R at the Ministry

Application and examples

Piotr Nowosielski
p.nowosielski@mz.gov.pl

Ministry of Health Republic of Poland

28/09/2019

About project & team

- The recent **influx of funds** (both of EU and local origin) in the public healthcare system necessitates their reasonable allocation.
- Our team is a part of EU-financed project which aims **to map health needs of Polish people** (i.e. to focus the decision makers on areas where it is needed the most).
- We have assembled a team of **ca. 30 data enthusiasts**, mainly of economic or STEM background.
- Over **50 scientific articles** have been published in Polish and international journals, combining Big Data with the medical research.

Data & Tools

- Our goal is to obtain a profound overview of the system so as **to enchance the evidence-based decision-making process**.
- Data we analyze:
  - **National Health Fund** (NFZ) registries (10-year span);
  - **Social Insurance Institution** (ZUS) registries;
  - **Ministry of Finance** registries;
  - Disease-specific registries, e.g. **National Cancer Registry** (KRN), **National Cardiac Surgeries Registry** (KROK).
- Tools we use:
  - Mainly `R`, Python;
  - We have created our own object-oriented `R` packages to **facilitate reporting, visualizing and database connection**.

What is a stroke?

A **stroke** is a medical condition in which poor blood flow to the brain results in cell death. If the patient does not undergo the treatment within several hours since the stroke, there's a **very low probability of survival** without any severe impairment.

There are two main types of stroke:

- **ischemic** due to lack of blood flow,
- and **hemorrhagic** due to bleeding.

What are the risk factors for 24-hour stroke mortality?

$$\log\left(\frac{Z_i}{1-Z_i}\right) = \beta_0 + \beta_1 age_i + \beta_2 sex_i + \beta_3 strokeward_i + \beta_4 distance_i$$

Why are the stroke ward and the distance important?

term	intercept	age	sex	stroke_ward	distance
beta	-6.212	0.056	-0.092	-2.002	0.029
std.error	0.262	0.001	0.026	0.025	0.006
statistic	-23.717	45.150	-3.491	-79.435	4.952
odds.ratio	0.002	1.057	0.913	0.135	1.030
or_conf_l	0.001	1.055	0.867	0.129	1.018
or_conf_u	0.003	1.060	0.961	0.142	1.042

We can infer from the model that:

- the increase in the distance to the hospital statistically significantly **enhances** the mortality risk;
- the undergone treatment in a stroke ward statistically significantly **decreases** the mortality risk.

Model assumptions

The decision function is linear:

$$f(x) = \sum_{i=j}^{N} c_j \cdot x_j$$

Problem constraints are as follows:

$$\sum_{j=1}^{N} a_{ij}x_j \leq b_i, i = 1, \ \ldots, \ m_{\mathrm{le}}$$

$$\sum_{j=1}^{N} a_{ij}x_j = b_i, i = 1, \ \ldots, \ m_{\mathrm{eq}}$$

$$\sum_{j=1}^{N} a_{ij}x_j \geq b_i, i = 1, \ \ldots, \ m_{\mathrm{ge}}$$

$$x_j \geq 0, j = 1, \ \ldots, \ N$$

where m denotes the number of given conditions.

Model variables

Decision variable

- $Build_d$ – a binary variable: $1 =$ a stroke ward is to be built and 0 otherwise;
- $Depend_{c,d}$ – a binary variable: $1 =$ all c common (Polish *gmina*) patients are to be referred into the d common and 0 otherwise.

Slack parameters

- Pop_c – Number of the c common patients;
- $Dist_{c,d}$ – the distance between the c common centroid and the d common centroid;
- Cap – the service capacity of a stroke ward;
- $WardsN$ – number of stroke wards.

Model

- Decision function:

$$\min \sum_{c=1}^{Ncom} \sum_{d=1}^{Ncom} Pop_c \cdot Dist_{c,d} \cdot Depend_{c,d}$$

- Problem constraints:

1. $\sum_{c=1}^{Ncom} Build_c = WardsN$
2. $\forall c \in Ncom \sum_{c=1}^{Ncom} Depend_{c,d} = 1$
3. $\forall c, d \in Ncom\ Depend_{c,d} \leq Build_d$
4. $\forall c \in Ncom \sum_{d=1}^{Ncom} Pop_c \cdot Depend_{c,d} \leq Cap \cdot Build_d$
5. $\forall c \in Ncom\ Build_c \in \{0, 1\}$
6. $\forall c \in Ncom\ Depend_{c,d} \in \{0, 1\}$

Results

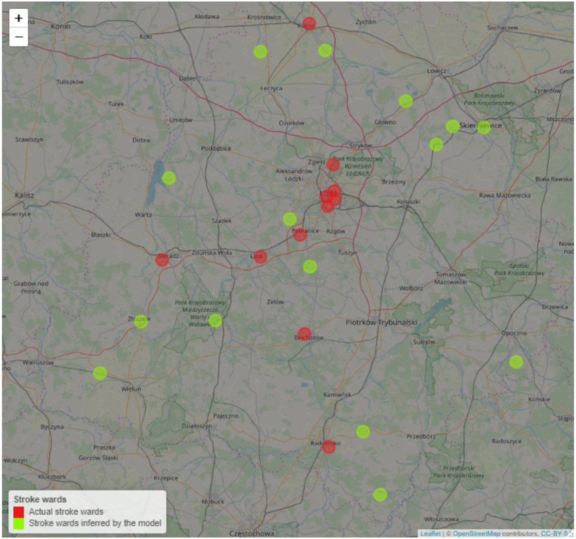


Figure 1: The distribution of stroke wards in the Lodz voivodeship

Nitty-gritty of heart failure (HF) disease

- Heart failure is a medical condition that occurs when your heart is **unable to pump blood well enough** for your body's needs to be fulfilled.
- In a majority of cases HF stems from other diseases (e.g. cardiovascular diseases, diabetes) **having been left untreated**.
- In view of population aging **people live longer**, and thus **develop HF more often**. Due to obscure etiological nature of HF there is hardly any systemic approach which results in huge, albeit **avoidable costs incurred by National Health Fund**.

