H1N1 – SCANDAL?

…proving former suspicions - based on an artificial-intelligence- and a big-data-driven approach…

Author: Robot Journalist (MY-X team, 2020.IV.01.)

Abstract: The vaccination should have preventive impacts in general. This prevention can be re-formulated as follows: the more is the ratio of the vaccinated persons the less should be the death rate. An explorative or AI-based model should therefore have parameters describing this relationship exactly in case of each risk groups. If the models will not reflect this expectation, then the preventive impact is not valid from statistical point of view. It does not mean that the vaccine as such is not effective. The production function between causes and consequences let derive an evaluation for each country. This evaluation means where is worth living – where is the life safer compared to other EU-countries. The model can also support simulations. The simulations deliver answers for the following questions: How would be the death rate in case of any changes of the independent variables? The developed results let declare following statements, presumptions: the preventive impact could not be validated! 9 of 27 countries have a relatively low safety: GR<EE<LV<FI<DK<LT<CZ<LU<SI

Keywords: corona virus + prevention + suspicion generating + robot analyst + automation

# Introduction

The following subchapters clarify: Why it is important to write about the topic for the society and for the author? What are the targeted groups? What are the objectives? What kind of utility can be expected after reading the article?

## Basic information

“Influenza A virus subtype H1N1 (A/H1N1) is the subtype of influenza A virus that was the most common cause of human influenza (flu) in 2009” – according to WIKIPEDIA (end of March 2020). In case of almost each governmental decision, there are suspicions. Suspicions will be handle based on the legal ways. Legal ways use the magic of words. Here and now, suspicion generating will be demonstrated based on a big-data- and artificial intelligence-driven approach. This approach can be automated. Therefore, this article can be seen as a product of a robot journalist.

## Objectives

The goals of the article are simple. To identify structured data about H1N1. To model death rate based on nine independent variables: like vaccination rates of six different risk groups, population density and urbanisation rate, and infection rate. To derive impacts of the independent variables to the consequence variable. To compare impacts to the expectations. To estimate death rate for each involved country of the EU-28 where Croatia should be excluded because of lack of information. To evaluate estimations and fact in order to see which country is safer.

Further goals are to have steps of data collecting, modelling, analysing in a reproducible form. Besides, to conclude in an automated way based on a prepared interpretation rule set without any human influence after involving the real data where conclusions were searched for existing prevention impacts and the difference of estimated and measured death rates of the countries. Parallel, to use solver-based online analytical tools for exploring relationships between the death rate as dependent variable and each independent one.

## Targeted groups

Investigative journalists work on suspicion generations in a professional way. The suspicion generation should also be driven by legal experts. The robotizing of this activity is important of the entire society. The robotizing brings more efficiency and more objectivity. A good presumption is trivial or self-evident. The proofs and the whole logical can immediately be interpreted by the Readers. The suspicion generating process must be reproducible. The logic of an evidence-based suspicion generating process must be existing before the needed data are available.

## Expected utilities

The most trivial expected benefit behind the automation is the increasing the efficiency. Efficiency means what kind of impact can be expected concerning a given volume of resources. The articles about possibilities of robotizing supports shifting paradigms especially in case of relevant phenomena like the corona pandemic process or even the predecessor – the H1N1 process. The credibility of the global process can not be evaluated based on personalized experiences – only on statistics and system theoretical approaches. The same is valid for the global warming.

## Personal motivations

The article tries to demonstrate in case of a closed pandemic process what kind of controls should always be executed. The consequences of the H1N1 database should be transferable to the pandemic process caused through corona virus infections. The Industry 4.0 strategy it means the robotizing is valid for each segment of the human life. Valid therefore for the suspicion generation. Valid for the investigative journalism. Valid for the teaching and learning strategies as such.

# Literature

It is enough in order the have a kind of continuity if we can identify one single object in the media containing suspicion generating items: <https://abcnews.go.com/Health/SwineFlu/swine-flu-pandemic-world-health-organization-scientists-linked/story?id=10829940>

The source and the publisher (c.f. <https://en.wikipedia.org/wiki/ABC_News>) do not any matter if a publication can be seen as reproducible.

If the source and publisher should be evaluated, then it should be done based on the big-data. Google Trends is a good data asset: <https://trends.google.com/trends/explore?date=all&q=%2Fm%2F01v3sw,%2Fm%2F02k8p8,%2Fm%2F0ps53,%2Fm%2F07k2d,%2Fm%2F01hddb>

Figure Nr.1 shows that ABC can be seen as an average-like objects with increasing trends concerning the curiosity of the potential users.

It is important, that the article has an author and a timestamp. It is also to highlight that the article works with citations. In the magic of words, citations can be seen as data. Real data-driven parts are not given – therefore, the reproducibility can not be automated. On the other hand, the magic of words has an other consistent frame: the logic – as the legal processes tries to use it…

Validation seems to be given based on references to other organisations.

