

AmsterdamUMCdb: Our experiences in sharing data responsibly

P.J. Thoral, intensivist, Amsterdam UMC, location VUmc

OSIM

Open Science
in Medicine





Disclosures

- Amsterdam UMC receives royalties from Pacmed for jointly developed models.

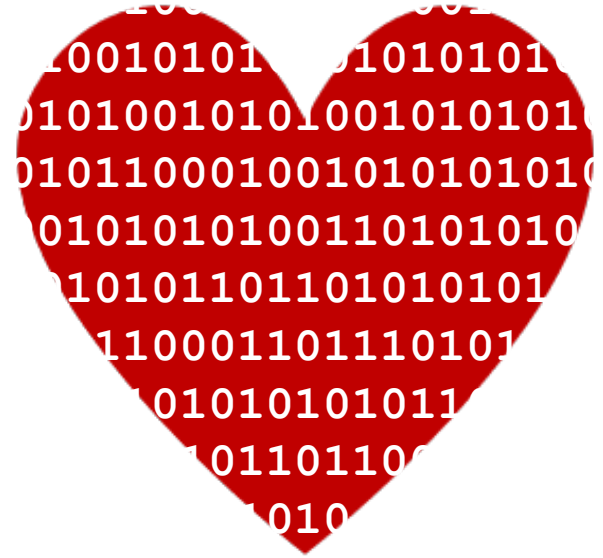




Photo by Massimo Catarinella, CC BY-SA 3.0





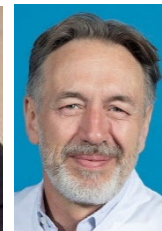
Paul Elbers
intensivist



Jan Peppink
clinical IT specialist



Ronald Driessen
clinical IT specialist



Eric Sijbrands
internist



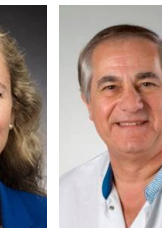
Erwin Kompanje
ethicist



Lewis Kaplan
intensivist



Heatherlee Bailey
intensivist



Jozef Kesecioglu
intensivist



Maurizio Cecconi
intensivist



Matthew Churpek
intensivist



Gilles Clermont
intensivist



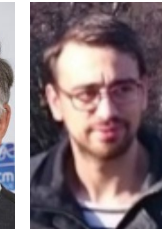
Mihaela van der Schaar
data scientist



Ari Ercole
intensivist



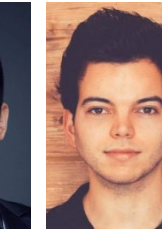
Armand Girbes
intensivist



Luca Roggeveen
physician-researcher



Tingjie Guo
PhD student



Lucas Fleuren
physician-researcher



Diederik Gommers
intensivist



Lilian Vloet
HAN/FCIC



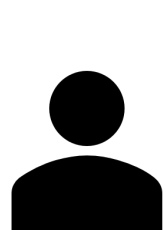
Bas van Bussel
intensivist



Iwan van der Horst
intensivist



Olaf Cremer
intensivist



Sander Rigter
intensivist



Tim Frenzel
intensivist



Hans van der Hoeven
intensivist



Rob Bosman
intensivist



Peter Pickkers
intensivist



Leo Heunks
intensivist



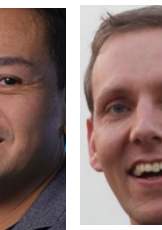
Arjen Slooter
intensivist



Nicole Juffermans
intensivist



Leo Celi
intensivist



Mark Hoogendoorn
data scientist



Amsterdam UMC
Laboratory for Critical Care
Computational Intelligence

Thorar PJ, Peppink JM, Driessen RH, et al. Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example. Crit Care Med. 2021;49(6):e563-e577. doi:10.1097/CCM.0000000000004916

Translational gap

93%
stays at

Model
prototyping and
development

MIND
× THE ×
GAP

Fine Food & Ale

MIND THE GAP

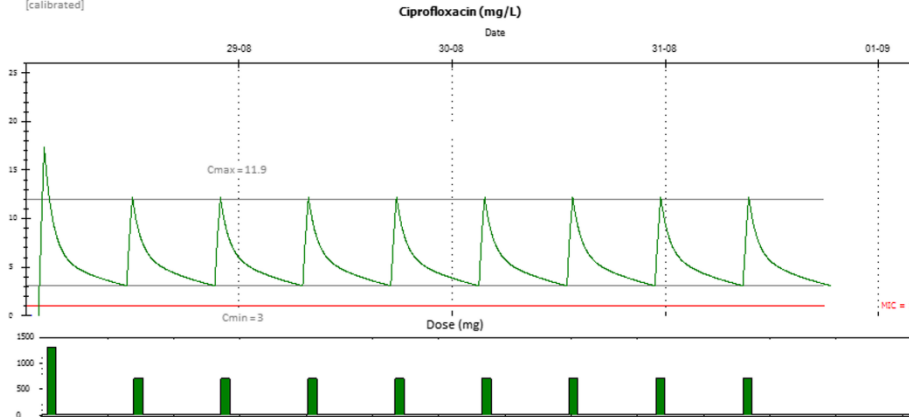




AutoKinetics & Pacmed Critical

AutoKinetics v1.2.0 - © Amsterdam UMC and OLVG

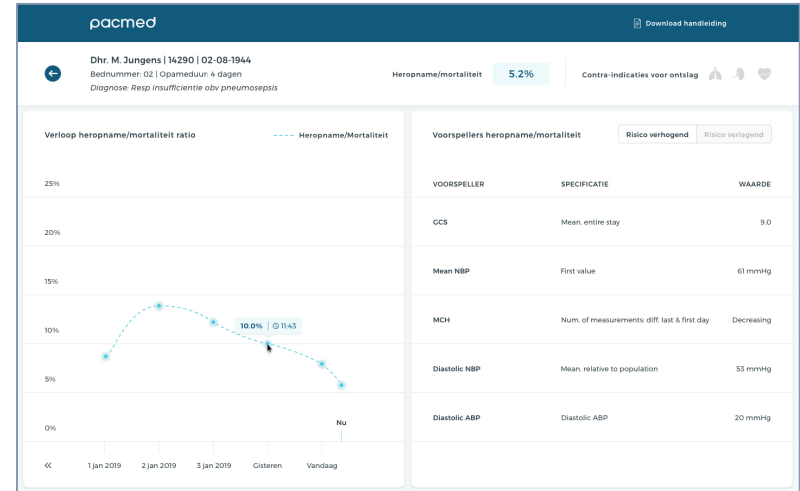
Model: Ciprofloxacin Khachman 2011
[calibrated]



Start with a loading dose Ciprofloxacin of 1450 mg on 28-08-2019 at 14:34. Continue with a maintenance dose of 850 mg every 8 hours. Start the maintenance dose on 28-08-2019 at 22:34.



