1st of July, Ljubjana

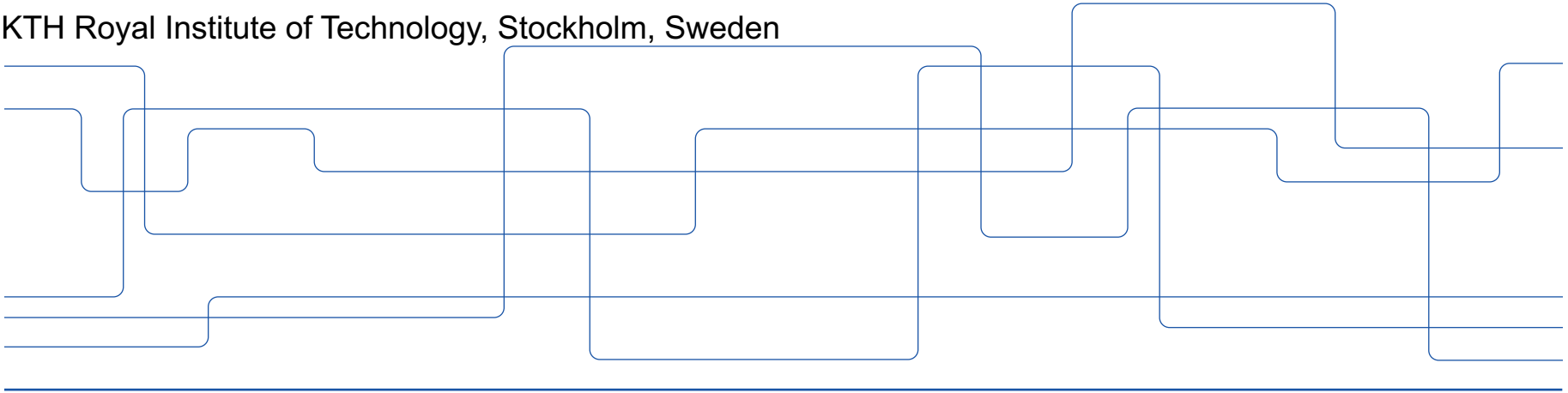
30th PAGE meeting

Session: Real-World Data in Oncology

Introducing a novel analytical framework for risk stratification of real-world data with survival and unsupervised machine learning.
A small cell lung cancer SCLC study

Luca Marzano, PhD candidate,

KTH Royal Institute of Technology, Stockholm, Sweden



Real-World Data

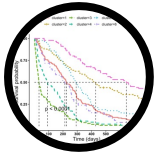
RCT

- Safety
- Pivotal studies
- Pharmacokinetics and dose finding
- Efficacy in controlled patient samples

RWD

- EHR
- Registries
- Patient-generated data
- Healthcare production

RWE



Risk stratification

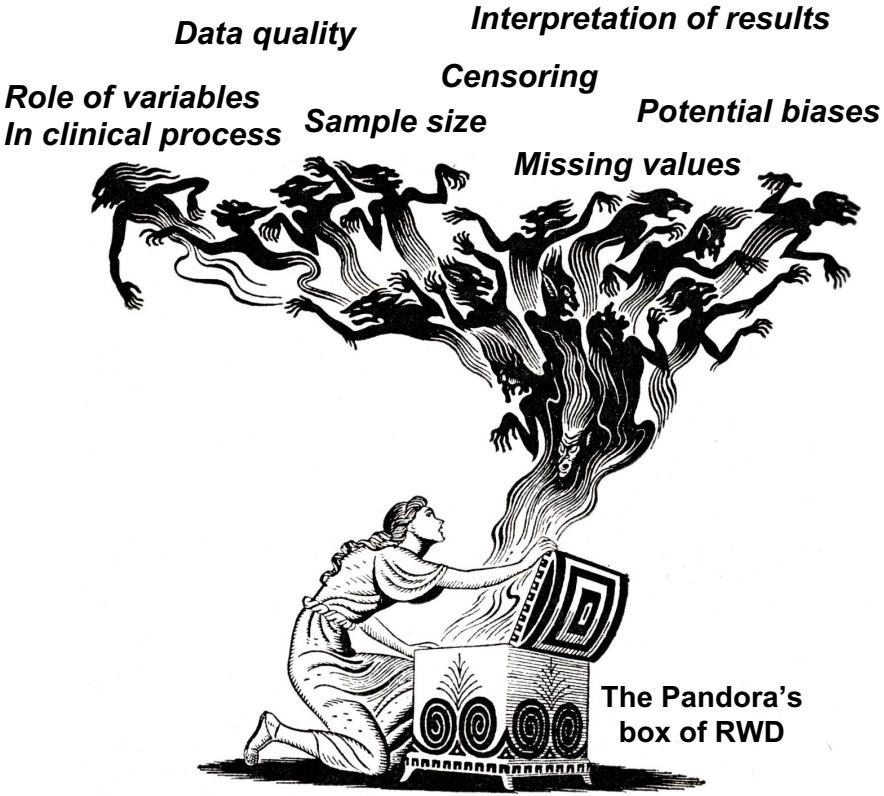
https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

Challenges and Aim of the work

However, RWD poses a series of practical challenges

Aim of the work: propose an approach to stratify RWD cohorts in clinical meaningful subgroups

