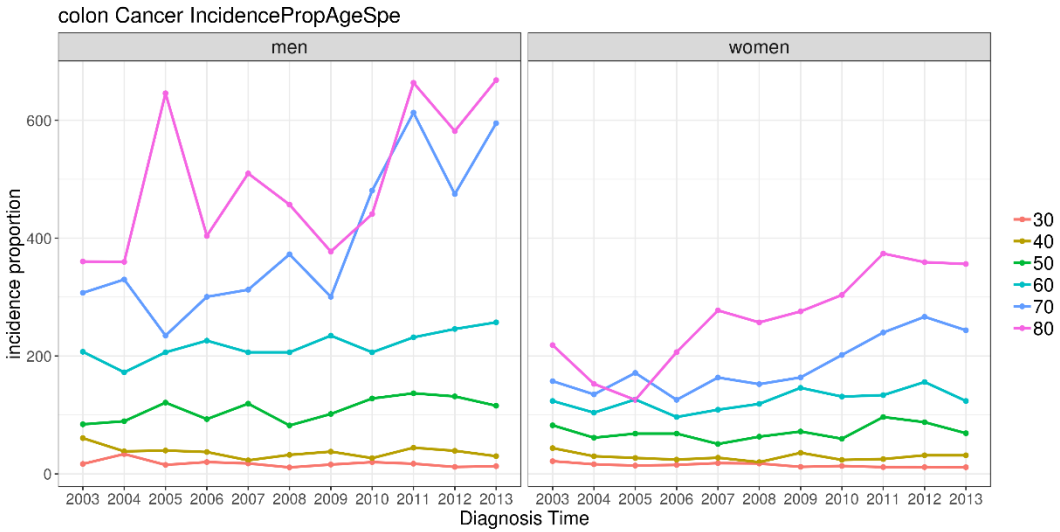


Though overall age standardized incidence of breast cancer increases, the incidence of breast cancer decreases in women born after 1960

Temporal trends of incidence according to year:

For which age group should we promote screening for prevention of colon cancer?





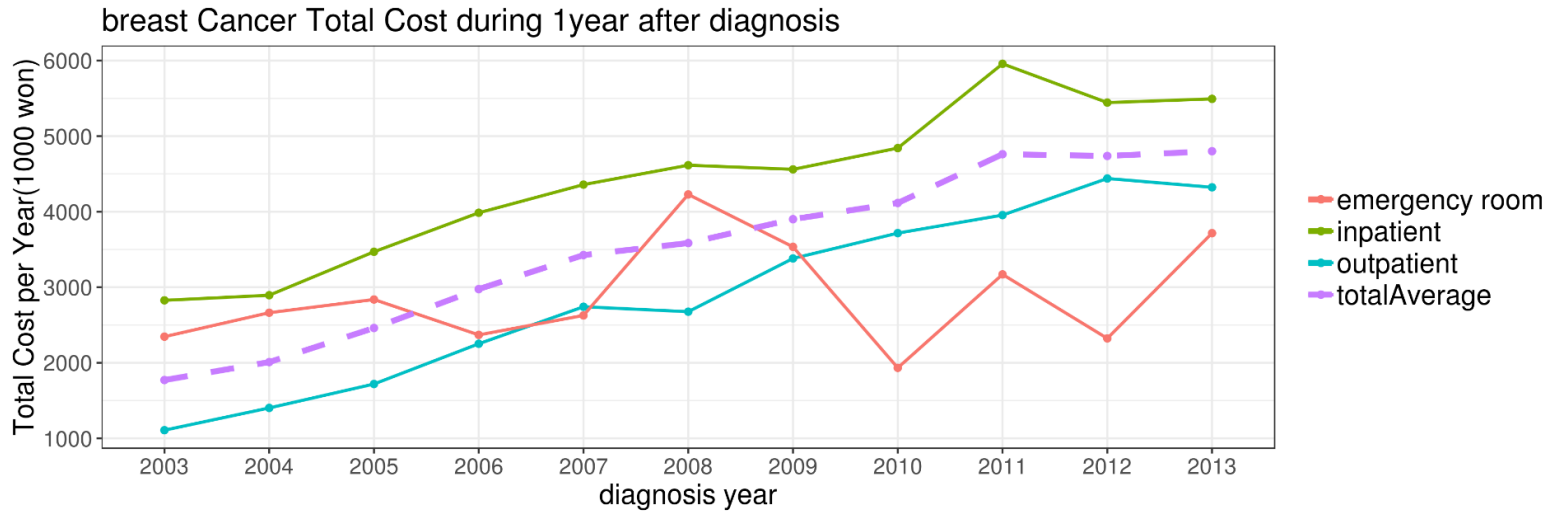
After rapid increase of incidence in colon cancer, overall incidence of colon cancer decreased in Korea.