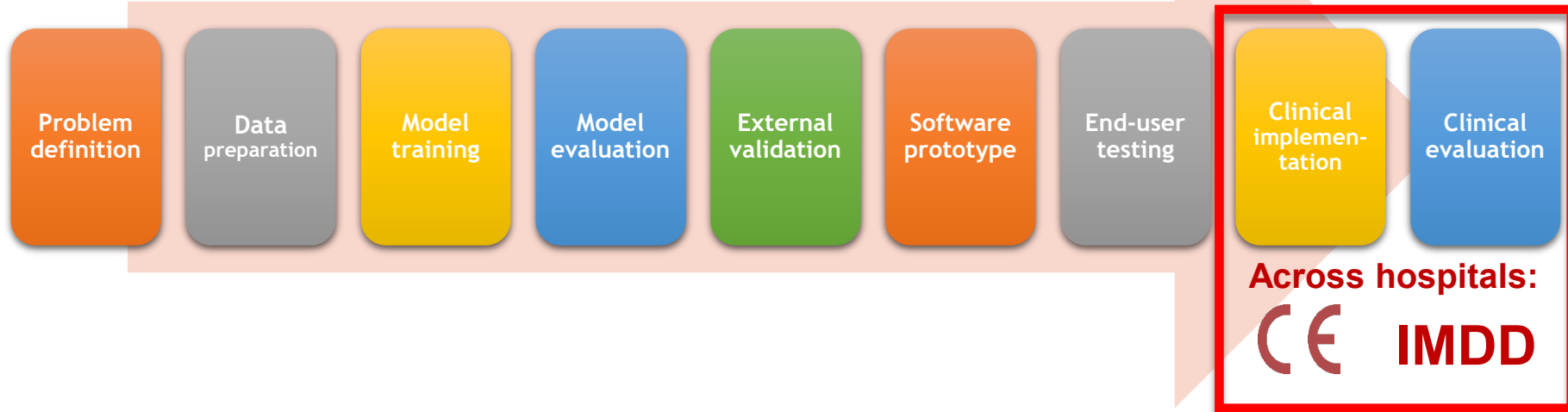
Roggeveen LF, et al. Right Dose Right Now: bedside data-driven personalized antibiotic dosing in severe sepsis and septic shock - rationale and design of a multicenter randomized controlled superiority trial. *Trials*. 2019 Dec 18;20(1):745.