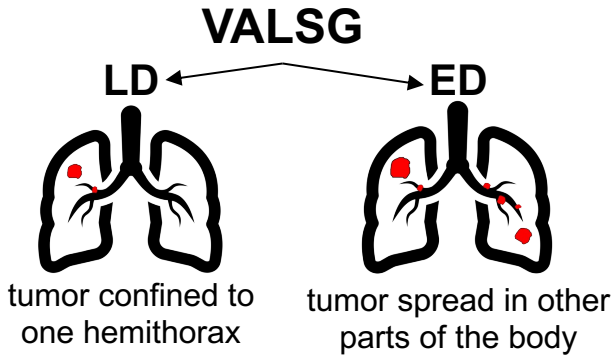
Focus: Small Cell Lung Cancer RWD



Credits: https://eruanna317.medium.com/pandoras-box-2017-278cb0373cb8

Case study: Small Cell Lung Cancer

Veterans' Administration Lung Study Group (VALSG) staging



the 8th version of the International Association for the Study of Lung Cancer (IALSC) TNM staging

8th TNM (T8, N8, M8 ->ST8)

T/M	Subcategory	N0	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	IIIB
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	M1a	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

ED

Detterbeck, F. C. (2018). The eighth edition TNM stage classification for lung cancer: What does it mean on main street?. *The Journal of thoracic and cardiovascular surgery*, *155*(1), 356-359.

Sample size: majority of patients ED (TNM: IVB T4N3M1C)

Validation of 8th TNM with the present cohort: Tendler, S., Grozman, V., Lewensohn, R., Tsakonas, G., Viktorsson, K., & De Petris, L. (2018). Validation of the 8th TNM classification for small-cell lung cancer in a retrospective material from Sweden. *Lung Cancer*, *120*, 75-81.

SCLC Real-World Data

Cases diagnosed and treated at Karolinska University Hospital between 2008 and 2016.

SCLC cohort
IIIA-IVB stage
n=636

VALSG LD = I-III TNM
VALSG ED = IV TNM

LD
ED
LD with large tumor size

CT+RT n=84

CT n=463

advanced age
comorbidities
ECOG PS 3(or higher)
refuse of treatment

special selected cases with ECOG PS 3 could receive CT

No Treatment n=89

35% **missing values** Were imputed with missing forest algorithm

PAGE 29 (2021) Abstr 9742 [www.page-meeting.org/?abstract=9742]

BASELINES
- age
- sex (M/F)
- TNM staging
  - T8
  - N8
  - M8
  - ST8
- ECOG PS (0-3)
- Lab measures
  - log LDH
  - log CRP
  - HB
  - Na
  - Albumin
- PET CT (y/n)
- Brain CT (y/n)

Role of variables in the clinical process

Chemotherapy cycles

1L | **2L** | **3L** | **4L**

PE (n=521)

PE (n=64)
IP(n=66)
Mono (n=20)

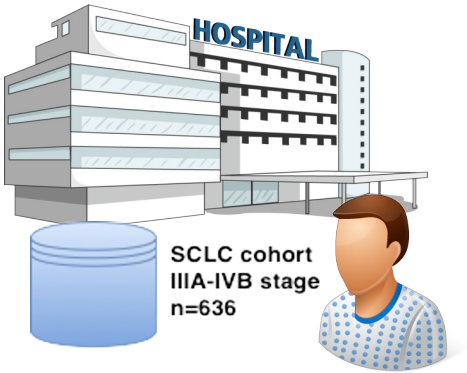
PE (n=4)
IP(n=16)
Mono (n=28)

PE (n=1)
IP (n=1)
Mono(n=3)

Other strategies (n=26)
received 3L (n=5) from which n=1 received 4L

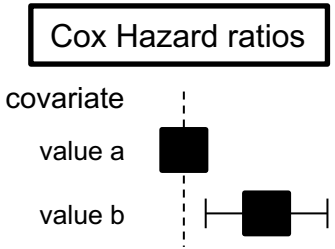
PCI (n=146)
starting within 4-8 weeks after completion of first line therapy
good response
sign of regression
ECOG PS 0-1

Traditional Survival and Unsupervised analysis



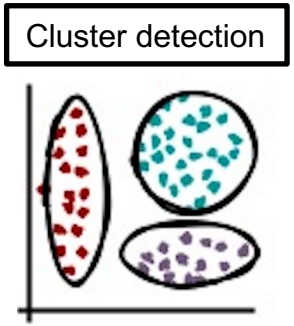
Censoring

Survival analysis
Covariate prognostic impact



Emerging of Machine Learning Techniques to handle censored data.
e.g., Random Survival Forest