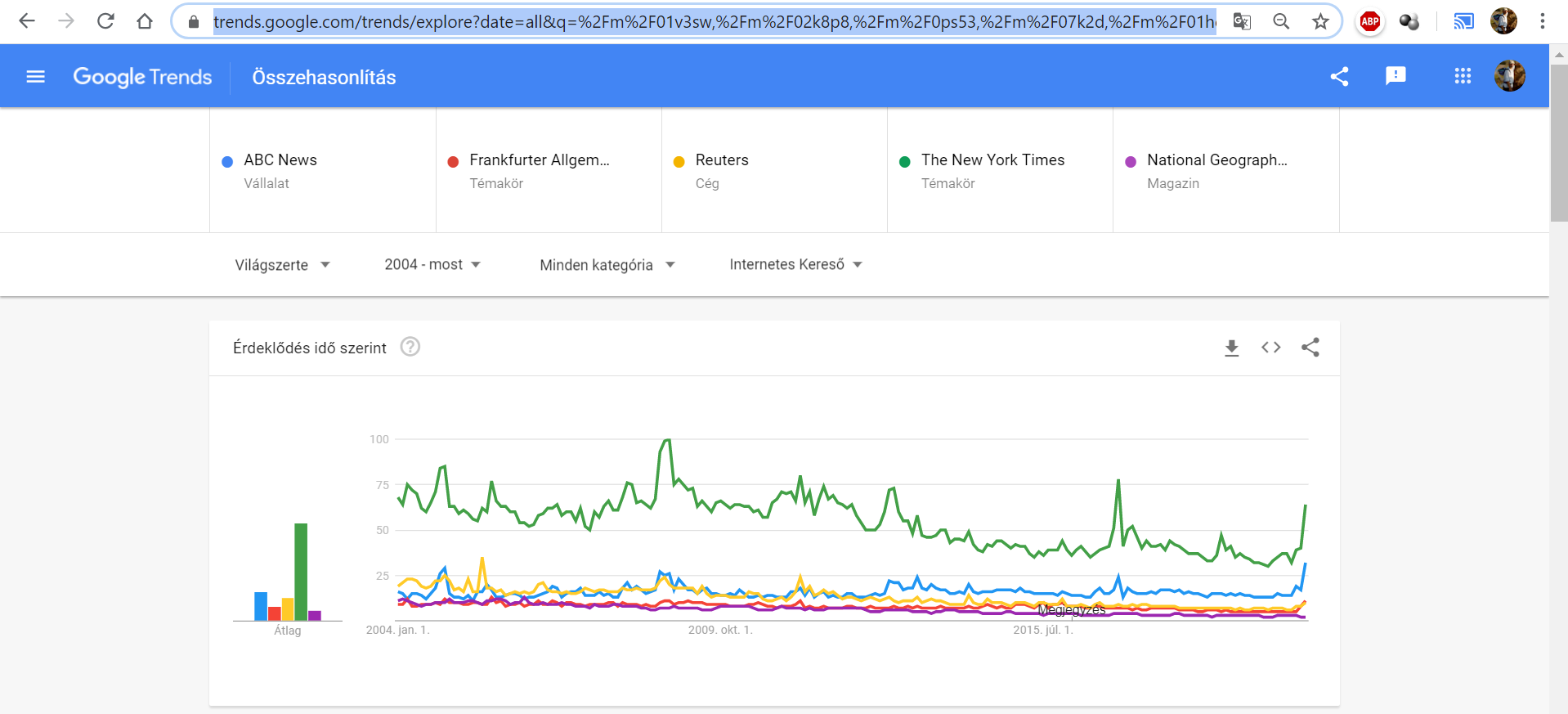


Figure Nr.1 – News agencies and newspapers compared to each other (source: Google Trends – where the horizontal axis is the time and the vertical axis is a relativized index being capable of demonstrating a kind of worldwide curiosity concerning the other objects)

The article does not use the keyword “prevention/preventive”. The other keyword “vaccine\*” is used 13 times. This article does not speak about rationality or irrationality of the vaccination as such. This article does speak about potential conflict of interests. However, good advices can even come from conflicted parties. If the vaccination does not have the expected impact, then advices from independent experts are inacceptable (c.f. the wise rabbi and the dying geese – see Figure Nr.2a/2b):