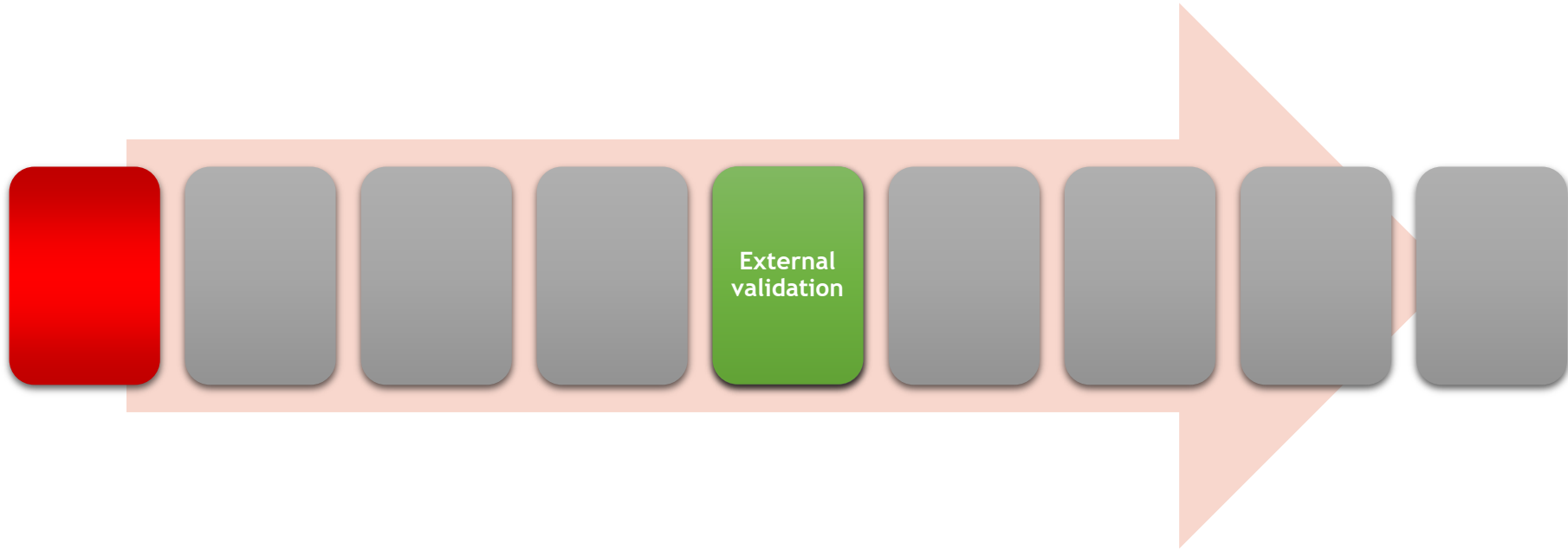


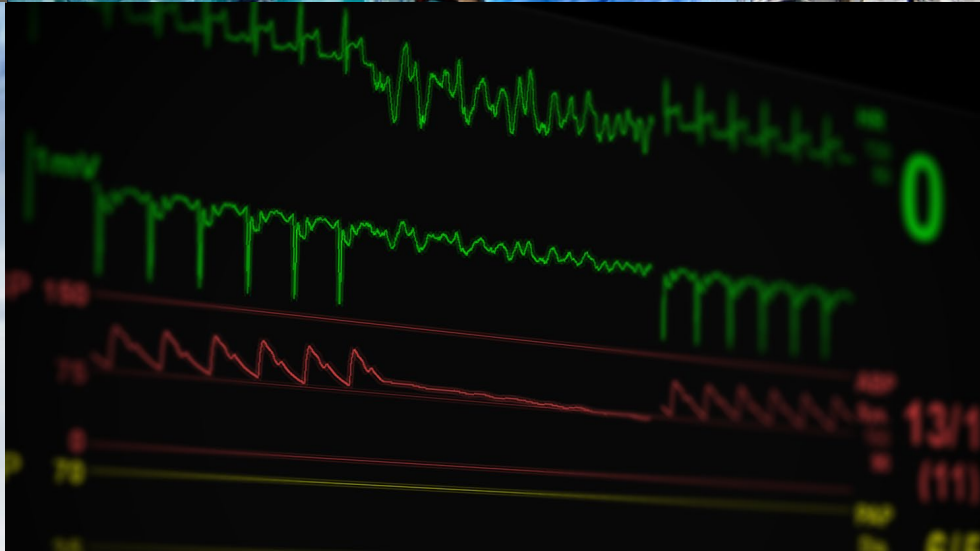
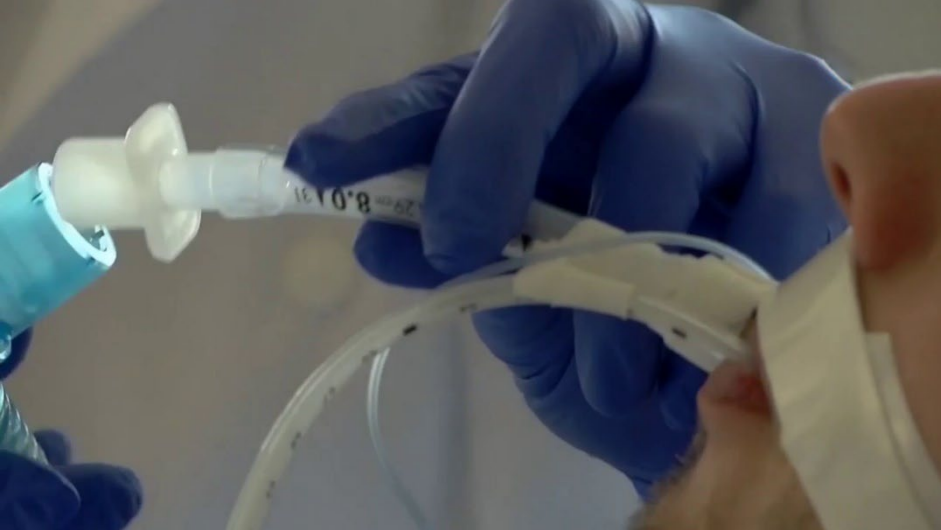
Thoral, P. J. et al. Explainable Machine Learning on AmsterdamUMCdb for ICU Discharge Decision Support: Uniting Intensivists and Data Scientists. *Critical Care Explorations* 3, e0529 (2021).