Unsupervised learning
Grouping of observations/patients

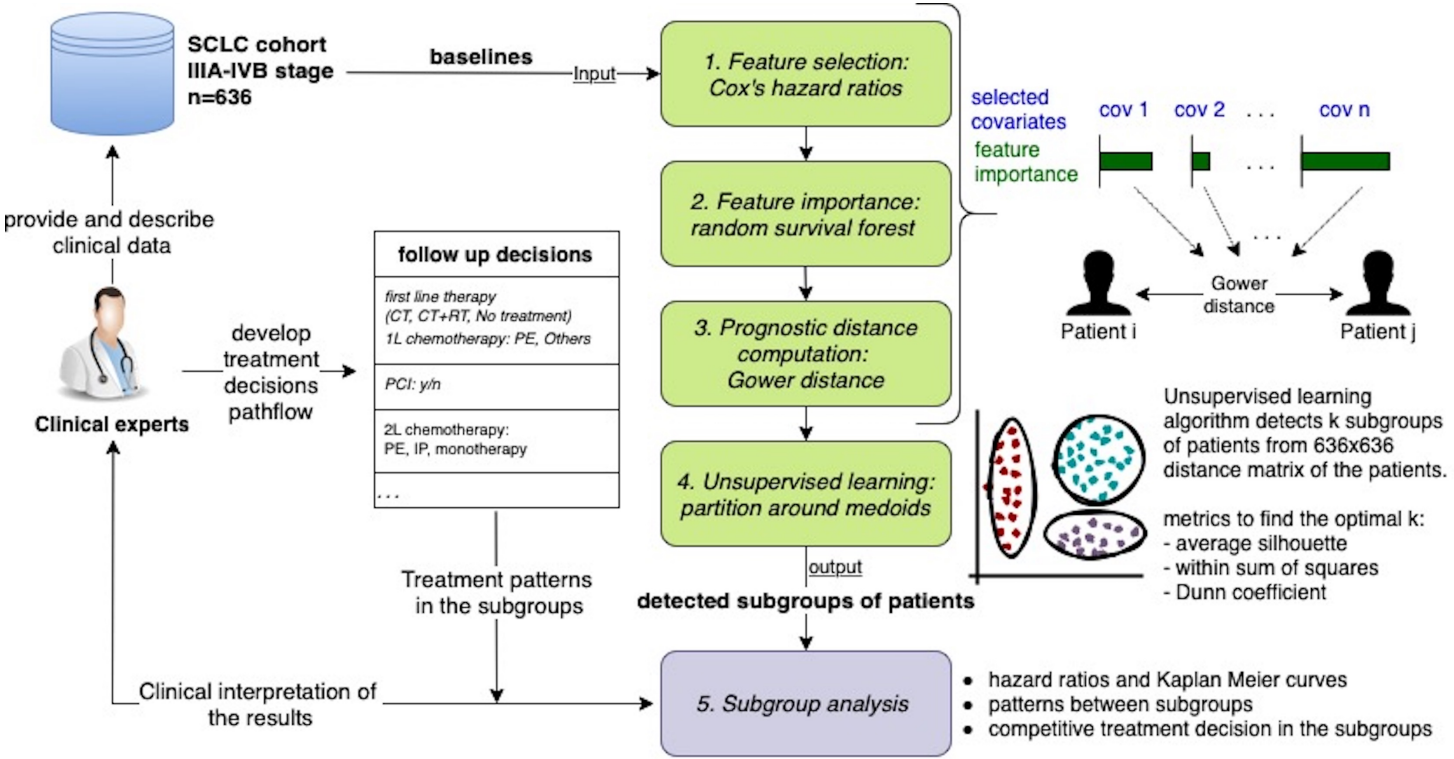


Definition of a pairwise similarity score between patient covariates

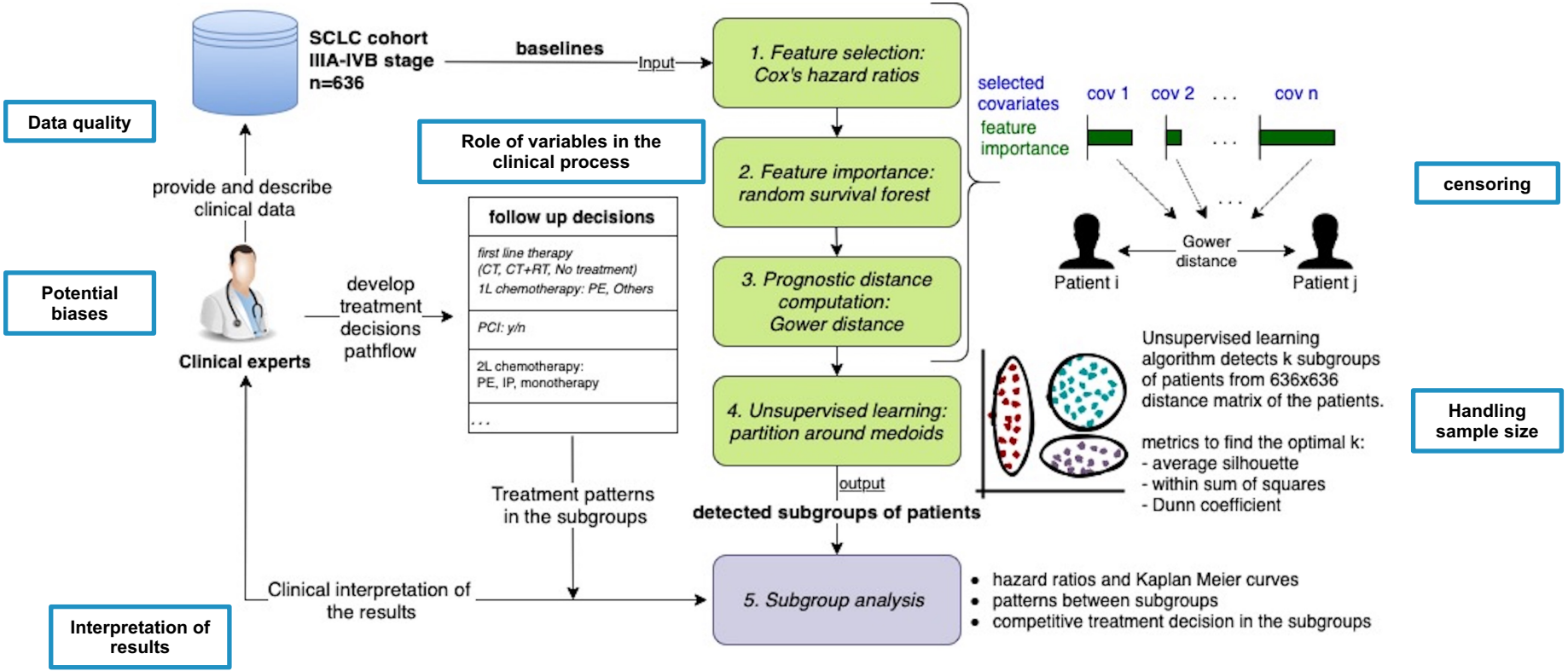
Detection of subgroups of patients according to their similarity

