

# Big Data a Data Science v AIM



Nemocnica AGEL  
Krompachy

merea®

**Štefan Trenkler**

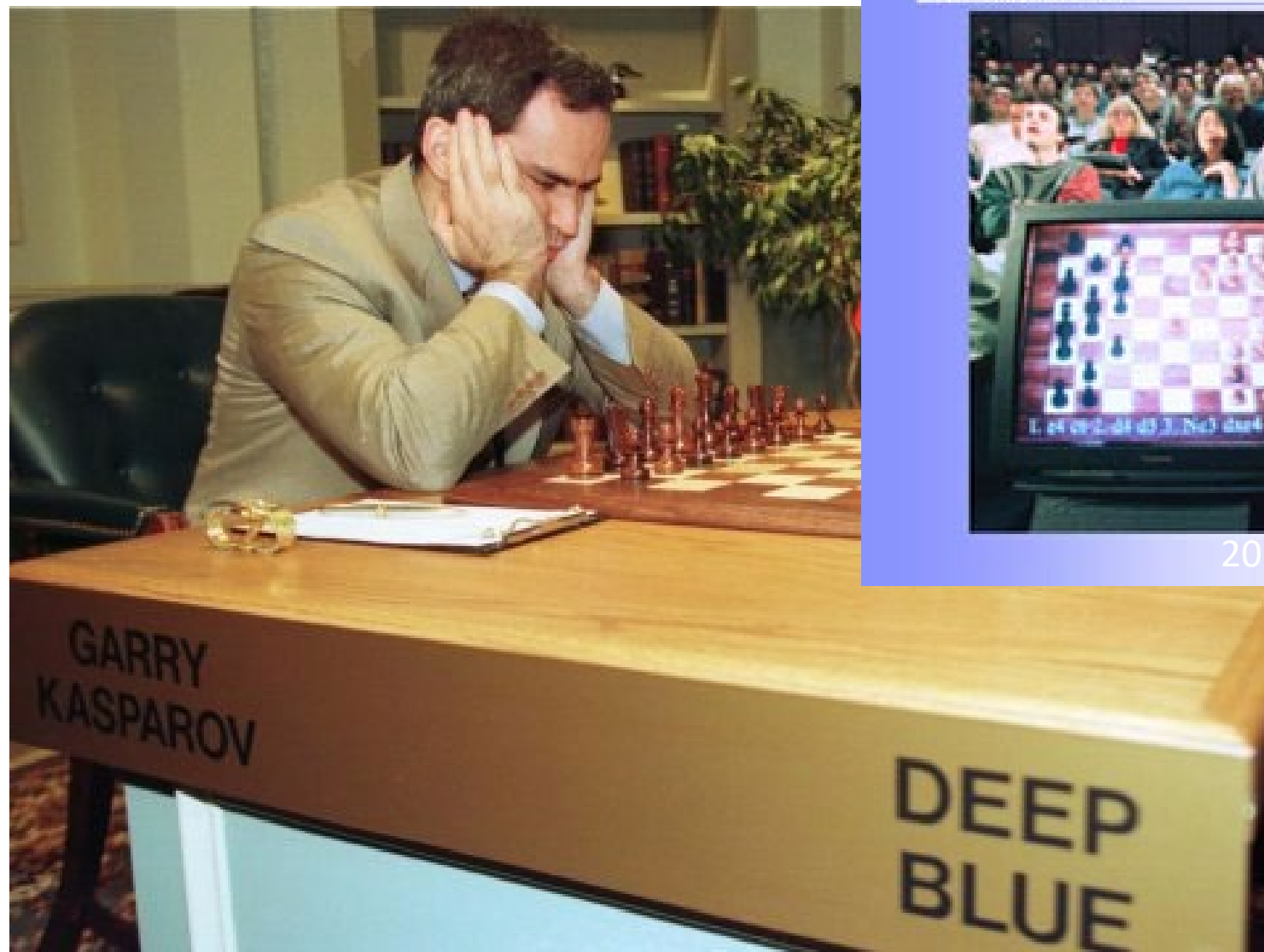
OAIM Nemocnica Agel Krompachy

ETC Merea Bratislava



Kongres CSARIM Brno 2022

# 1997: umelá inteligencia?



May 11th, 1997  
**Computer won world champion of chess**  
(Deep Blue) (Garry Kasparov)



2022

(Reuters = Kyodo News)

# Čo sa udialo

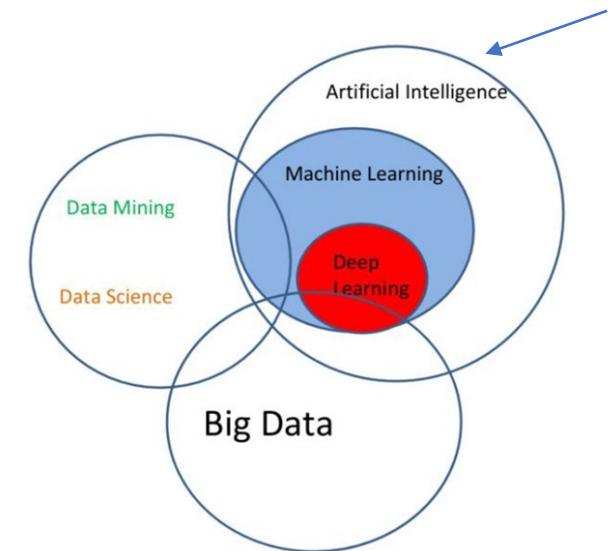
- 1. Počítače** – menšie, rýchlejšie, lacnejšie
- 2. Softvér** výkonnejší
- 3. Internet** + sociálne siete, networking, IoT
- 4. Miniaturizácia** - senzory ako mikročipy, tatoo, IoT
- 5. Digitalizácia všetkého** → **BIG Data**