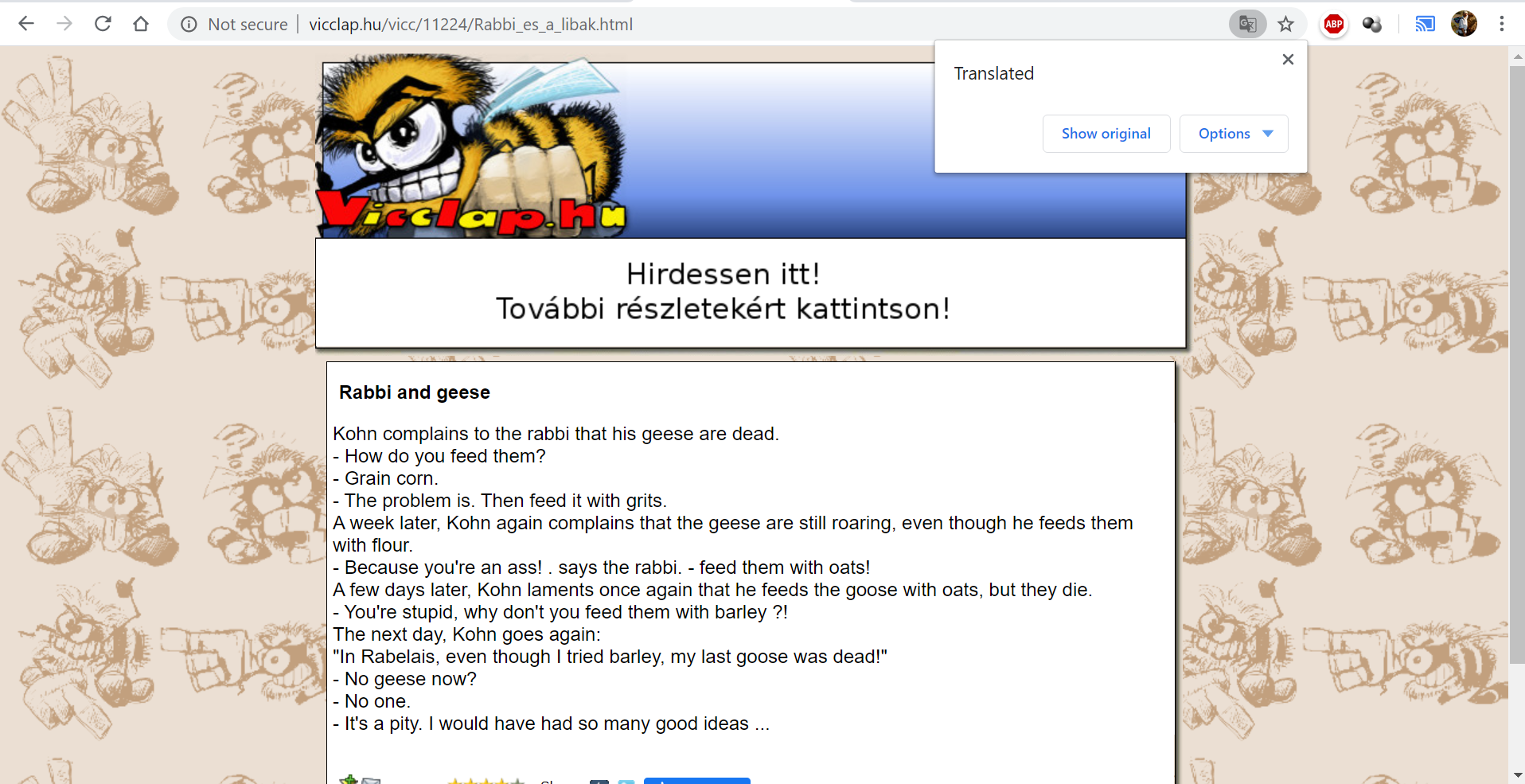


Figure Nr.2: It is not enough to have ideas – the forecasts should become reality (source: own presentation)

Figure Nr.3 demonstrates in the its first paragraph the first presumption…



Figure Nr.3 – Vaccination-related suspicions (source: ABC News)

This article of a robot journalist will present the reproducible way, how a big-data-oriented and artificial intelligence-related robot investigation can be processed by each Reader/Student/Teacher/Citizen…

# Data assets

Each relevant data management steps can be seen in the background XLS-file: <https://miau.my-x.hu/miau/quilt/2020/projekt_h1n1/OAM1_h1n1.xlsx>

Used sources = raw data (see XLS-file – sheet ‘db’ – range(H2:H14 – CTRL+K for entire URLs):

* Vaccination rate of 6 different risk groups: <https://www.ecdc.europa.eu/sites/portal/files/documents/influenza-vaccination-2007%E2%80%932008-to-2014%E2%80%932015.pdf#page=37>
* Number of deaths caused by H1N1 in case of countries: <https://de.statista.com/statistik/daten/studie/71422/umfrage/todesfaelle-durch-die-schweinegrippe-in-europa/>
* Number of infected persons in case of countries: <https://de.statista.com/statistik/daten/studie/28467/umfrage/influenza-h1n1-in-europa/>
* Population density: <https://de.statista.com/statistik/daten/studie/74693/umfrage/bevoelkerungsdichte-in-den-laendern-der-eu/>
* Grade of urbanisation:
* <https://de.statista.com/statistik/daten/studie/249029/umfrage/urbanisierung-in-den-eu-laendern/>
* Population: <https://ec.europa.eu/eurostat/databrowser/view/tps00001/CustomView_1/table?lang=en>

Calculated variables:

* Ratio of deaths (caused by H1N1) = Number of deaths / Population
* Ratio of infections (caused by H1N1) = Number of infected persons / Population

Preconceptions:

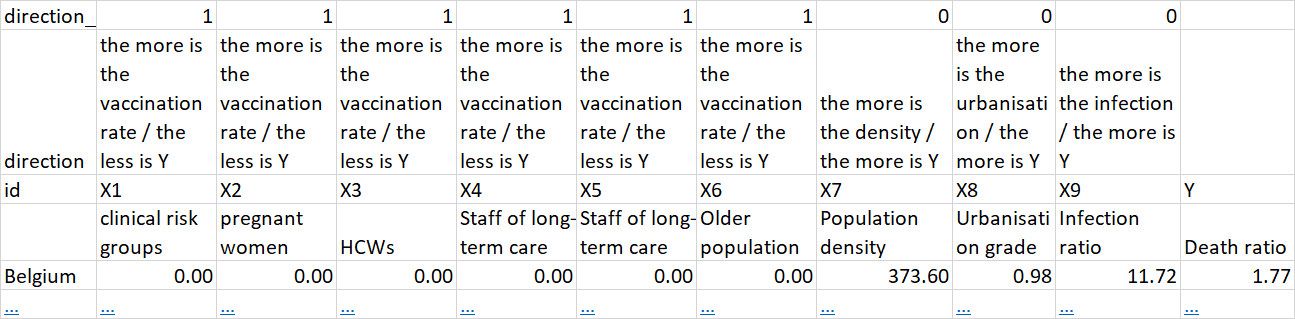


Figure Nr.4 – Expectation concerning relationships between Xi and Y (source: own presentation)

Figure Nr.4 shows the expected impacts. The preventive impact of the vaccinations can be seen as well-known (see: <https://en.wikipedia.org/wiki/Preventive_healthcare> - keyword „vaccin\*” - used 25 times).

Figure Nr.5 demonstrates how we can derive easy and fast a preconception about the relationship of the phenomena like density (population/urbanization) and infection/death.