The road ahead: big data to the bedside

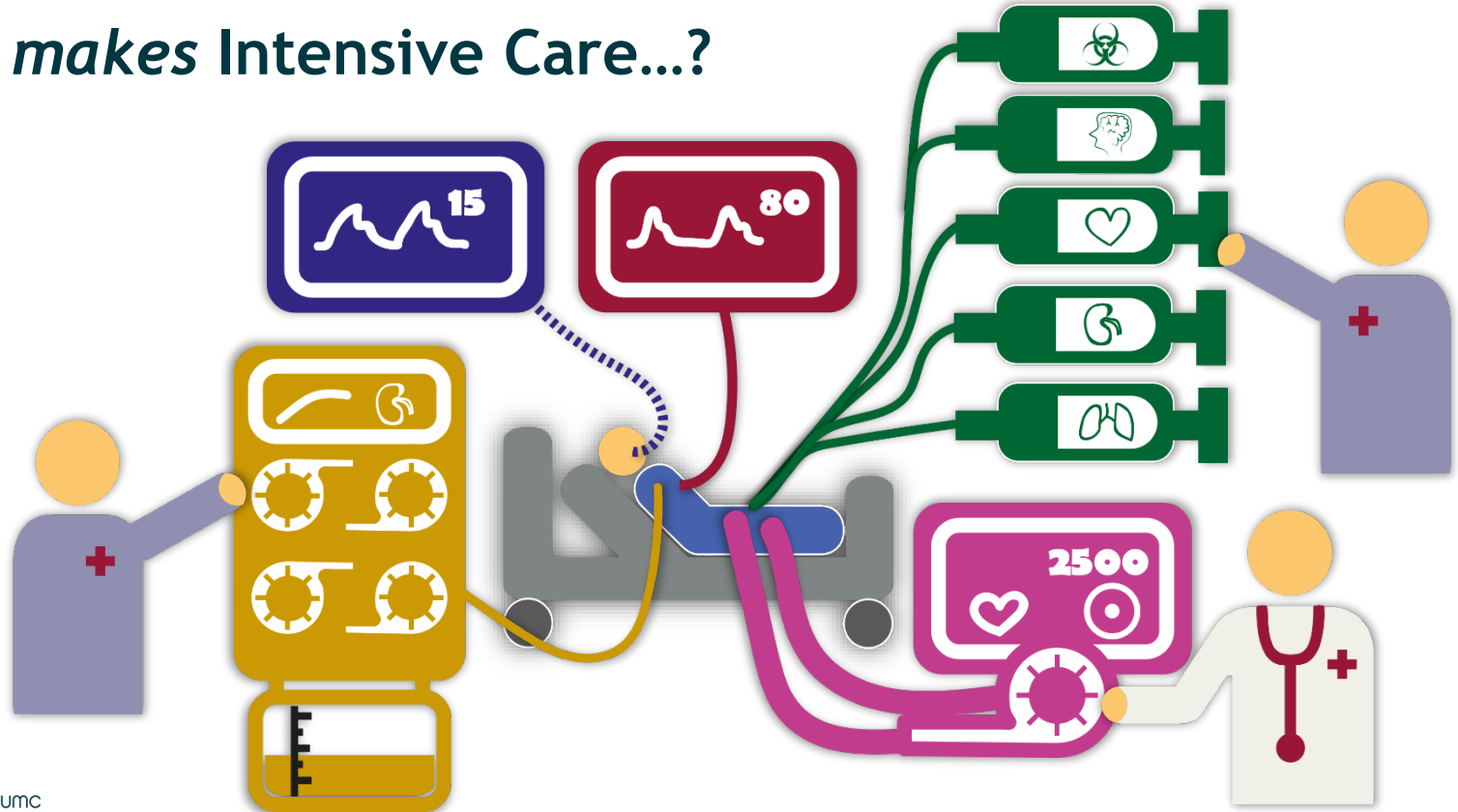






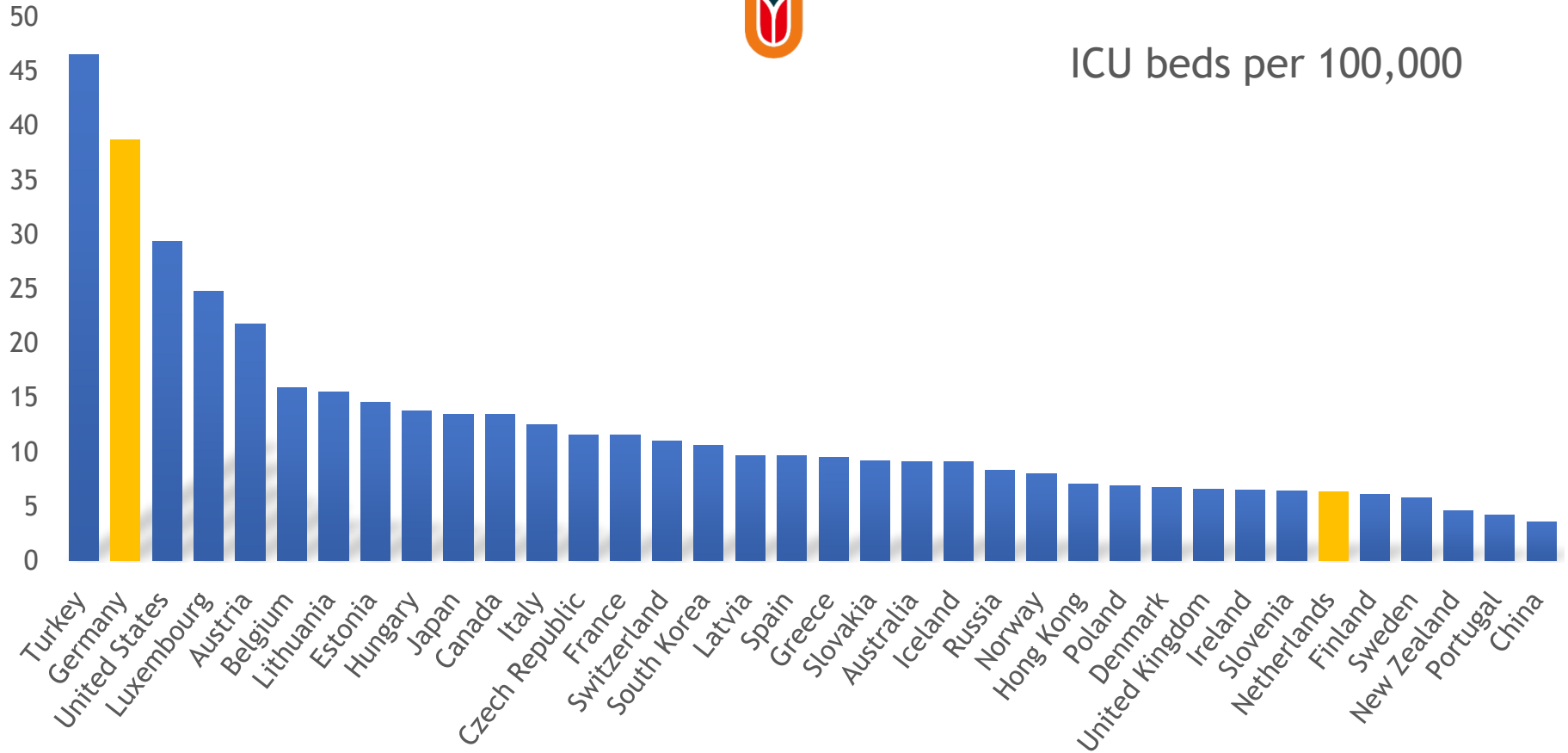


What *makes* Intensive Care...?