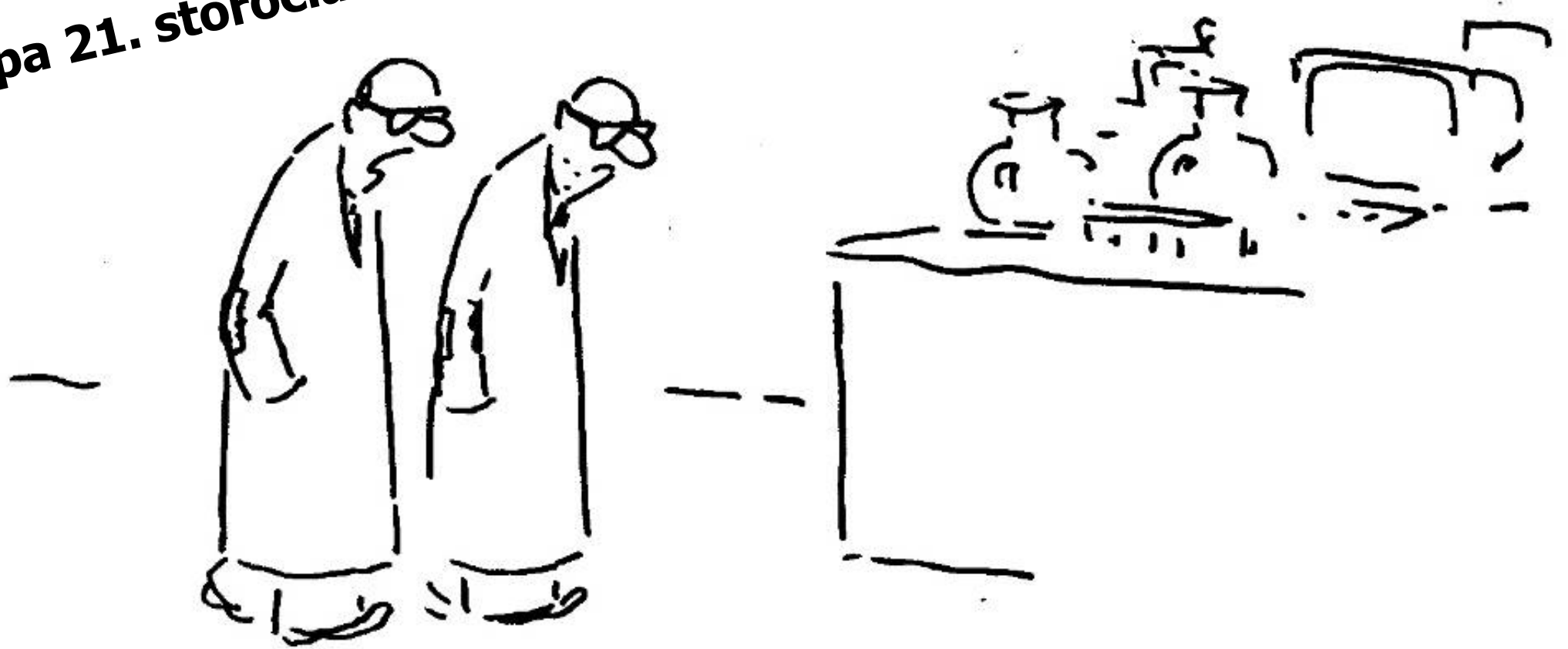
**21. storočie - čas na umelú inteligenciu**

# Big data

- **Digitálne** dáta, ktoré sú **tera, peta, exa, zetta, yotta ( $10^{24}$ )**
  - veľkom **objeme**
  - veľkej **rozmanitosti**
  - **veľkou rýchlosťou (real time)**
- nie je možné ich spracovať tradičnými systémami na spracovanie údajov
- Data science
- **Umelá inteligencia** – analýza dát
  - strojové učenie (štatistické modelovanie)
  - umelé neurónové siete
    - hlboké učenie
    - data mining



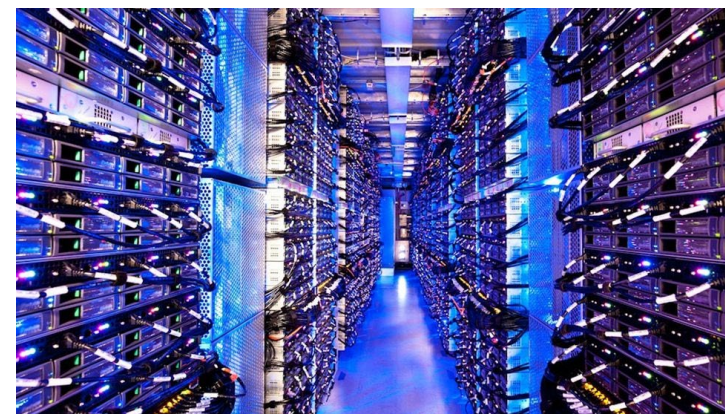
Dáta = ropa 21. storočia



MUSÍME SE SMÍRIŤ S TÍM, ŽE NAŠIM Hlavným <sup>DOMATSTVÍM</sup>  
JSOU INFORMACE.

# Dátové centrá

tera, peta, exa, zetta, **yotta** ( $10^{24}$ )



# Analýza big data - UI



## 1. Machine learning:

- **predikcia** na základe známych (big) údajov z tréningu
- reprodukovať existujúce znalosti

Príklad: Watson, IBM - analýza obrazov (RTG, MRI, CT, sono...); šach ...





## 2. Data mining/knowledge discovery:

- **odhalenie skrytých súvislostí v (big) dátach**

Príklad: Google Deepmind Health Project

A to všetko bleskovo na naozaj veľkých údajoch

## Za dva týdny nejlepším hráčem. Umělá inteligence porazila mistra v DOTA2

14. srpna 2017 5:56    



Hráči trénují měsíce.  
Umělá inteligence se  
za dva týdny stala mistrem.