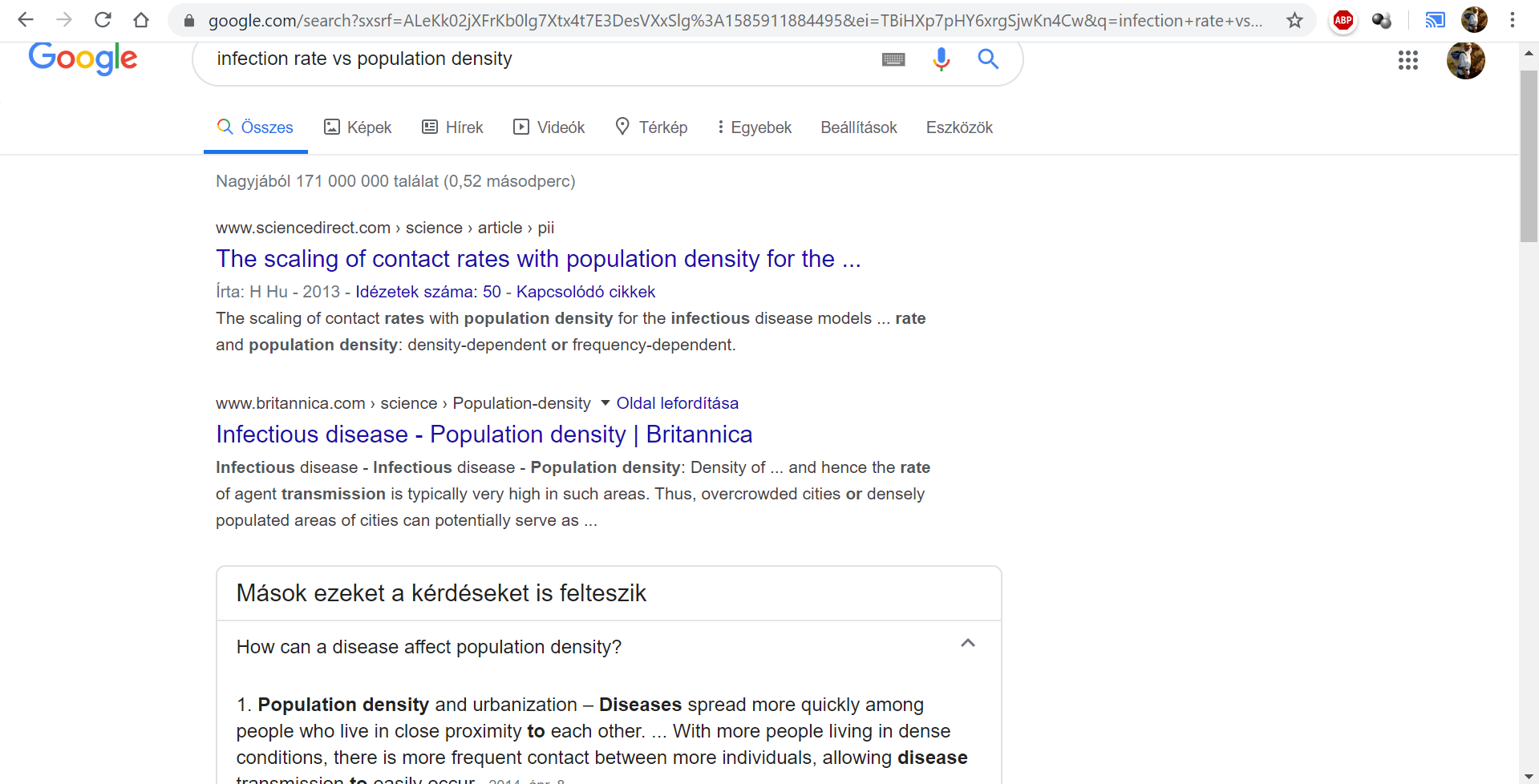


Figure Nr.5 – Naïve derivation of expected impacts (source: own presentation)

It is important to highlight: the Liebig-principle speaks about changing of impacts, about bubbles of effects (c.f. <https://miau.my-x.hu/miau/253/traffic-simulations.pdf>).

The used data sources are different concerning the processability of the data:

* The data about the vaccination rates could not be transformed into a database, it means each relevant value should be retyped in a manual way.
* The other sources could be converted into Excel-tables with more or less IT-knowledge.

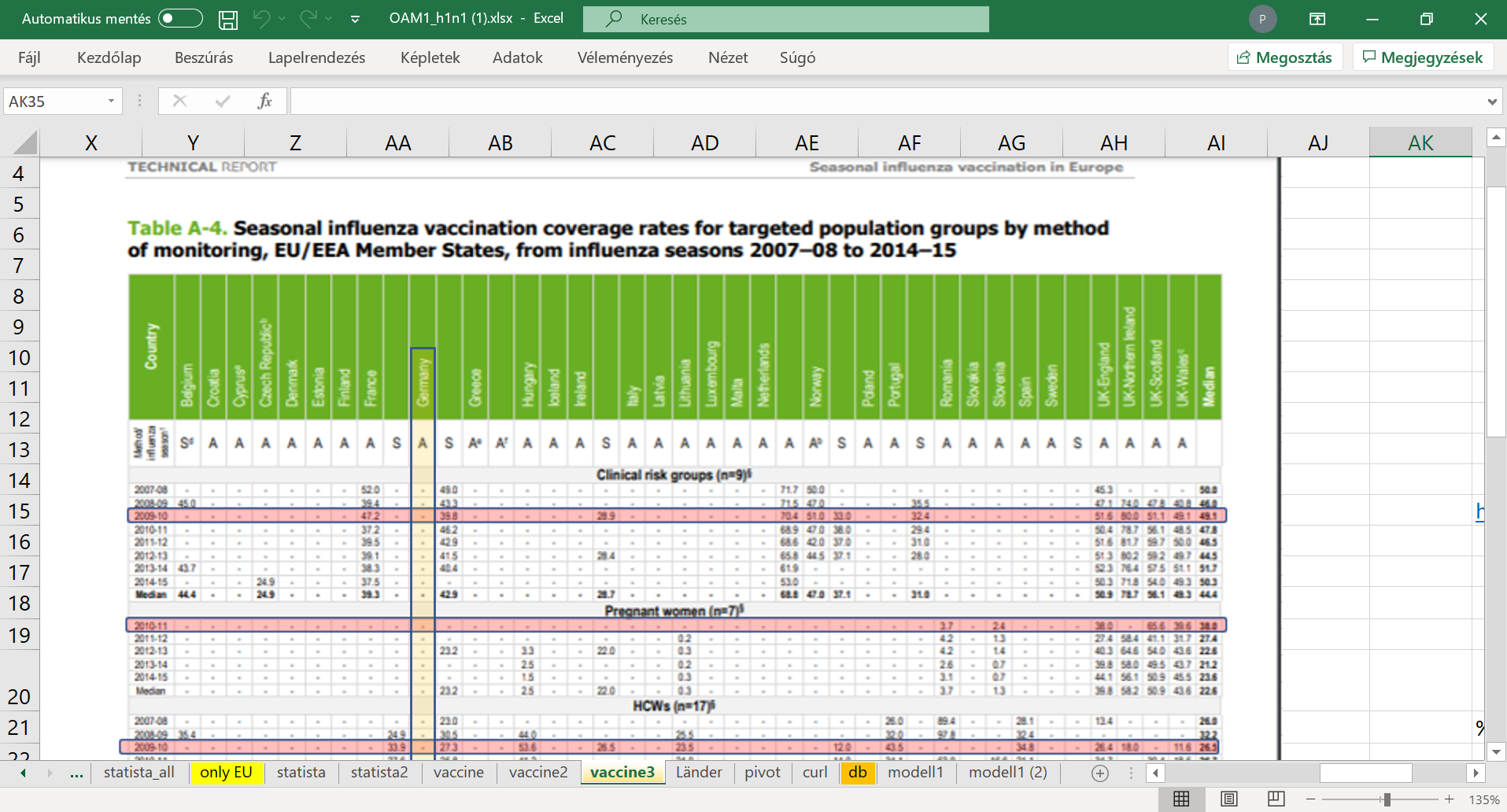


Figure Nr.6 – Filter-effects for retyping (source: own presentation)

Figure Nr.6 presents how it is possible to support the retyping in case of image-based sources.