Previous study based on Cox model with the present cohort:
Tendler, S., Zhan, Y., Pettersson, A., Lewensohn, R., Viktorsson, K., Fang, F., & De Petris, L. (2020). Treatment patterns and survival outcomes for small-cell lung cancer patients–a Swedish single center cohort study. *Acta Oncologica*, *59*(4), 388-394.

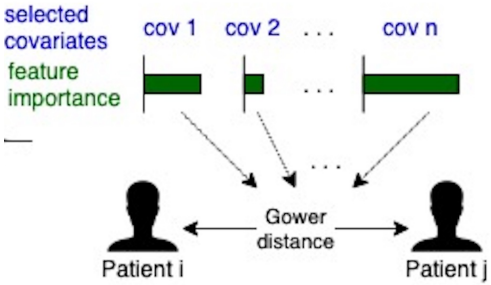
The framework: combining survival and unsupervised learning



The framework: combining survival and unsupervised learning



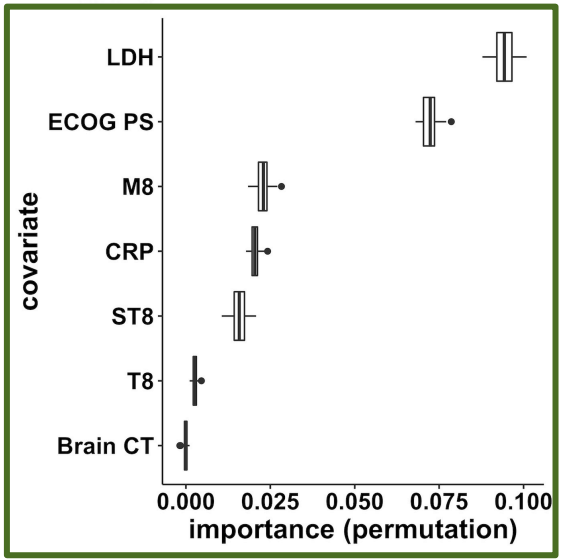
Results: covariate impact and detected clusters

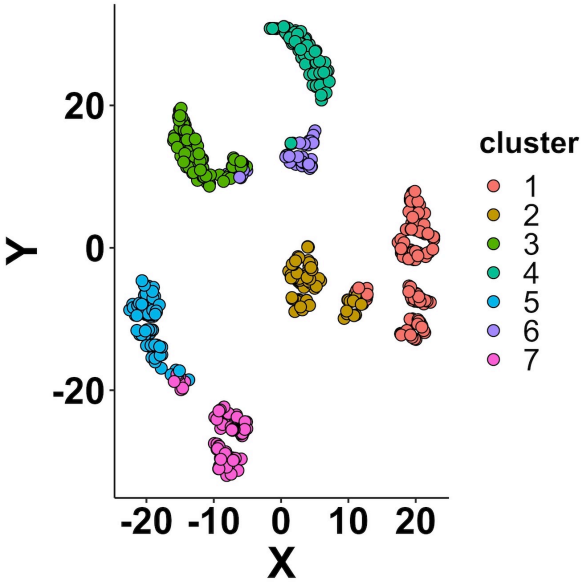


Unsupervised learning algorithm detects k subgroups of patients from 636x636 distance matrix of the patients.

metrics to find the optimal k:
- average silhouette
- within sum of squares
- Dunn coefficient