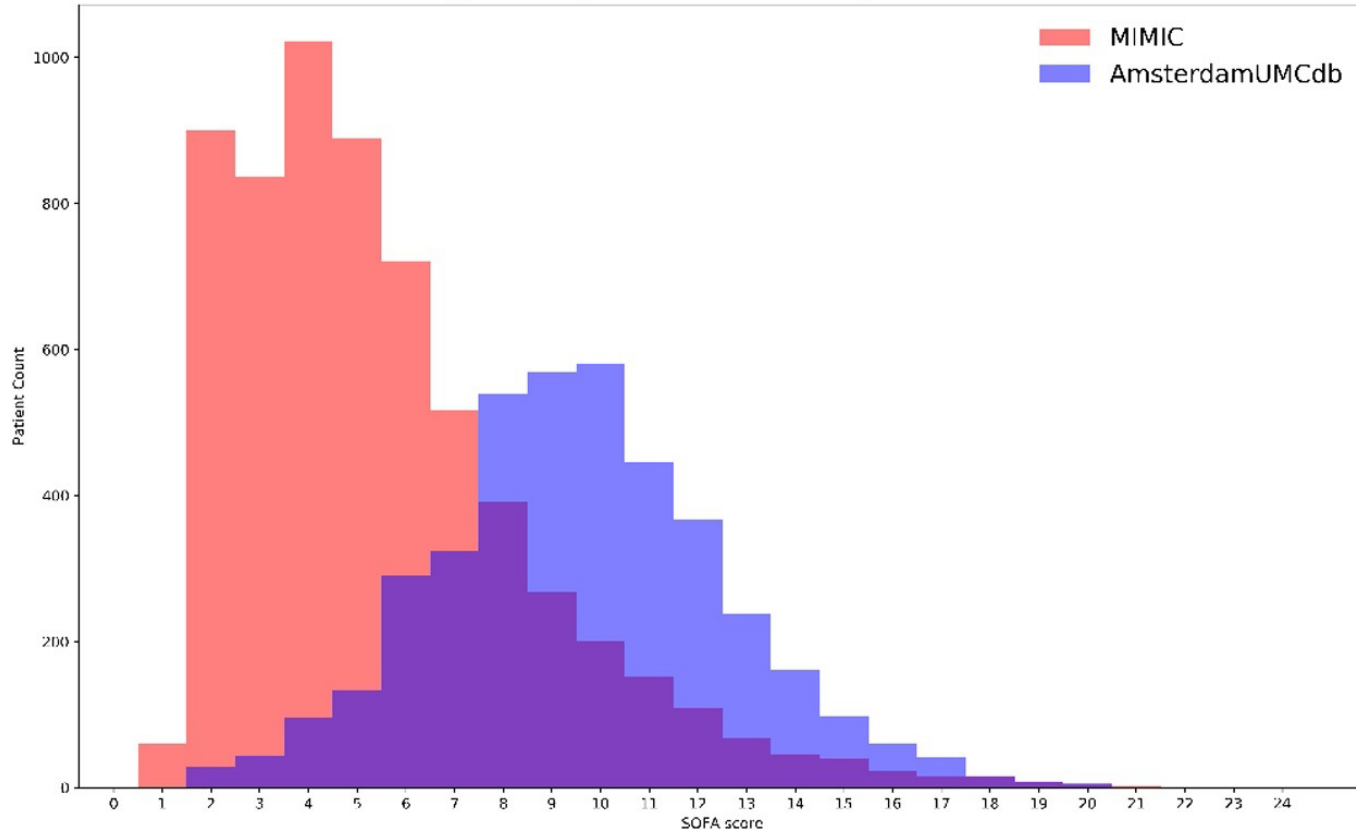


ICU beds per 100,000



SOFA scores

Degree of dysfunction of six organ systems

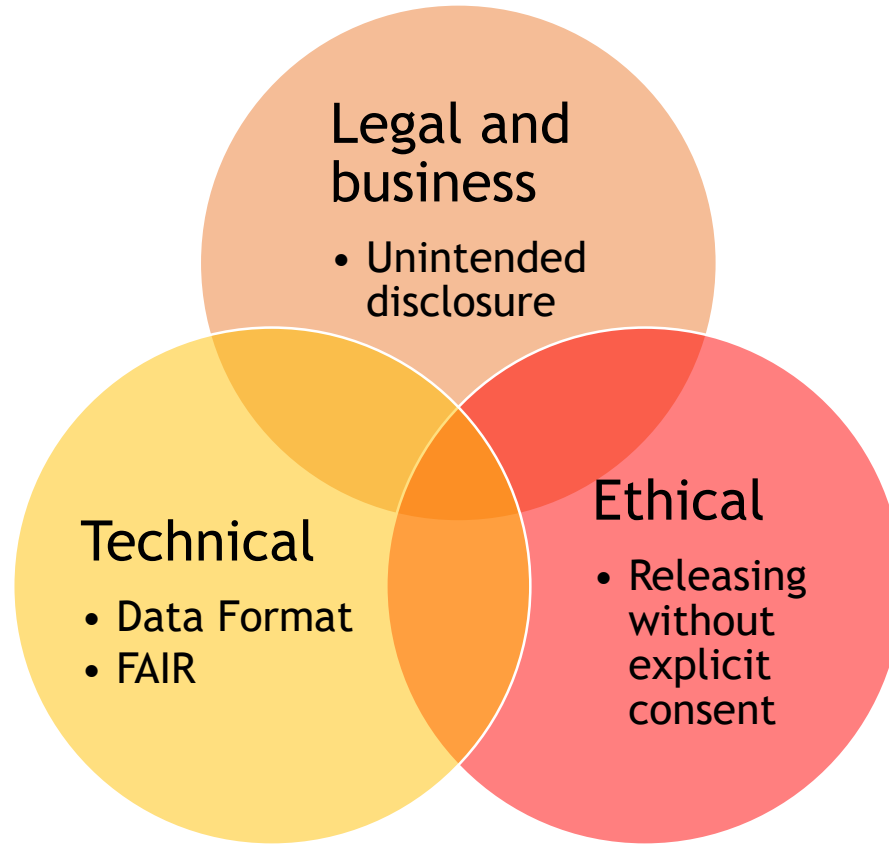


AmsterdamUMCdb

- First European freely accessible ICU database
November 2019
- Joint Data Science Initiative of SCCM and ESICM
- 23,106 ICU admissions from 2003 to 2016
- Tertiary referral academic medical center ICU
- High-resolution clinical data
- Fully GDPR (EU) and HIPAA (US) compliant



The hurdles



Motivate Stakeholders



Marcel van der Haagen
Privacy information officer



Nicole de Haan
Science Information Offices



Mark Kramer
Executive Board

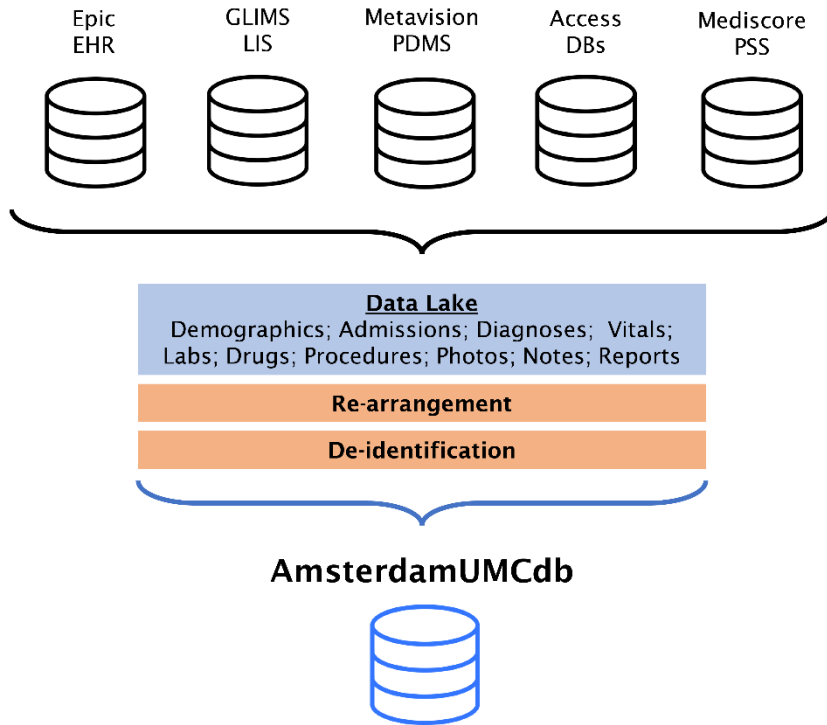


Michel Paardekooper
Privacy information officer



Jan Hol
Director Communications





drugitems	
admissionid	INTEGER
orderid	BIGINT
ordercategoryid	INTEGER
ordercategory	VARCHAR
itemid	INTEGER
item	VARCHAR
...	...

procedureorderitems	
admissionid	INTEGER
orderid	BIGINT
ordercategoryid	INT
ordercategory	VARCHAR
itemid	INT
item	VARCHAR
...	...

admissions	
patientid	INTEGER
admissionid	INTEGER
origin	VARCHAR
admittedat	BIGINT
admissionyeargroup	VARCHAR
lengthofstay	SMALLINT
...	...

numericitems	
admissionid	INTEGER
itemid	BIGINT
item	VARCHAR
tag	VARCHAR
value	FLOAT
unitid	INT
...	...

freetextitems	
admissionid	INTEGER
itemid	BIGINT
item	VARCHAR
value	VARCHAR
comment	VARCHAR
measuredat	BIGINT
...	...

listitems	
admissionid	INTEGER
itemid	BIGINT
item	VARCHAR
valueid	INT
value	VARCHAR
measuredat	BIGINT
...	...

processitems	
admissionid	INTEGER
itemid	BIGINT
item	VARCHAR
start	BIGINT
stop	BIGINT
duration	BIGINT