More details about data processing (problems and tricks concerning the H1N1-databse): <https://miau.my-x.hu/miau/quilt/2020/quilt2/launching2020III25/part1a.html>

# Methodology

A general view can be downloaded here: <https://miau.my-x.hu/miau/196/My-X%20Team_A5%20fuzet_EN_jav.pdf>

The H1N1-related modelling is a simple/standard case in the series of the production functions. A production function is a more or less formula where based on the independent variables (vaccination rate of six risk groups, population density, grade of urbanisation, rate of infections) one single dependent variable (death rate) will be derived (with more or less fitting). Production functions can be derived in unlimited ways. The chosen online, solver-based engine (COCO-STD - <http://miau.my-x.hu/cocostd>) is a part of flexible frame system where the so-called learning process can be stop based on an optimization process in an automated way. The robot journalist needs flexibility and automation (<https://miau.my-x.hu/myx-free/coco/>).

The standard similarity analysis (COCO-STD) is very simple. The inputs are always ranking values of the independent variables (where the ranking is made based on the expected relationships between Xi and Y). The Y values, it means the values of the consequence or dependent variable have always a common unit (death rate pro 1000000 capita). The modelling process searches for optimal substitution values for each ranking value where the better/higher impact may not have lower substitution values.

In frame of one model it is also possible to involve the same variable set with direct (expected) and indirect (potential) directions.

The interpretation rules of the estimations from a COCO-STD optimization are existing independent from the real data about H1N1. These rules were born parallel to the modelling process as such before decades. These rules are simple and are basing on the system theory:

* In ideal case, the more robust relationships are the expected relationships.
* An object (country) can be evaluated as a better object, if the estimation value for negative phenomena (like death) is higher than the real value, because the negative impacts could have been more.
* The most relevant attributes of a particular object can be ranked based on the substitution values of the input ranking numbers of the given object.

# Results

The following visual effects can be seen as an extract of the modelling results:

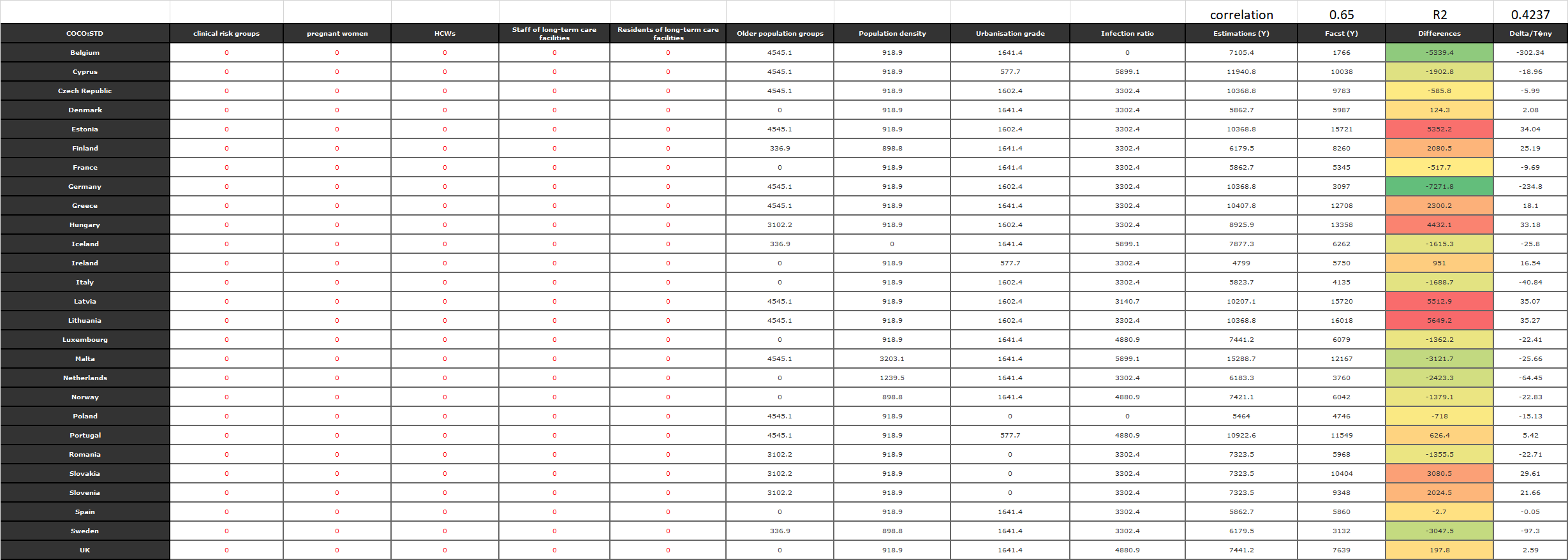


Figure Nr.7 – Model-parameters and estimations based on expected relationships (source: own presentation – where the red coloured numbers highlight the lack of the expected impacts of prevention)