# GROWTH OF HEALTHCARE DATA

(one exabyte = one billion gigabytes)

tera, peta, exa, zetta, **yotta...**

48% increase annually

153  
exabytes

2,314  
exabytes

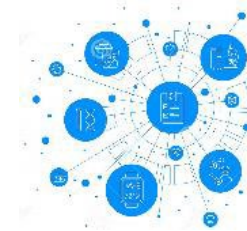


2013

2020

Source: <http://ow.ly/2okF30fti5O>

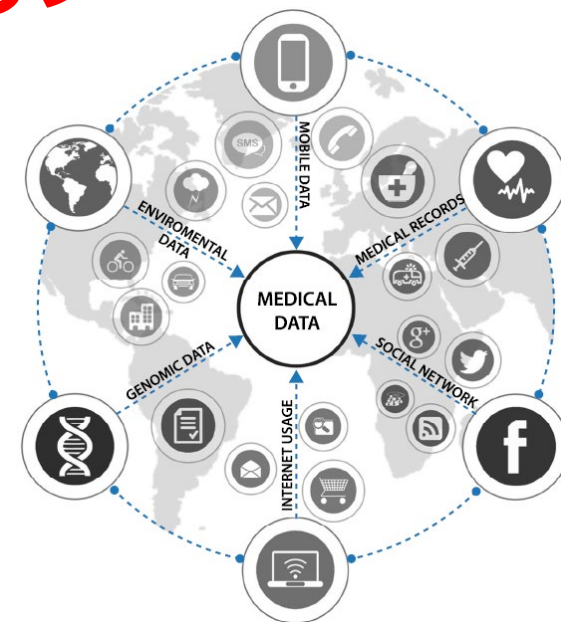
# Medicínske (big) dáta - zdroje



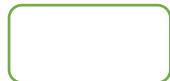
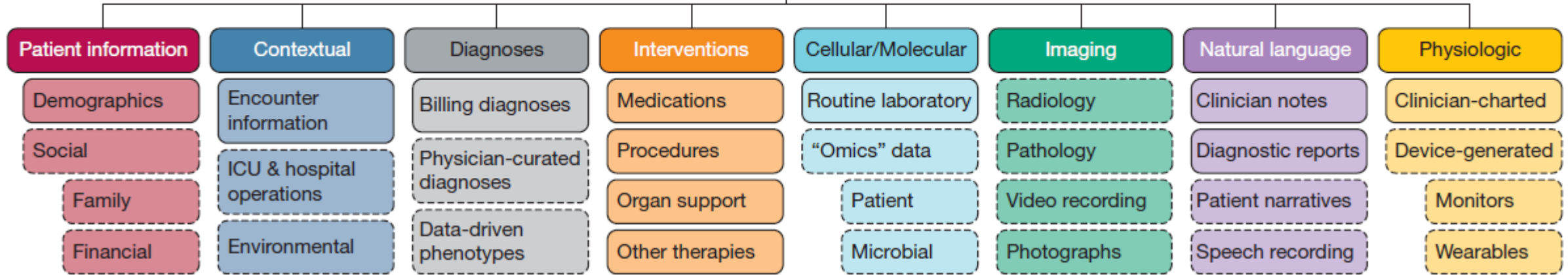
1. Digitalizovaný chorobopis – **EHR**, voľný text
2. Údaje z prístrojov, monitorov (on-line), nositeľnej elektroniky, sociálnych sietí (5G)...
3. Zobrazovacie metódy
4. Laboratória, genomika (omics), mikrobióm...
5. Administratívne dáta
6. Výsledky (observačných) štúdií, medicínska literatúra

...heterogénne...

- Medicína: data science (driven) podporovaná lekármi
- Lekár: Klinický data scientists (+vedomosti, múdrosť)  
**Interpretácia** - kauzalita, validácia



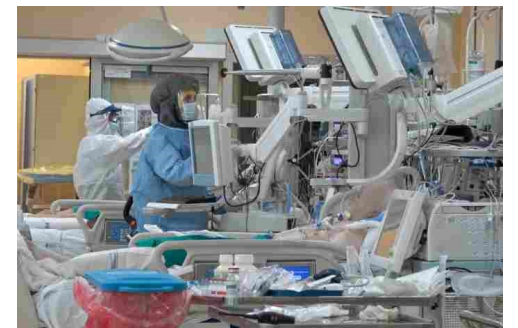
# Big Data in the ICU



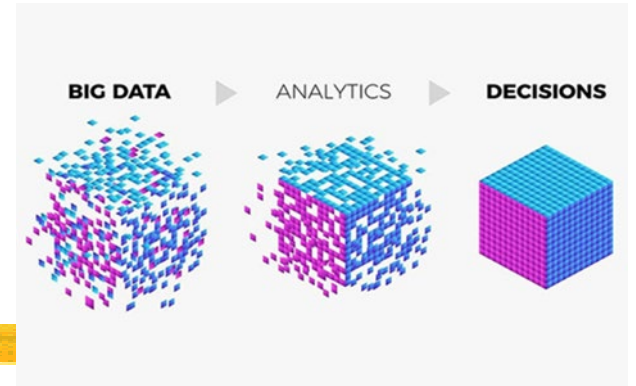
Všeobecne dostupné

# Intenzívna medicína – optimálne prostredie pre UI

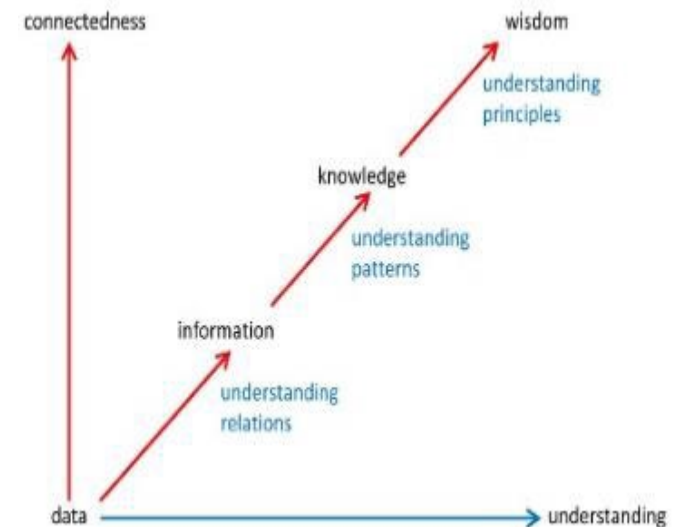
- Heterogénni, nestabilní, kriticky chorí pacienti, MOF; neistota
- Komplexné prostredie
- Prostredie s „big data“ (popisy, monitory, laboratória, obrazy...) - veľké, rôznorodé, rýchle...
- Viacero liečebných postupov (lieky)
- Zlyhávanie rozsiahlych RCT, odporúčania?
- **Realita:**
  - **variabilita**, náklady
  - postup **podľa** skúsenosti, inštinktu...