Care pathways

- Care pathways can be described as **patient's progression through the healthcare system**.
- It is pivotal to keep track of individuals' care pathways in order **to identify the blind spots of the system**.
- Care pathways can be aggregated by means of a discrete-time stochastic process in order to provide a holistic view of a treatment regimen for a specific disease.
- In an ideal state of affairs people with HF would be treated mainly in **primary care** (PC) and **ambulatory care** (AC), whereas at present the bulk of those people **end up being hospitalized**.

Preliminaries

- Let S denote the set of states $\{s_1, s_2, ..., s_t\}$. Let p_{ij} denote the transition probability from state s_i to state s_j.
- Current step (i.e. visit in the healthcare system) is **conditionally dependent only on the previous one**, i.e. the process is memoryless and thus satisfies the Markov property:

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, ..., X_0 = x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1})$$

- Discrete Markov process consists of a starting vector **u** (an aggregate of initial states of patients' care pathways) and transition probability matrix **P**. The prediction after n steps can be computed as:

$$\mathbf{u}^{(\mathrm{n})} = \mathbf{u}\mathbf{P}^{\mathrm{n}}$$

.

- Transition probability matrix can be derived from individuals' sequences through **maximum likelihood estimation**.

Discrete-time Markov process for HF

- In order to ensure clarity suppose that the set of states $S = \{PC, AC, Hosp, EMS\}$ is exhaustive.
  (PC – primary care, AC – ambulatory care, Hosp – hospitalization, EMS – emergency medical services)

- The structure of the process is as follows:

$$\mathbf{u}^{\mathrm{T}} = \begin{array}{c} \mathrm{PC} \\ \mathrm{AC} \\ \mathrm{Hosp} \\ \mathrm{EMS} \end{array} \begin{pmatrix} 0.50 \\ 0.13 \\ 0.35 \\ 0.02 \end{pmatrix} \qquad \mathbf{P} = \begin{array}{c} \\ \mathrm{PC} \\ \mathrm{AC} \\ \mathrm{Hosp} \\ \mathrm{EMS} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{PC} & \mathrm{AC} & \mathrm{Hosp} & \mathrm{EMS} \end{array} \\ \begin{pmatrix} 0.86 & 0.04 & 0.09 & 0.01 \\ 0.25 & 0.66 & 0.08 & 0.01 \\ 0.53 & 0.08 & 0.3 & 0.09 \\ 0.41 & 0.04 & 0.46 & 0.09 \end{pmatrix} \end{array}$$

- The probability of ending up in a given state can be plotted as a function of the number of steps. The model converges to the stationary distribution π satisfying following equation: $\pi = \pi\mathbf{P}$.

Complete directed graph

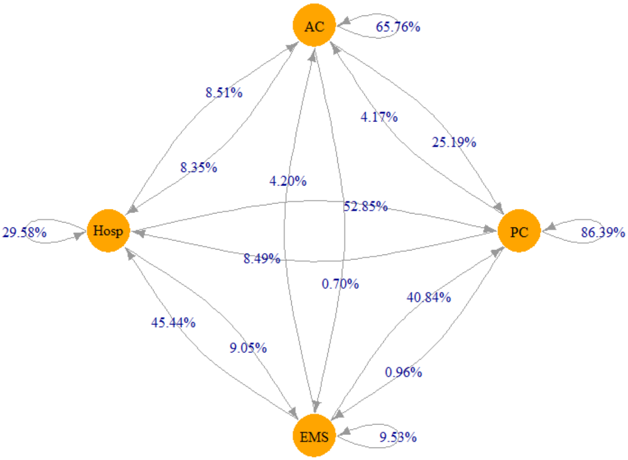
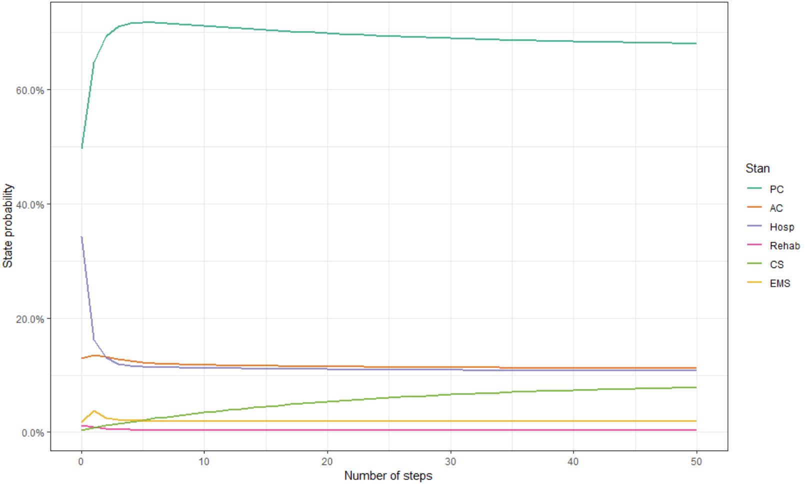


Figure 2: Care pathways of patients with heart failure (I50)

Prediction from the model



Summary

- Thank you for your attention
- Have a nice paRty