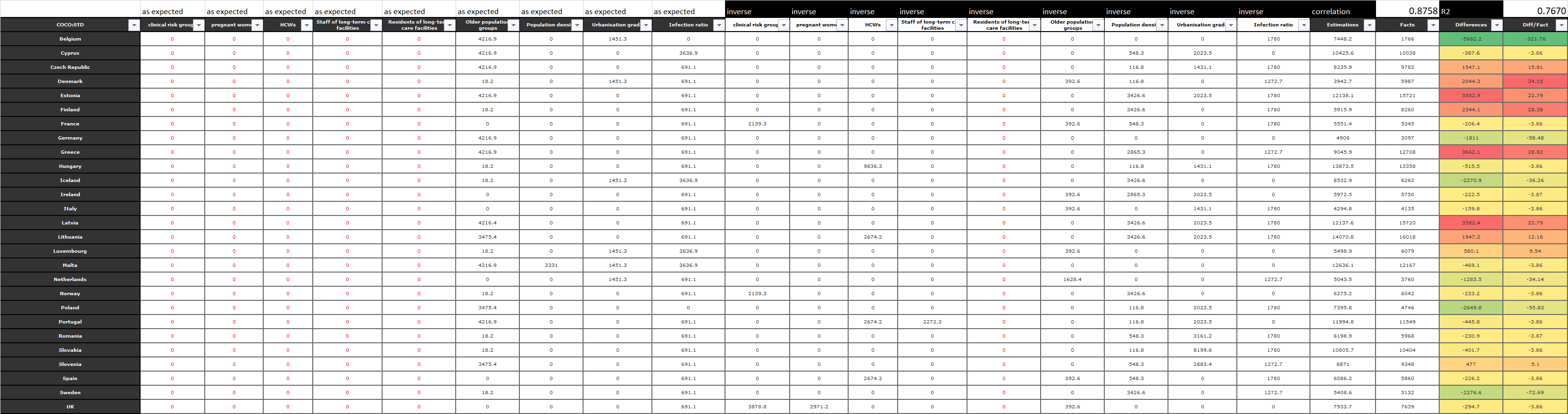


Figure Nr.8 – Model-parameters and estimations based on doubled relationships (source: own presentation – where the red coloured numbers highlight the lack of the expected impacts of prevention)

Figure Nr.7 demonstrates a relatively weak model (correlation = 0.65, R2=0.4237). The more flexible model (Figure Nr.8) with a doubled variable set has an acceptable high correlation of 0.8758 and an R2 of 0.767, where a correlation can be accepted close to 0.9 without any questions.

The objects, it means the countries has an estimation and a published death rate as Y and the differences of the estimations and the facts can be interpreted as absolute values and or after a standardization where the differences will be divided through the published death rate. The standardized view uses the published level for relativizing during the absolute differences do not take in account what should have been, if…

The best country is Belgium. The most weak countries are Greece according to the absolute differences and Denmark in case of the relativized differences. Germany seems to be on the good side and seemingly Hungary too.

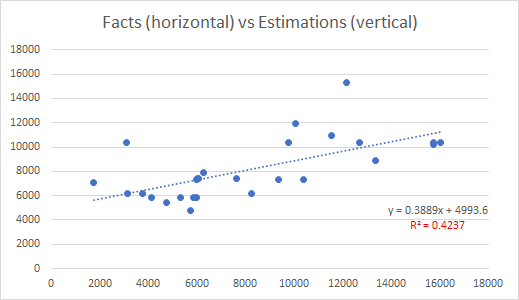
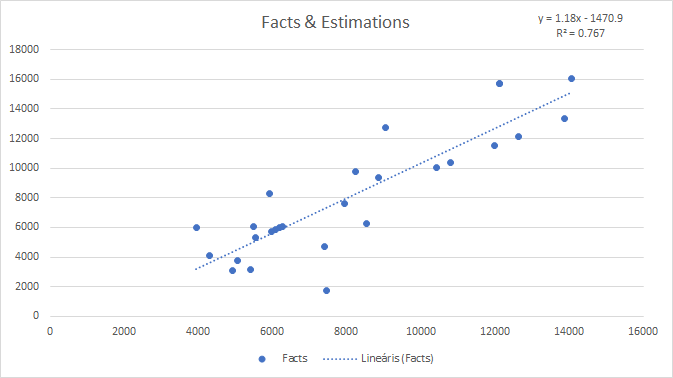
 

Figure Nr.9 – R2 values for the simple and the doubled models (source: own presentation – where the R2 values can be compared to the same values in case of Figure Nr.7 and Figure Nr.8).