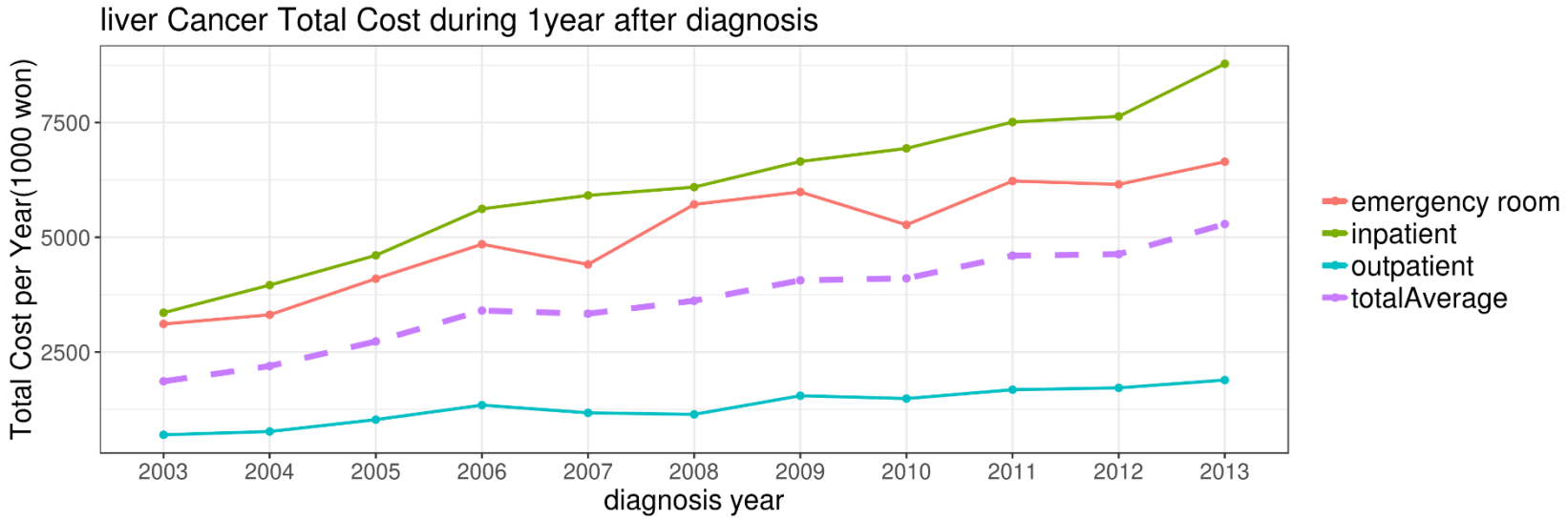
Still, the incidence of colon cancer increases in old ages.