# Data - nestačia



- Organizovať – štruktúrovať – analyzovať – zobratíť – použiť
- Údaj  $\Rightarrow$  informácia  $\Rightarrow$  vedomosť  $\Rightarrow$  **múdrost'**
- Big data analytics (BDA)
- Zmysel, užitočnosť, hodnota, **relevancia** pre konkrétnu potrebu, situáciu, kontext
- Použiteľná a užitočná informácia, **user friendly**



# ICU a UI – možné uplatnenie

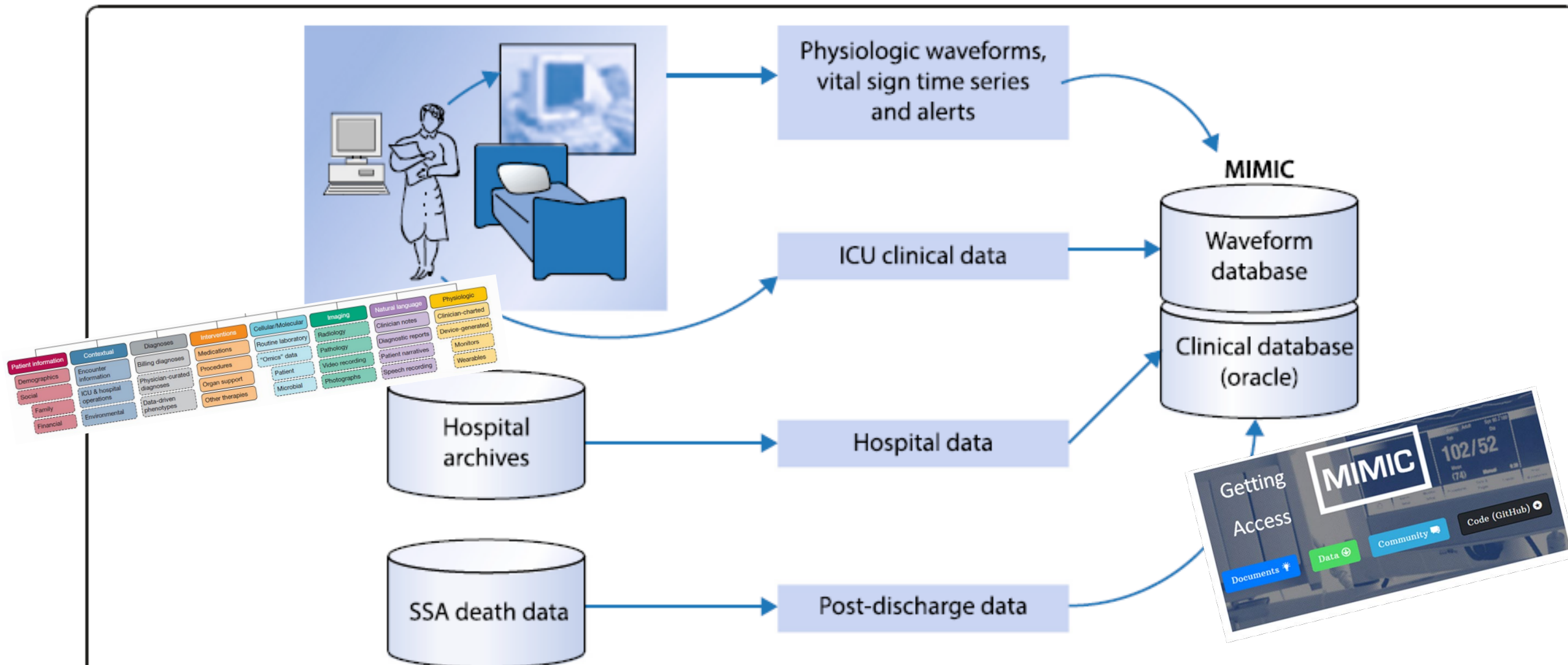
- 1. Nájsť komplexné vzťahy** v (big) heterogénnych dátových súboroch
- 2. Analýza** údajov z prístrojov v reálnom čase (on-line)
- 3. Predikcia** rizika, priebehu, výsledku  
- prevencia MOF, NZS, infekcie....
- 4. Podpora klinického rozhodovania** (kognitívny asistent)
- 5. Personalizácia** medicíny (precision, personal)



# Databázy a intenzívna medicína

- **MIMIC III** (MIT Boston), BIDMC ICU 5x populácia  
(Medical Information Mart for Intensive Care)
  - **kompletné** data z hospitalizácie
  - 53 500 hospitalizácií; otvorený zdroj
- **eICU** - Collaborative Research Database (Philips)
  - 128 nemocníc, dostupné na požiadanie
- .....
- .....
- Extrapolácia? Vlastný (EU) MIMIC III?





**Figure 2 The MIMIC database.** SSA: social security administration (figure courtesy of the Laboratory of Computational Physiology, Massachusetts Institute of Technology).



 Database Credentialed Access

# MIMIC-IV

Alistair Johnson , Lucas Bulgarelli , Tom Pollard , Steven Horng , Leo Anthony Celi ,

Published: June 12, 2022. Version: 2.0

**When using this resource, please cite:** [\(show more options\)](#)

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2022). MIMIC-IV (version 2.0). *PhysioNet*. <https://doi.org/10.13026/7vcr-e114>.

**Please include the standard citation for PhysioNet:** [\(show more options\)](#)

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. 101 (23), pp. e215–e220.