Cohort



Distinct patients:

20,109

ICU admissions:

23,106

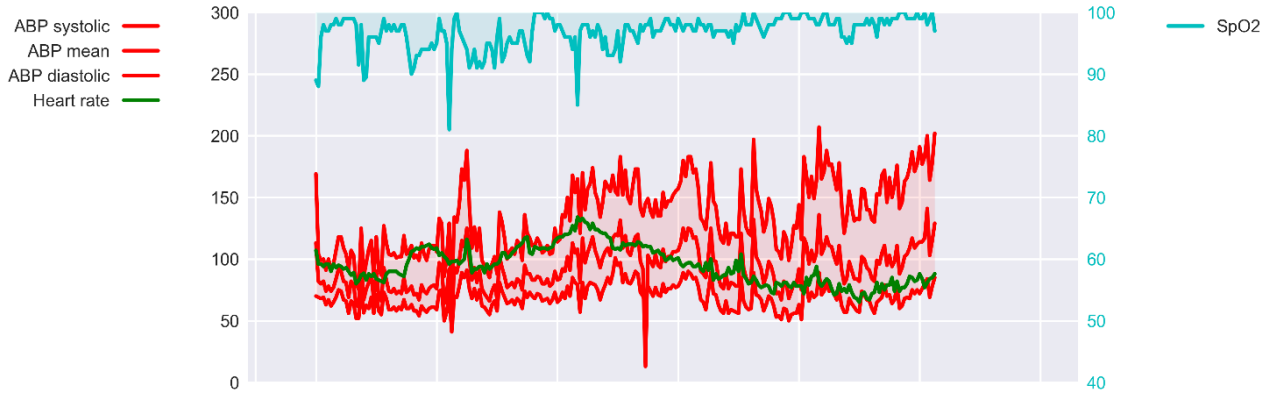
Resolution:

Up to one data point per minute (device data)

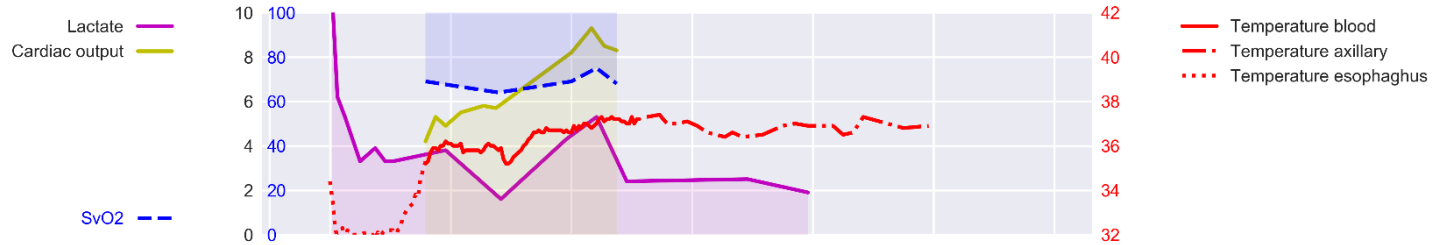
Reason for admission	
Cardiothoracic surgery, no. (%)	5935 (25.69)
Sepsis, no. (%)	3136 (13.57)
Respiratory failure, no. (%)	1568 (6.79)
Neurosurgery, no. (%)	1619 (7.01)
Trauma, no. (%)	902 (3.90)
Gastro-intestinal surgery, no. (%)	1149 (4.97)
Vascular surgery, no. (%)	1037 (4.49)
Cardiac arrest, no. (%)	959 (4.15)
Neurologic disorders (non-traumatic), no. (%)	628 (2.72)
Cardiac disorders (incl. cardiogenic shock) no. (%)	538 (2.33)



Vitals



Circulation



Observations

Glasgow Coma Scale	E1M1V1	E4M1V1	E4M5V1	E4M6V1	E4M6V1	
Heart rhythm	Sinus Tac	Sinus Tac	Sinus Tac	Sinus Tac	NSR	NSR
Respiratory support	PC	PS/CPAP	PS/CPAP	PS/CPAP	PS/CPAP	
FI02	100	60	64	49	39	
P peak	35	28	25	27	11	
PEEP	10	14	14	14	8	
Oxygen delivery device					HME	Prongs
Oxygen flow					3	5





Why the US beat us: Safe Harbor (HIPAA)

Safe Harbor method: removal of specific elements
Names
All geographic subdivisions smaller than a state
All elements of dates (except year)
Telephone numbers
Vehicle identifiers and serial numbers, including license plate numbers
Fax numbers
Device identifiers and serial numbers
Email addresses
Web Universal Resource Locators (URLs)
Social security numbers
Internet Protocol (IP) addresses
Medical record numbers
Biometric identifiers, including finger and voice prints
Health plan beneficiary numbers
Full-face photographs and any comparable images
Account numbers
Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of the Privacy Rule



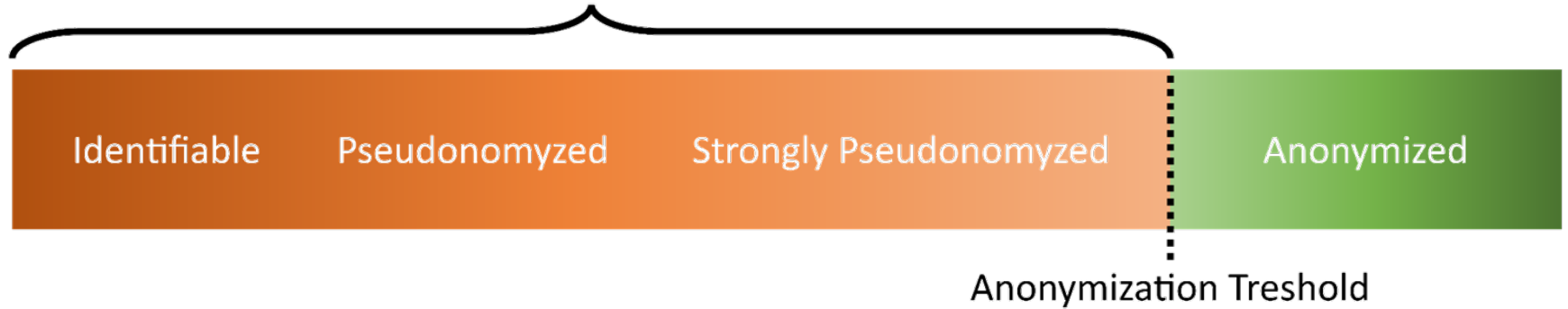


GDPR. Recital 26

- “**the principles of data protection should not apply to personal data rendered anonymous** in such a manner that the data subject is not or **no longer identifiable**. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, **considering all objective factors**, such as the costs of and the amount of time required for identification, the available technology at the time of the processing and technological developments”