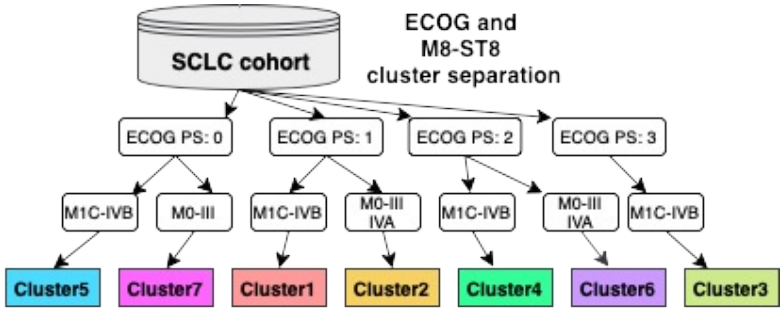
Variable		N	Hazard ratio	p
CRP		636	1.32 (1.11, 1.58)	0.002
LDH		636	3.81 (2.53, 5.74)	<0.001
ECOG PS	0	175	Reference	
	1	231	1.19 (0.95, 1.49)	0.130
	2	126	2.25 (1.71, 2.96)	<0.001
	3	104	3.52 (2.58, 4.79)	<0.001
Brain CT	No	416	Reference	
	Yes	220	1.22 (1.02, 1.47)	0.032
ST8	IVB	373	Reference	
	IIIA	54	0.82 (0.28, 2.45)	0.725
	IIIB	63	0.62 (0.21, 1.81)	0.383
	IIIC	68	0.72 (0.26, 2.05)	0.543
	IVA	78	3.04 (1.05, 8.84)	0.041
T8	T4	397	Reference	
	T0	7	1.22 (0.52, 2.85)	0.647
	T1a	5	0.68 (0.25, 1.89)	0.463
	T1b	26	0.74 (0.47, 1.16)	0.183
	T1c	34	1.23 (0.83, 1.81)	0.300
	T2a	45	0.83 (0.59, 1.18)	0.309
	T2b	38	0.65 (0.45, 0.93)	0.018
	T3	82	1.16 (0.89, 1.50)	0.272
	TX	2	0.82 (0.19, 3.50)	0.788
M8	M1C	373	Reference	
	M0	184	0.75 (0.27, 2.10)	0.585
	M1A	51	0.27 (0.09, 0.81)	0.019
	M1B	28	0.20 (0.07, 0.58)	0.003