We are neglecting these population in national screening system

https://github.com/ABMI/Argos

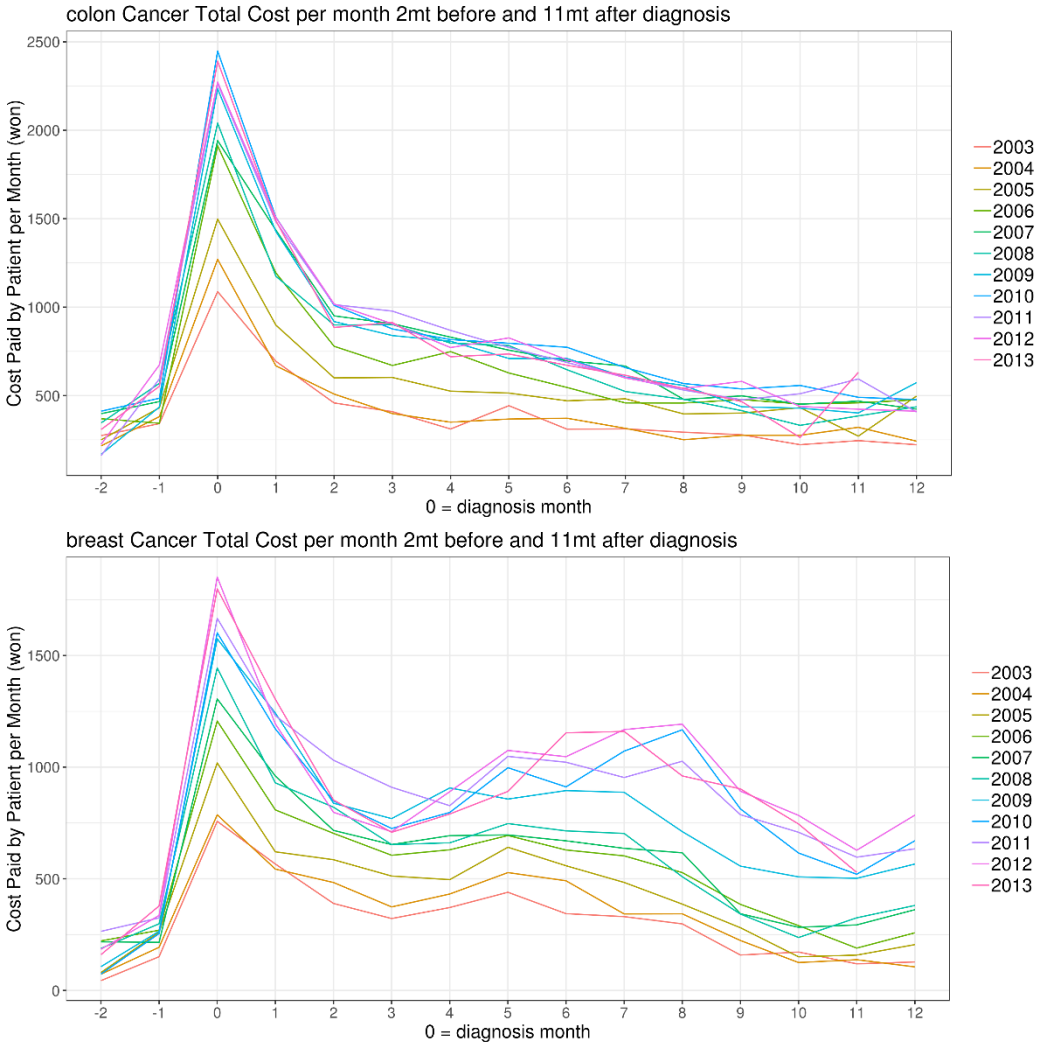
Temporal trend of overall cost in cancer patients