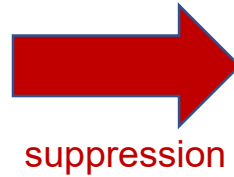
Personal Data



De-identification techniques

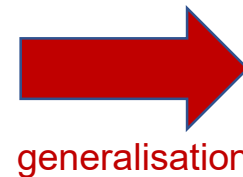


Patientid	Weight (kg)
0	198
1	78
2	56
3	84



Patientid	Weight (kg)
0	NULL
1	78
2	56
3	84

Patientid	Weight (kg)
0	198
1	78
2	56
3	84



Patientid	Weight (kg)
0	110+
1	70-80
2	50-60
3	80-90

De-identification metrics



Quasi-identifiers

Patientid	Age	Sex	Diagnosis
0	18-40	M	Trauma
1	40-50	F	CABG
2	50-60	F	Cardiac arrest
3	70-80	M	Lymphoma
4	70-80	M	Lymphoma
5	70-80	F	Aorta-valve replacement
6	70-80	F	Sepsis
7	70-80	F	Pneumonia

k -anonymity = 1

k -anonymity = 1

k -anonymity = 1

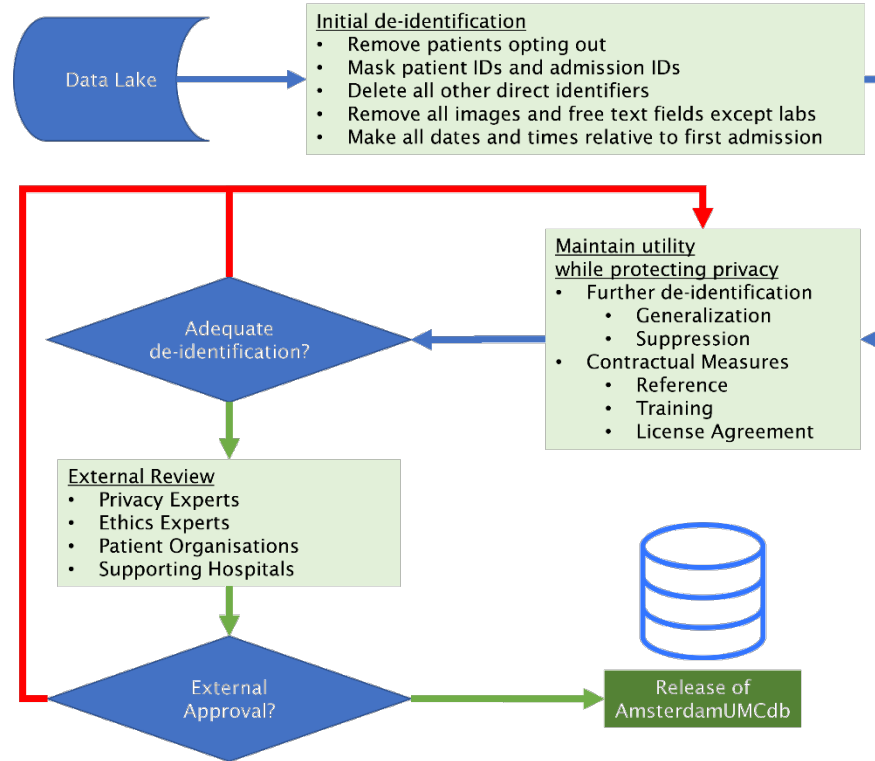
k -anonymity = 2

k -anonymity = 3

l -diversity = 1

l -diversity = 3

Know your adversary!



		Hypothetical Adversary		
		Friendly Researcher	Rogue Researcher	Rogue Insurance Company
Assumed Background Knowledge				
Gender		X	X	X
Age		X	X	X
Weight		X	X	
Height		X	X	
Admission date		X	X	X
Survival at discharge		X	X	X
Number of ICU admissions				X
Assessment of re-identification risk				
P(access)		1.00	1.00	0.27
P(intention)		0.20	0.10	0.10
Average Risk				
P(re-identification)		0.047	0.047	0.009
k-anonymity		89	89	682
l-diversity		26	26	65
Maximum Risk				
P(re-identification)		0.50	0.50	0.50
k-anonymity		2	2	2
l-diversity		2	2	2

Credentialed access:

- Course in Research Ethics
- Signature of an intensivist





Ethical considerations

- Lead by independent team
- Informed consent for secondary use of **data difficult to achieve** for (ICU) patients
 - Large number of patients
 - Relatively high mortality
 - Impaired levels of consciousness
- **Selection bias** due to refusals and untraceable patients
- Obtaining individual consent for sharing ICU database is neither feasible nor desirable
- Ethical principal: *duty of easy rescue*

Low burden

- Minimal risk when properly de-identified

Large benefit

- Potentially improving outcomes for future patients

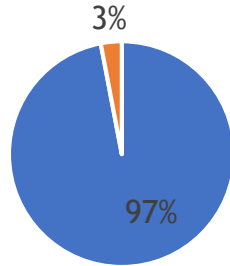
Act!





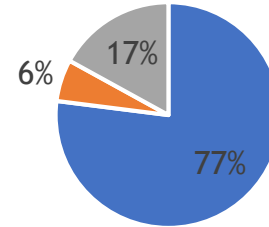
Sharing your data: Dutch Survey

Did you give permission to share your data and/or tissue? (n=1510)



■ Yes ■ No

Would you give permission to share your data and/or tissue? (n=4399)