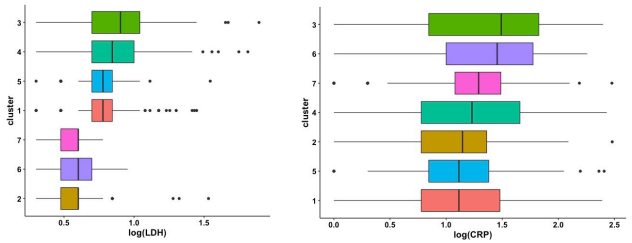


Results: subgroups analysis

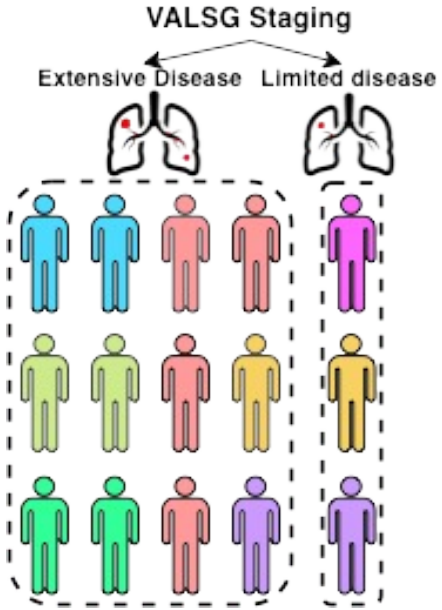
Covariate patterns



LDH and CRP distributions



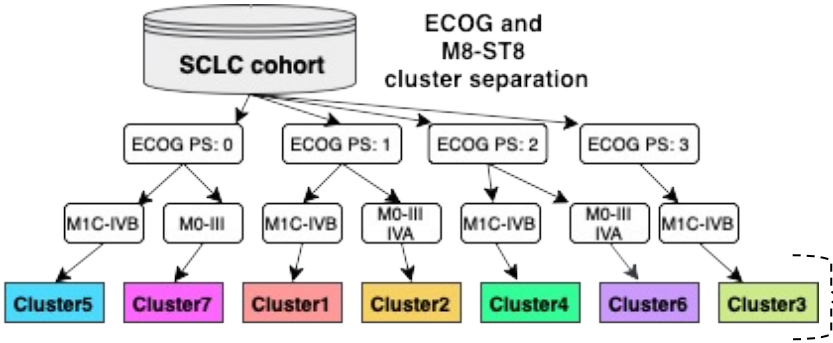
VALSG comparison



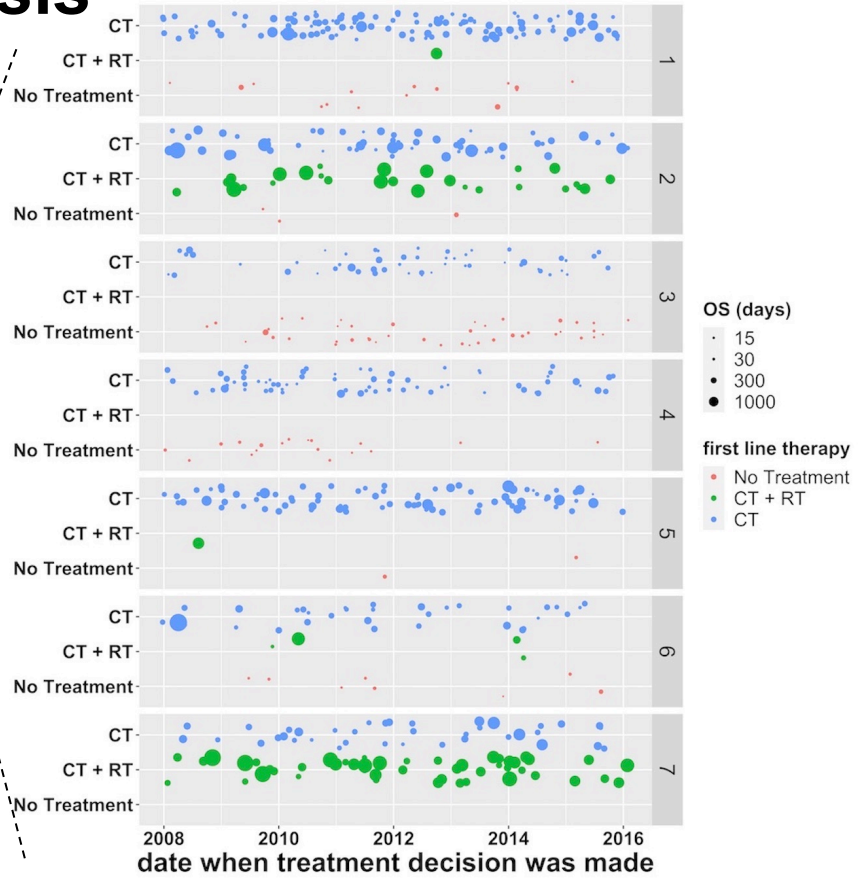
Cluster Hazard ratios

Variable	N	Hazard ratio	p
cluster	1 140	Reference	
2	91	0.44 (0.33, 0.58)	<0.001
3	97	3.47 (2.66, 4.53)	<0.001
4	87	2.18 (1.66, 2.87)	<0.001
5	85	0.75 (0.57, 0.98)	0.04
6	46	1.04 (0.74, 1.45)	0.83
7	90	0.30 (0.22, 0.40)	<0.001

Results: subgroup analysis



Treatment patterns

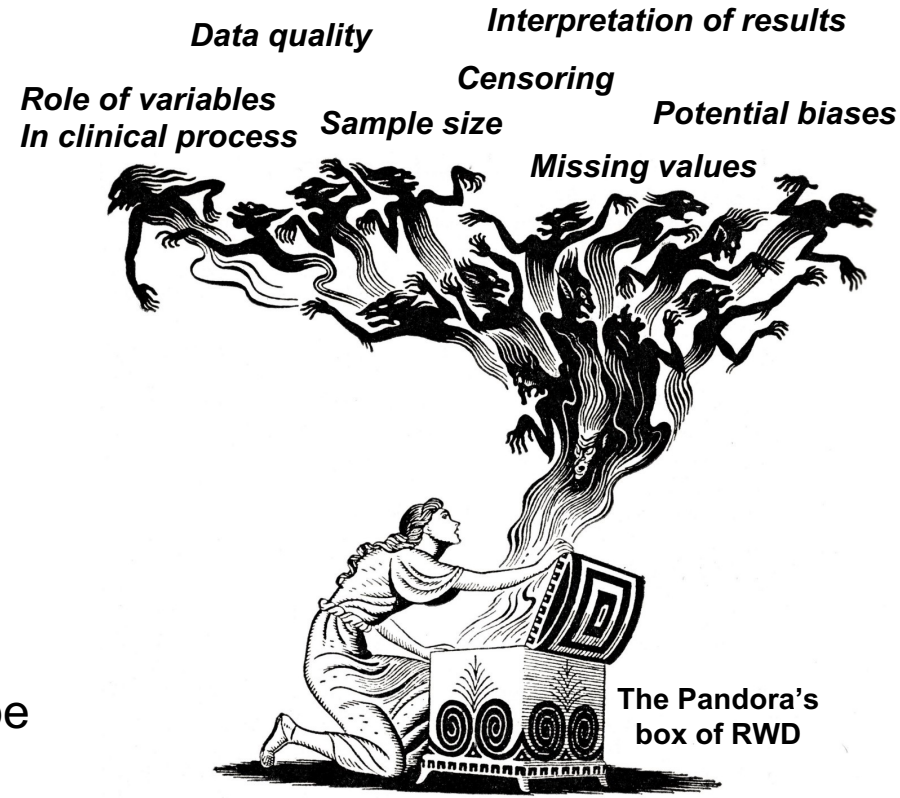


Variable	N	Hazard ratio	p
cluster	1 140	Reference	
	2 91	0.44 (0.33, 0.58)	<0.001
	3 97	3.47 (2.66, 4.53)	<0.001
	4 87	2.18 (1.66, 2.87)	<0.001
	5 85	0.75 (0.57, 0.98)	0.04
	6 46	1.04 (0.74, 1.45)	0.83
	7 90	0.30 (0.22, 0.40)	<0.001

- Competitive treatment CT vs CT+RT (clust 2 and 7)
- Competive decision CT vs No Treat. (clust 3 and 4)
- Similar survival CT (clust 1, 5 and 6)

Summary

- Comprehensive separation of SCLC prognostic groups
- Handling RWD challenges
- Considerations regarding the new stage categories IIIC, IVA and IVB
- Competitive processes and treatment patterns
- Potential role of RWD to inform and shape future clinical trials



Credits: https://eruanna317.medium.com/pandoras-box-2017-278cb0373cb8

Acknowledgments

Division of Health Informatics and Logistics,
School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH),
KTH Royal Institute of Technology, Stockholm, Sweden

- ***Luca Marzano****
- Adam S. Darwich
- Jayanth Raghotama
- Sebastiaan Meijer

Dept. of Oncology-Pathology,
Karolinska Institutet and the Thoracic Oncology Center,
Karolinska University hospital, Stockholm, Sweden.