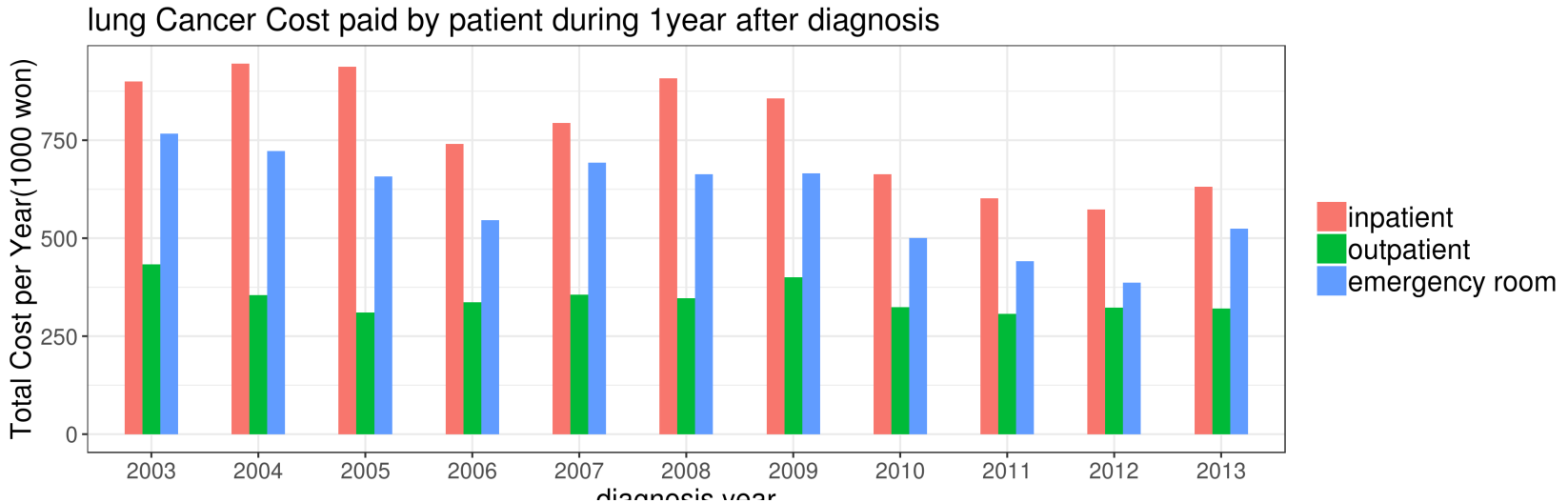


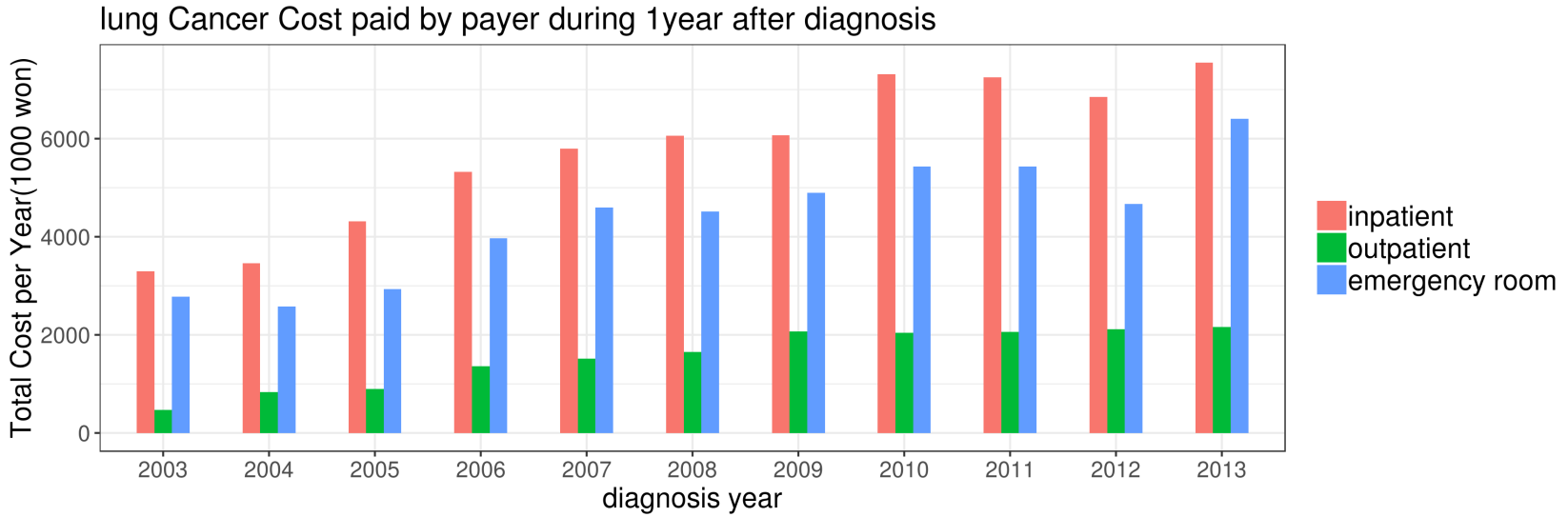
https://github.com/ABMI/Argos

Cost plot according to the time before