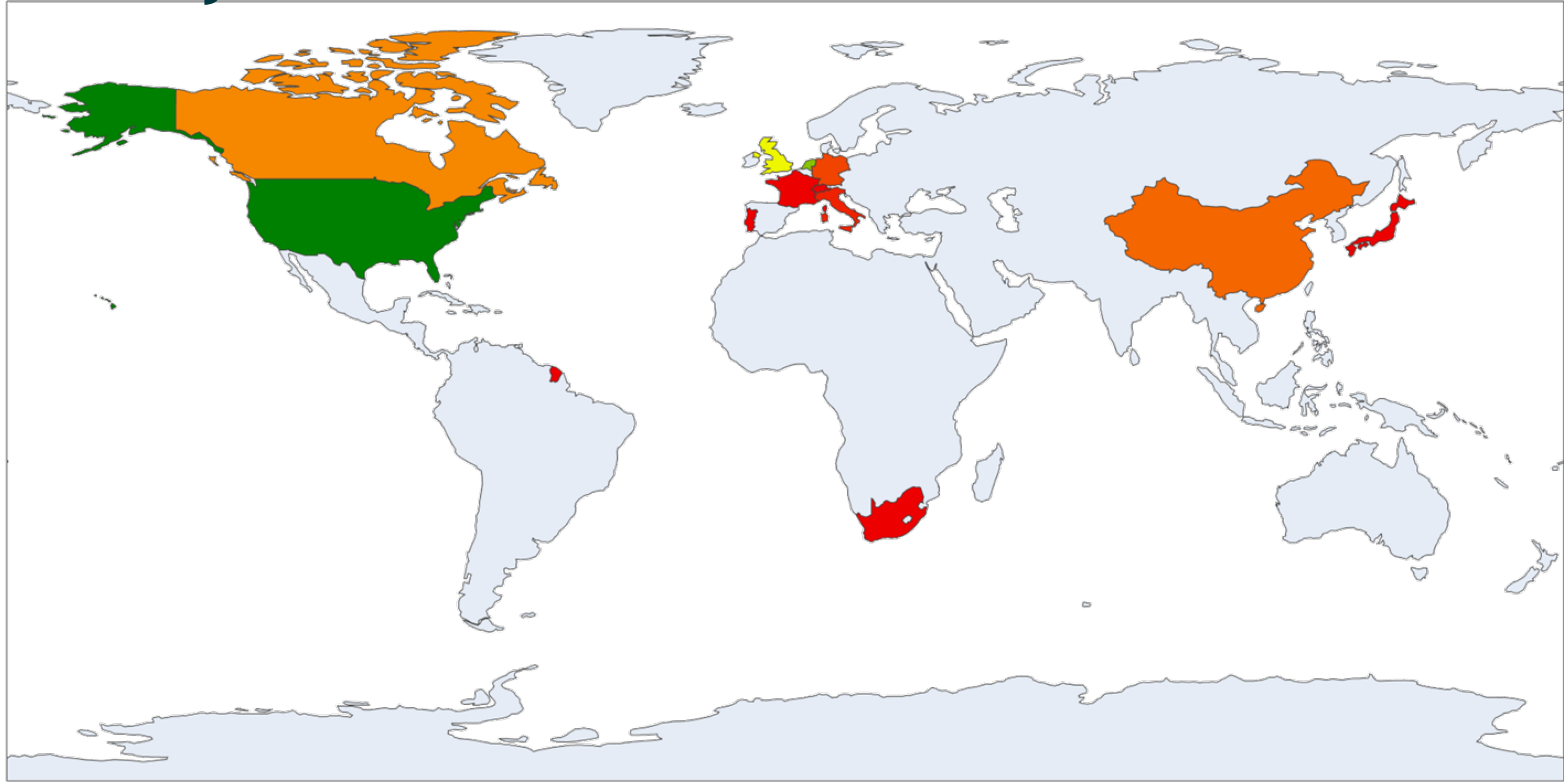
■ Yes ■ No ■ Unsure



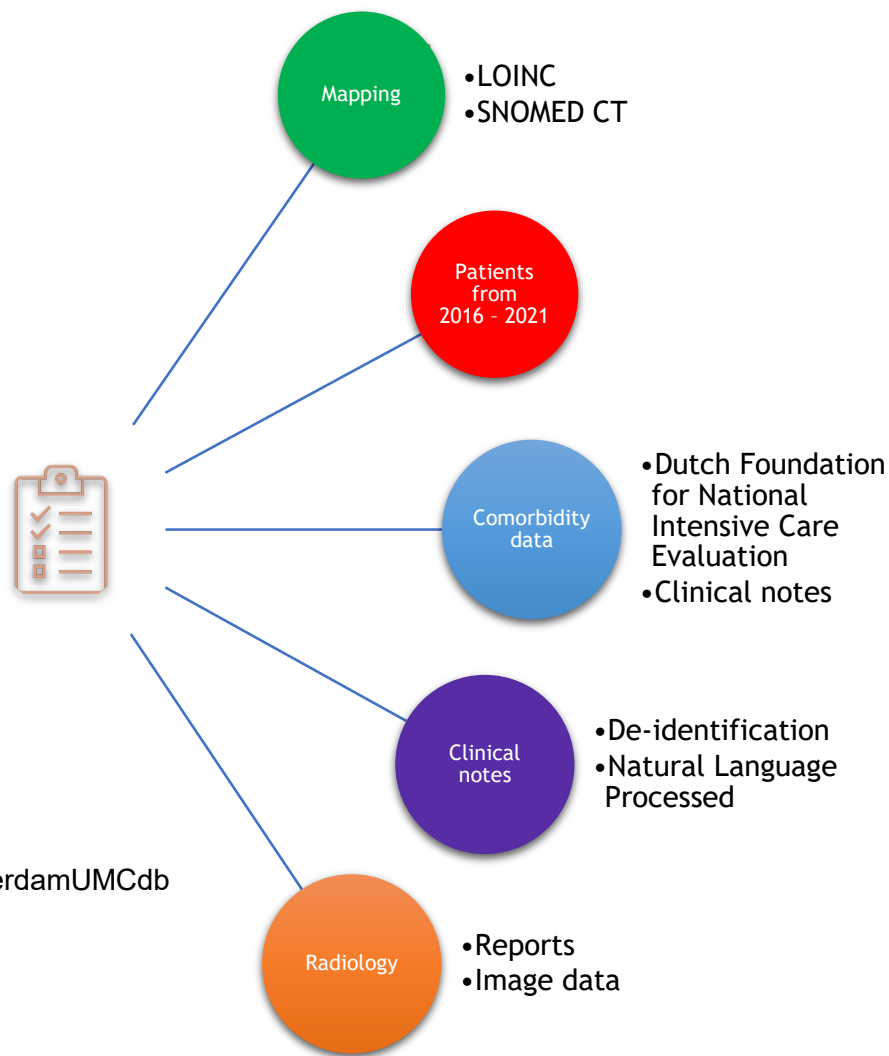
AmsterdamUMCdb requests per country



Requests granted



Future work



github.com/AmsterdamUMC/AmsterdamUMCdb

Image © Pixabay.com

COVID-19 ICU Data Sharing

Treatment uncertainty:
the need to Share Data

Image © ANP





The Dutch ICU Data Warehouse

Dutch ICUs will start collaborating by sharing large amounts of routinely collected data to improve the quality of care and treatment strategies for future critically ill patients





Conclusions

- AmsterdamUMCdb addressed technical, legal, ethical and privacy challenges
- Risk-based de-identification strategy for US and European regulations
- SCCM and ESICM encourage other ICUs to follow the example to share data responsibly



Request access