# Príklady analýzy big data

1. Riadenie pumpy pri anestézii
2. Predikcia sepsy
3. Liečba sepsy
4. Podskupiny sepsy
5. Nositeľná elektronika



# 1. Predikcia

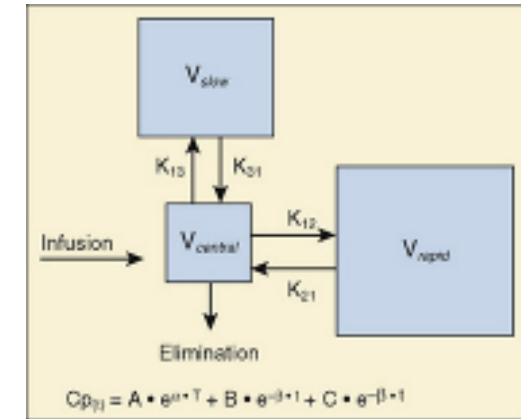
## Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanyl

(ANESTHESIOLOGY 2018; 128:492-501)

### A Deep Learning Approach

Hyung-Chul Lee, M.D., Ho-Geol Ryu, M.D., Ph.D., Eun-Jin Chung, M.D., Chul-Woo Jung, M.D., Ph.D.

- TIVA: propofol + remifentanyl (TCI)
- BIS: predikcia verzus skutočná hodnota
  1. FK + FD modely (Schinder/Minto)

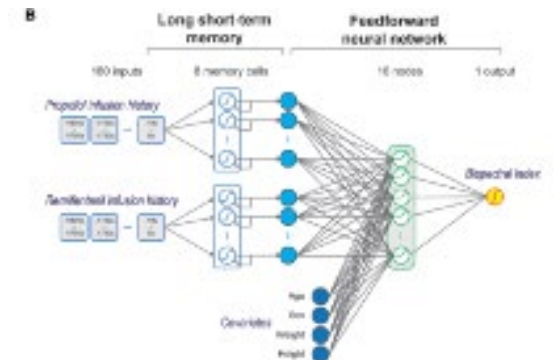


## 2. Metóda hlbokého učenia

- long short term memory (koncentrácie)
- feed forward neural network (demografia)

Žiadna znalosť FK a FD

ale **2 milióny údajov** (big data) z pump a BIS



# 1. Predikcia

## Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanyl

(ANESTHESIOLOGY 2018; 128:492-501)

*A Deep Learning Approach*

Hyung-Chul Lee, M.D., Ho-Geol Ryu, M.D., Ph.D., Eun-Jin Chung, M.D., Chul-Woo Jung, M.D., Ph.D.

- TIVA: propofol + remifentanyl (TCI)
- BIS: predikcia verzus skutočná hodnota

1. FK + FD modely (Schindler)

2. Metóda hĺbkového učenia

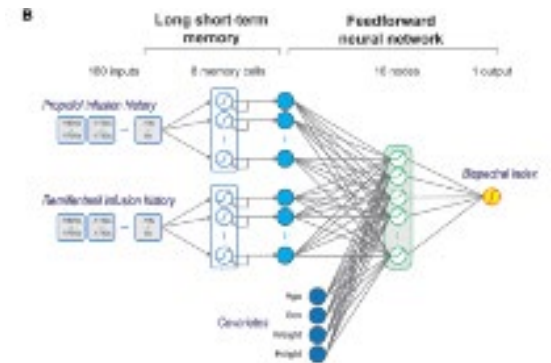
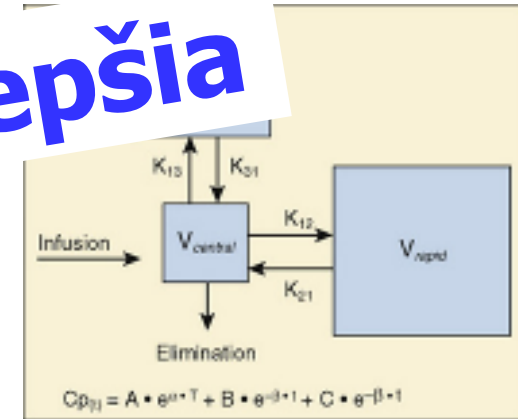
- long short term memory (koncentrácie)
- feed forward neural network (demografia)

Žiadna znalosť FK a FD

ale **2 milióny údajov** (big data) z pump a BIS



Umelá inteligencia bola lepšia



## 2. Predikcia

# InSight – 6 vitálnych parametrov, sepsa

		InSight	MEWS	SOFA	SIRS
AU-ROC		0,92	0,84	0,76	0,75
Senzitivita	Sepsa	0,98	0,98	0,82	0,82
	Ťažká sepsa	0,99	0,98	0,90	0,81
	Šok	1,00	1,00	0,99	0,91
Špecificita	Sepsa	0,95	0,72	0,32	0,51
	Ťažká sepsa	0,85	0,72	0,37	0,50
	Šok	0,99	0,91	0,58	0,49