- Luigi De Petris
- Rolf Lewensohn
- Salomon Tendler
- Asaf Dan

This project is a contribution to the Centre for Data-Driven Health (CDDH), KTH Royal Institute of Technology (https://www.kth.se/cddh)

***Contact information**

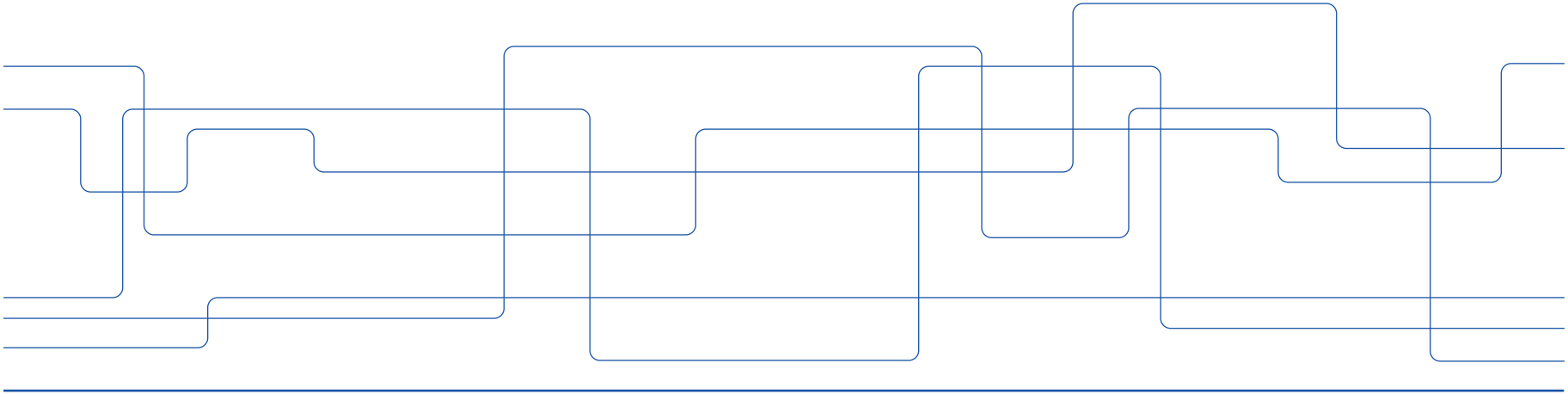
Luca Marzano

lmarzano@kth.se

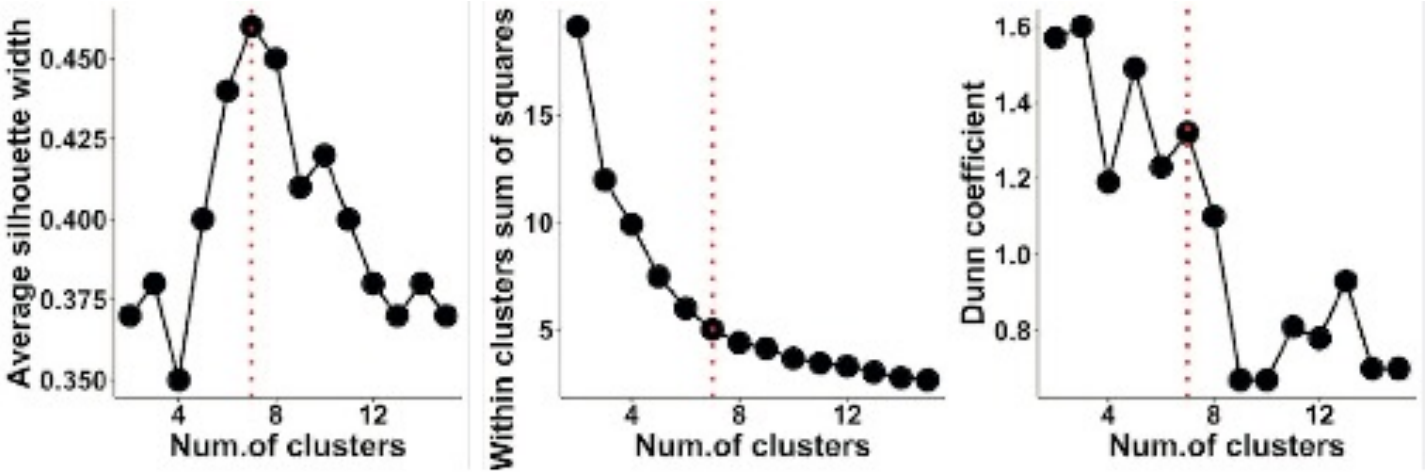
lmarzano@ug.kth.se

Luca Marzano

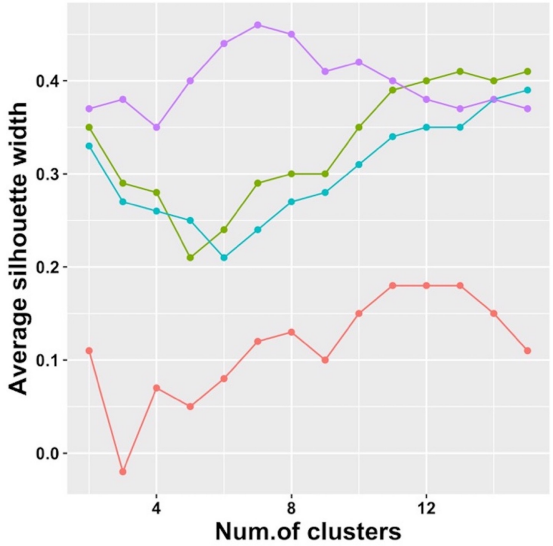