and after diagnosis

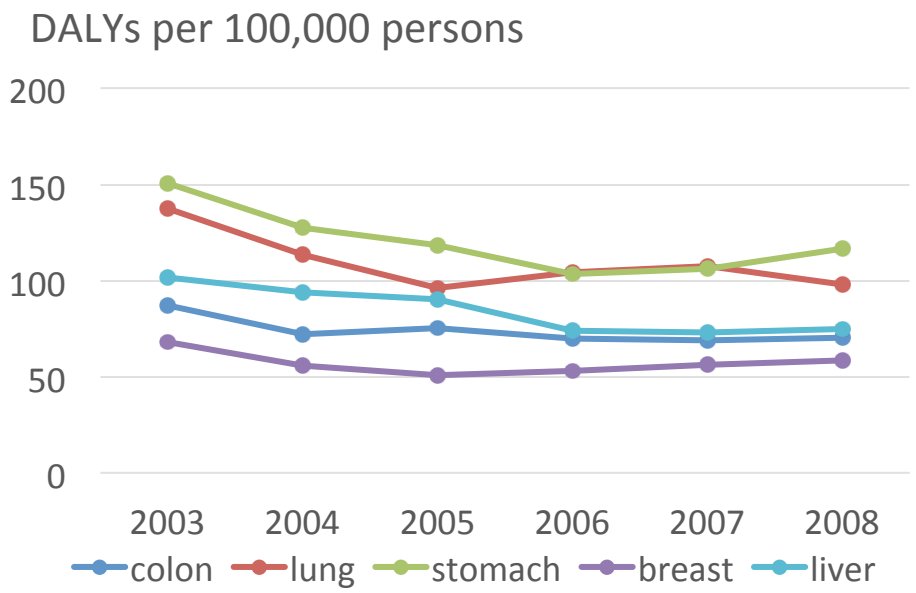


Temporal trend of overall cost in lung cancer paid by patient and national insurance