Možnosť prenosu vlastných údajov

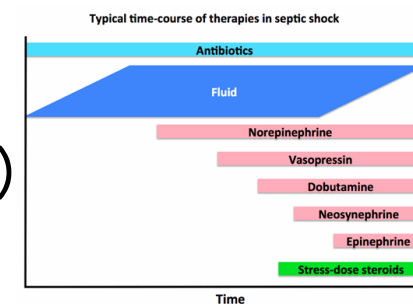
Mao, BMJ Open 2018

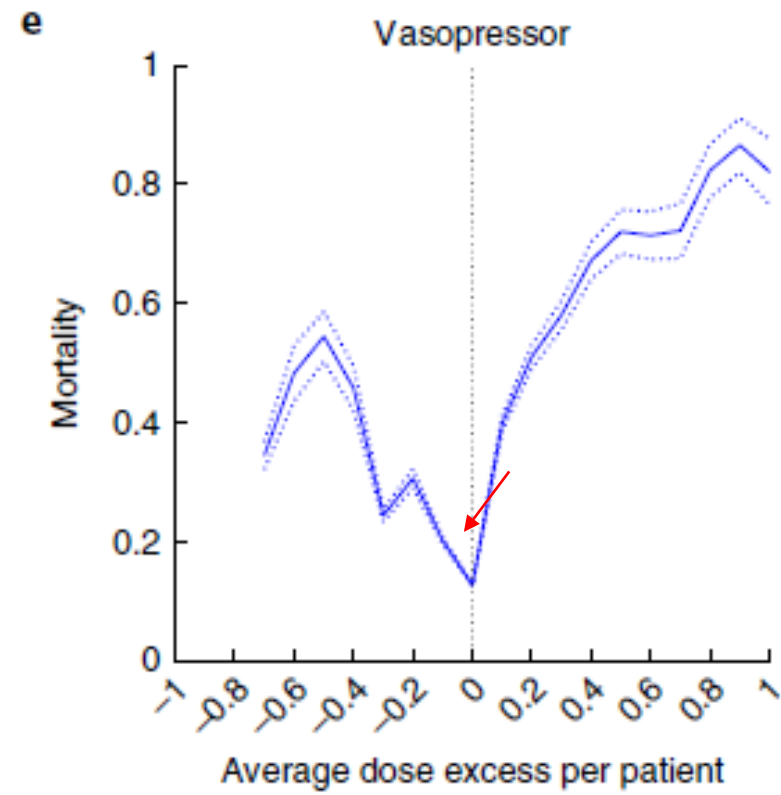
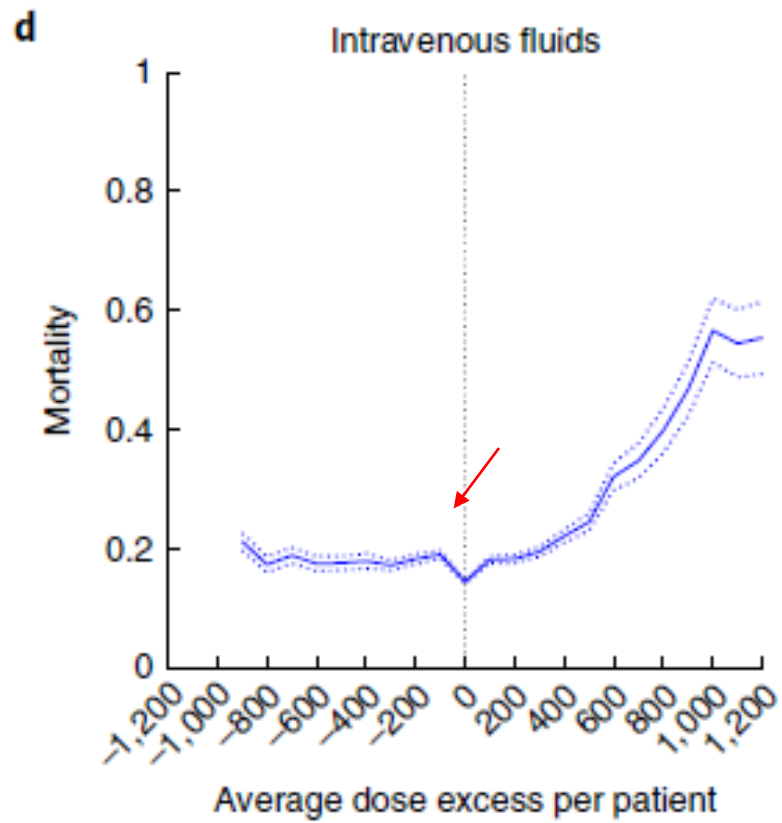
## The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care

Matthieu Komorowski<sup>1,2,3</sup>, Leo A. Celi<sup>3,4</sup>, Omar Badawi<sup>3,5,6</sup>, Anthony C. Gordon<sup>1\*</sup> and A. Aldo Faisal<sup>2,7,8,9\*</sup>

NATURE MEDICINE | VOL 24 | NOVEMBER 2018 | 1716-1720 |

- Imperial College London
- Pacienti so sepsou, ICU
  - tréning: MIMIC-III (17 000, 5 ICU, jedna nemocnica)
  - validovanie: eICU (79 000, 128 nemocníc)
- 48 premenných; 90-dňová mortalita
- **Data o liečbe – tekutiny, vazopresory** (dávka a 4 hod)
- ML - Reinforcement learning
  - učí sa metódou pokus/chyba, hľadá pravidlá pre optimálny výsledok
- Markov decision process
- „Dolovali“ sa tie postupy lekárov, ktoré viedli k najlepším výsledkom
- Porovnanie: lekári vz. algoritmus





Porovnanie AIC a lekára v eRI

## 4. Data mining

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

# Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis

Christopher W. Seymour, MD, MSc; Jason N. Kennedy, MS; Shu Wang, MS; Chung-Chou H. Chang, PhD; Corrine F. Elliott, MS; Zhongying Xu, MS; Scott Berry, PhD; Gilles Clermont, MD, MSc; Gregory Cooper, MD, PhD; Hernando Gomez, MD, MPH; David T. Huang, MD, MPH; John A. Kellum, MD, FACP, MCCM; Qi Mi, PhD; Steven M. Opal, MD; Victor Talisa, MS; Tom van der Poll, MD, PhD; Shyam Visweswaran, MD, PhD; Yoram Vodovotz, PhD; Jeremy C. Weiss, MD, PhD; Donald M. Yealy, MD, FACEP; Sachin Yende, MD, MS; Derek C. Angus, MD, MPH

20 118 pacientov so sepsou, „dolovanie“ podobných podskupín

JAMA. 2019;321(20):2003-2017.