Thank you for your attention!

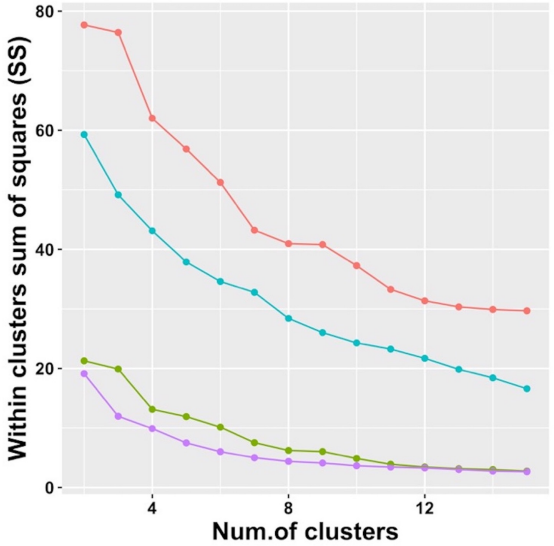


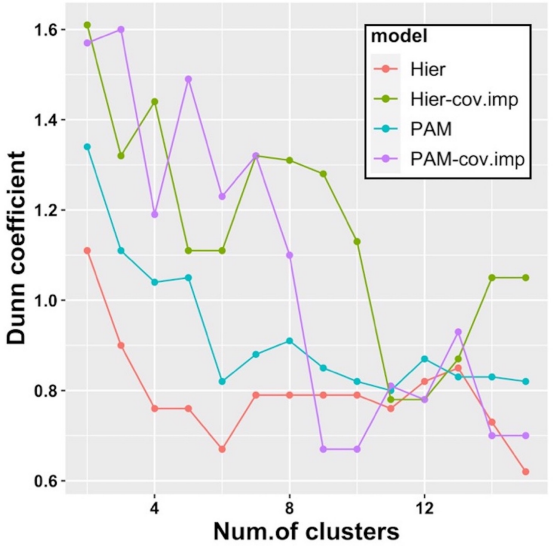
Back-up slide 1

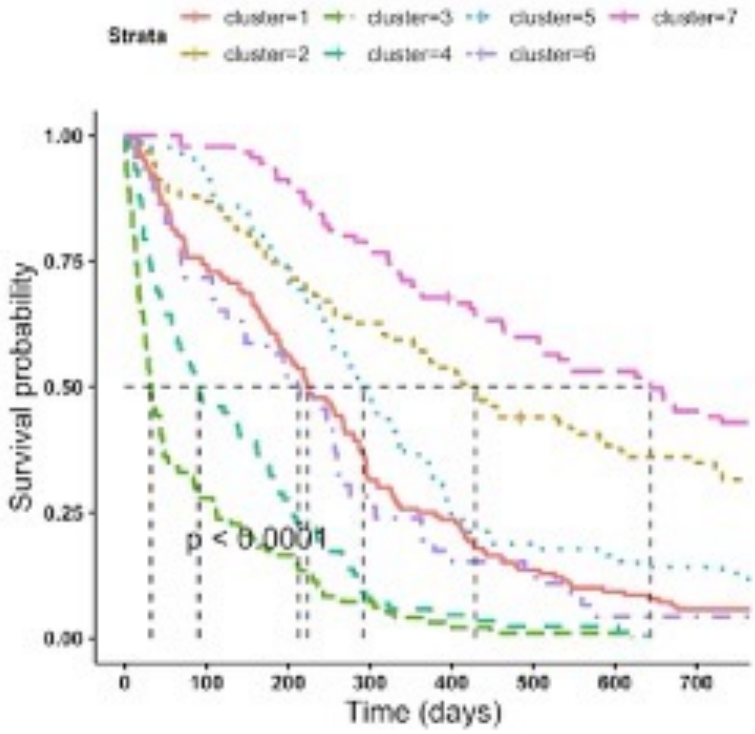


Back-up slide 2







Back-up slide 3

Variable	N	Hazard ratio	p
cluster	1 140	Reference	
2	91	0.44 (0.33, 0.58)	<0.001
3	97	3.47 (2.66, 4.53)	<0.001
4	87	2.18 (1.66, 2.87)	<0.001
5	85	0.75 (0.57, 0.98)	0.04
6	46	1.04 (0.74, 1.45)	0.83
7	90	0.30 (0.22, 0.40)	<0.001