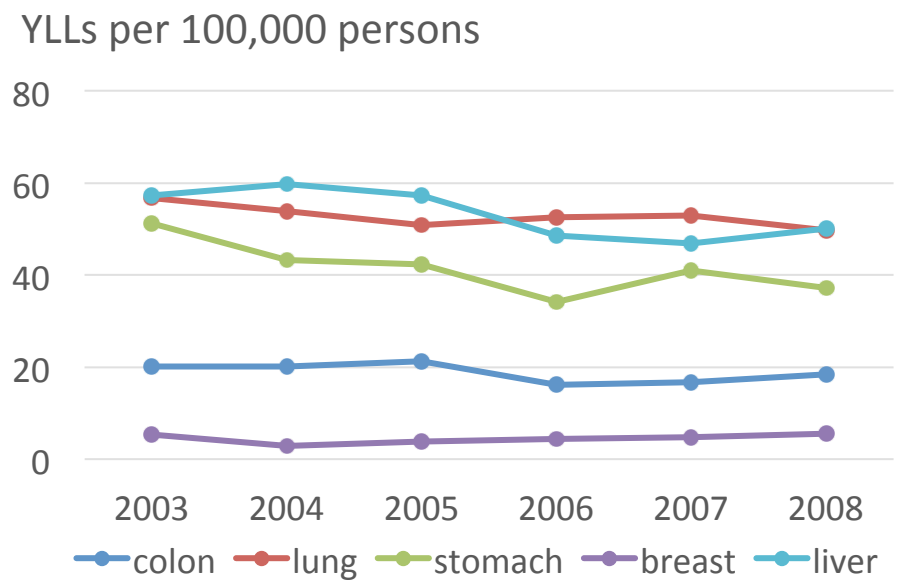




Assessing temporal trends in DALY based on incidence data

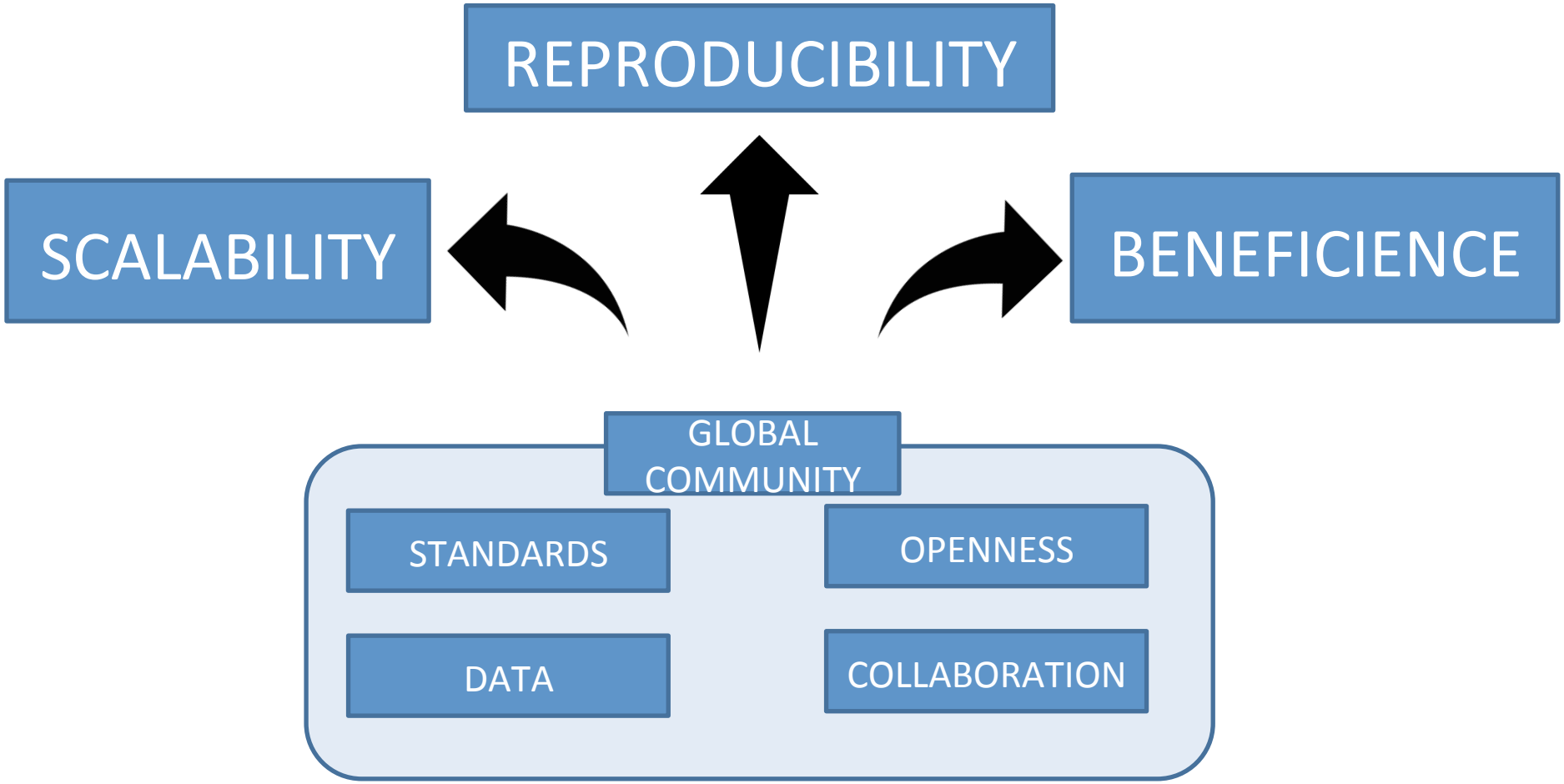


DALYs per 100,000 persons during 2003-2008



YLLs per 100,000 persons during 2003-2008

OHDSI: Open Innovation based on the open community



Mission, Vision, and Values of OHDSI

- Our Mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

- Our Vision

A world in which observational research produces a comprehensive understanding of health and disease.

Mission, Vision, and Values of OHDSI

- Innovation: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.
- Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.
- Community: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.
- Collaboration: We work collectively to prioritize and address the real world needs of our community's participants.
- Openness: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.
- Beneficence: We seek to protect the rights of individuals and organizations within our community at all times.


Thank You
for your time