# Data mining - fenotypy sepsy

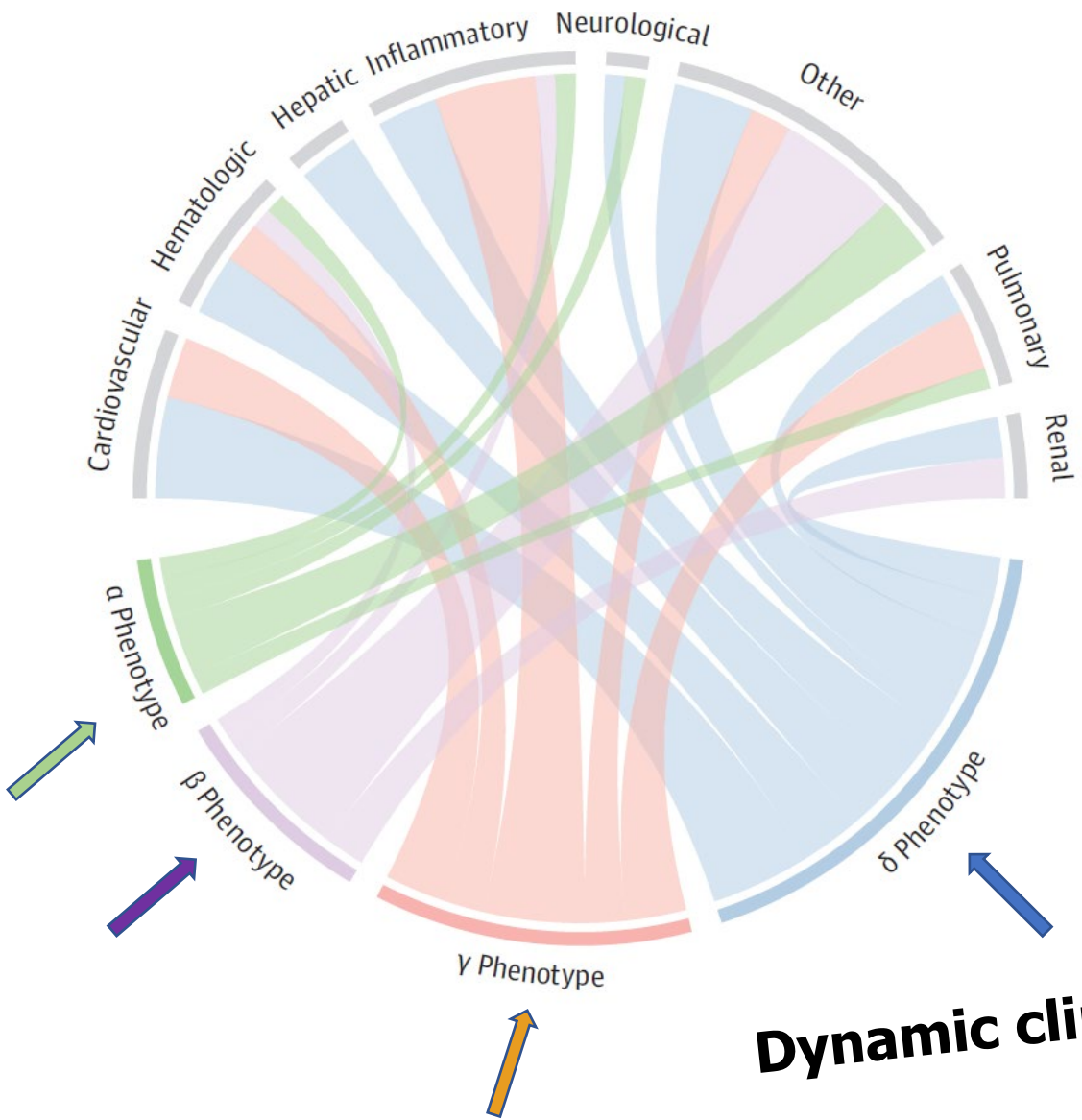


- Alfa (24 - 42 %): menej abnormalít
- Beta (19 - 30 %): starší, komorbidity, obličky
- Gama (23 - 50 %): zápalové markery, albumín, teplota
- Delta (5 - 23 %): laktát, transaminázy, hypotenzia, pečeň
  
- Fenotypy mutidimenzionálne, nezodpovedajú tradičnému deleniu (miesto, závažnosť, postihnutie orgánov)

# Data mining - fenotypy sepsy

- Alfa (24 - 42 %): menej abnormalít
  - Beta (19 - 30 %): starší kom
  - Gama (23 - 5%) **Jedinečné klinické podskupiny** markery, albumín, teplota
  - Delta (5 - 23 %): laktát, transaminázy, hypotenzia, pečeň
- 
- Fenotypy mutidimenzionálne, nezodpovedajú tradičnému deleniu (miesto, závažnosť, postihnutie orgánov)

**A** All phenotypes combined



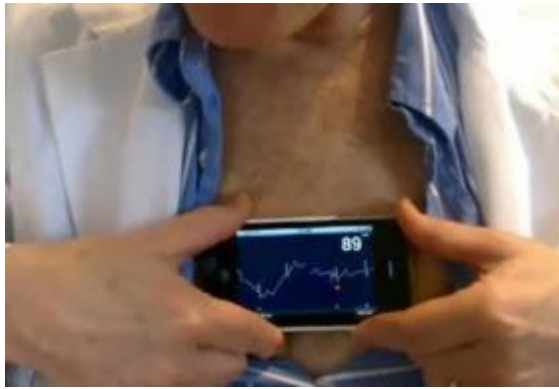
Mortalita:

- Alfa 2 %
- Beta 5 %
- Gama 15 %
- Delta 32 %

**Dynamic clinical data mining (DCDM) - podskupiny**

# 5. Wearable

# 2014



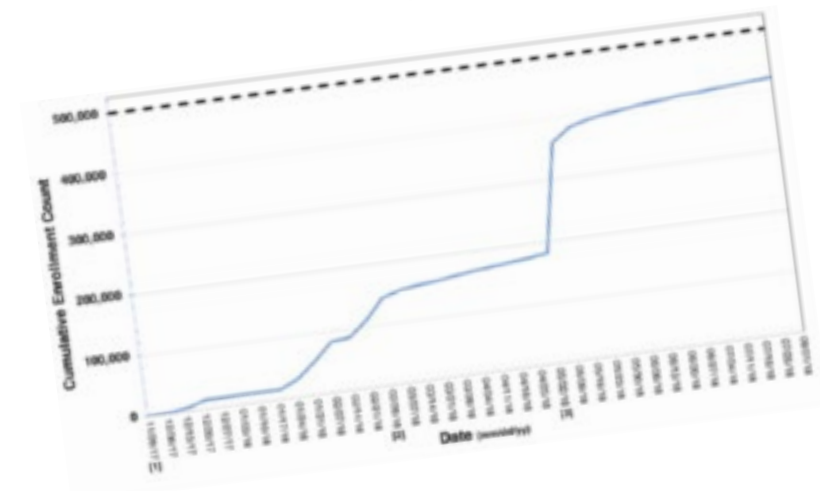
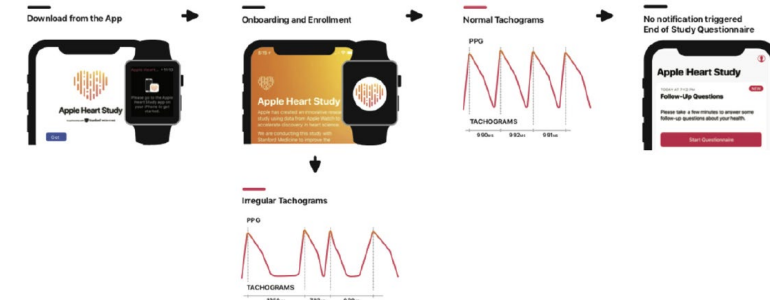
# The Apple Heart Study 16.3.2019

**Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study**

Mintu P. Turakhia, MD, MAS, <sup>a,b</sup> Manisha Desai, PhD, <sup>c</sup> Haley Hedlin, PhD, <sup>c</sup> Amol Rajmane, MD, MBA, <sup>d</sup> Nisha Talati, MBA, <sup>d</sup> Todd Ferris, MD, MS, <sup>e</sup> Sumbul Desai, MD, <sup>f</sup> Divya Nag, <sup>f</sup> Mithun Patel, MD, <sup>f</sup> Peter Kowey, MD, <sup>g</sup> John S. Rumsfeld, MD, PhD, <sup>h</sup> Andrea M. Russo, MD, <sup>i</sup> Mellanie True Hills, BS, <sup>j</sup> Christopher B. Granger, MD, <sup>k</sup> Kenneth W. Mahaffey, MD, <sup>l</sup> and Marco V. Perez, MD <sup>1</sup> *Stanford, Palo Alto, Cupertino, CA; Philadelphia PA; Denver Colorado; Camden NJ; Decatur TX; Durham NC*



- 68. kongres ACC, New Orleans, Louisiana
- Stanford University + Apple
- Prezentácia v otvárací deň
- Apple Watch (series 1-3), pletyzmografia
- **400 000 zaradených, 25 000 > 65 r**
- **0,5 % odhalená AF**
- Potenciál virtuálnych štúdií – rýchle zaradenie obrovského počtu pacientov (big data)
- Diskusia ...



# Problémy s dátami



1. Zber, uchovanie, spracovanie, výber dát, klinická otázka/kontext, model, analýza, vizualizácia, **interpretácia**, korelácia vz. kauzalita
2. **Heterogenita** - štandardizácia, harmonizácia
3. **Správnosť**, kompletnosť
4. Prenositel'nosť, **extrapolácia**
5. Kultúrne rozdiely, organizačné rozdiely
6. Ochrana súkromia, etické a **legálne** problémy; GDPR
7. **Anonymizácia/explicitný** súhlas, **dobrovoľné** zdieľanie
8. Komercializácia údajov vz. **otvorený** prístup

# Budúcnosť

- Big data – ropa 21. storočia
- Big data (real time data) + UI = predikcia, **prevencia** ...
- Learning health care system (**LHS**)
- Kvalita a **bezpečnosť**
- Riešiť **problémy** s dátami
- **Vzdelávanie** v UI má začať v ranom veku na školách, až po študentov LF a lekárov
- Clinical data scientist
- Máme nielen big data, ale aj **vlastné** medicínske vedomosti; integrovať
- Presnosť nástrojov AI sa musí **validovať** s ohľadom na výsledok a klinickú užitočnosť



# J.G.Mendel (1822-1884) a big data



GREGOR MENDEL

semeno		květ	lusk		stonek	
tvar	dělohy	barva	tvar	barva	umístění	velik
šedý & kulatý	žluté	bílá	plný	žlutý	lusk a květy podél stonku	dlouhý
bílý & svrasklý	zelené	fialová	příškrčený	zelený	koncové lusky, vrcholový květ	krátký
1	2	3	4	5	6	7

Tab. 2 Počet rostlin v pokusech s hrachem

Rok/ Pokus	Čisté linie	Křížení rostlin lišících se v			Zpětné křížení	Celkem
		jednom	dvou	třech		
		párech znaků				
1854	850					850
1855	850					850
1856	550	25				575
1857	550	531				1081
1858	550	1309	30	30		1919
1859	550	2413	15	24		3002
1860	550	4739	529	639		6457
1861	550	3240	80	4730		8600
1862	550	1360	80	160	100	2250
1863	550	300	80	160	551	1641
<b>Celkem</b>	6100	<b>13917</b>	814	5743	651	<b>27225</b>



# Budúcnosť



- Big data – ropa 21. storočia
- Big data (real time data) + UI = predikcia, **prevencia** ...
- Learning health care system (**LHS**)
- Kvalita a **bezpečnosť**
- Riešiť **problémy** s dátami
- **Vzdelávanie** v UI má začať v ranom veku na školách, až po študentov LF a lekárov
- Clinical data scientist
- Máme nielen big data, ale aj **vlastné** medicínske vedomosti; integrovať
- Presnosť nástrojov AI sa musí **validovať** s ohľadom na výsledok a klinickú užitočnosť
- **Nebojme** sa technológií, osvojme si ich

# Ďakujem za pozornosť



[strenkler@gmail.com](mailto:strenkler@gmail.com)