The 2D-figures (see Figure Nr.9) make possible to check and to interpret the meaning and derivation of the correlation values.

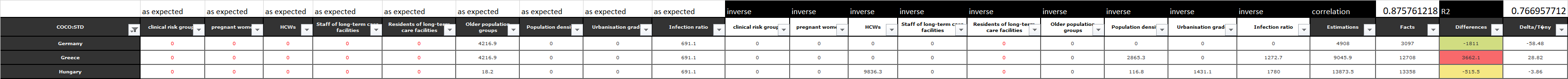


Figure Nr.10 – Impact mechanisms of three highlighted countries (source: own presentation)

Figure Nr.10 does highlight the above-mentioned 3 countries: DE, GR, HU in order to give more detailed interpretation about these typical objects.

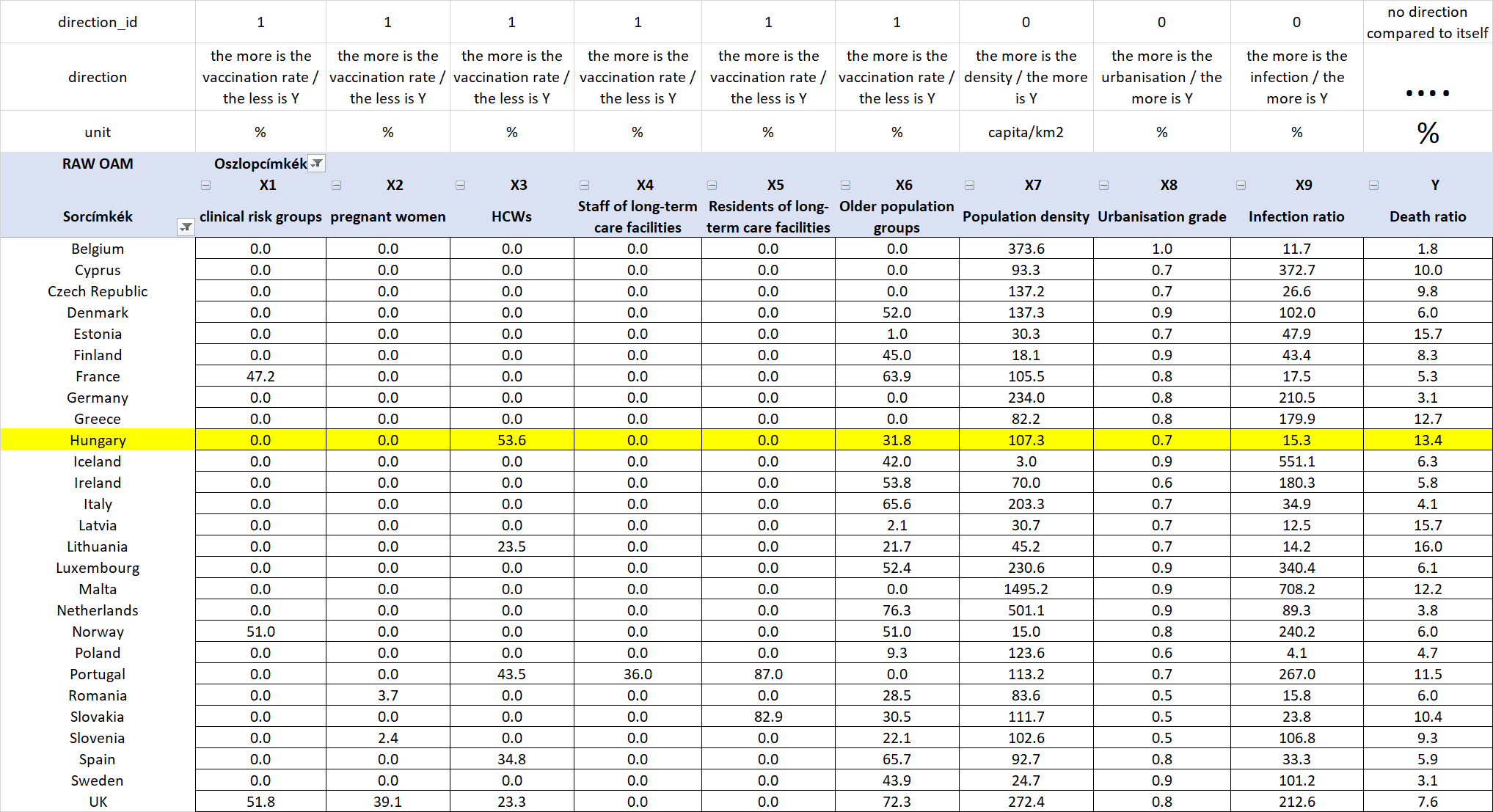


Figure Nr.11 – Case study of Hungary – Phase I – raw data (source: own presentation)

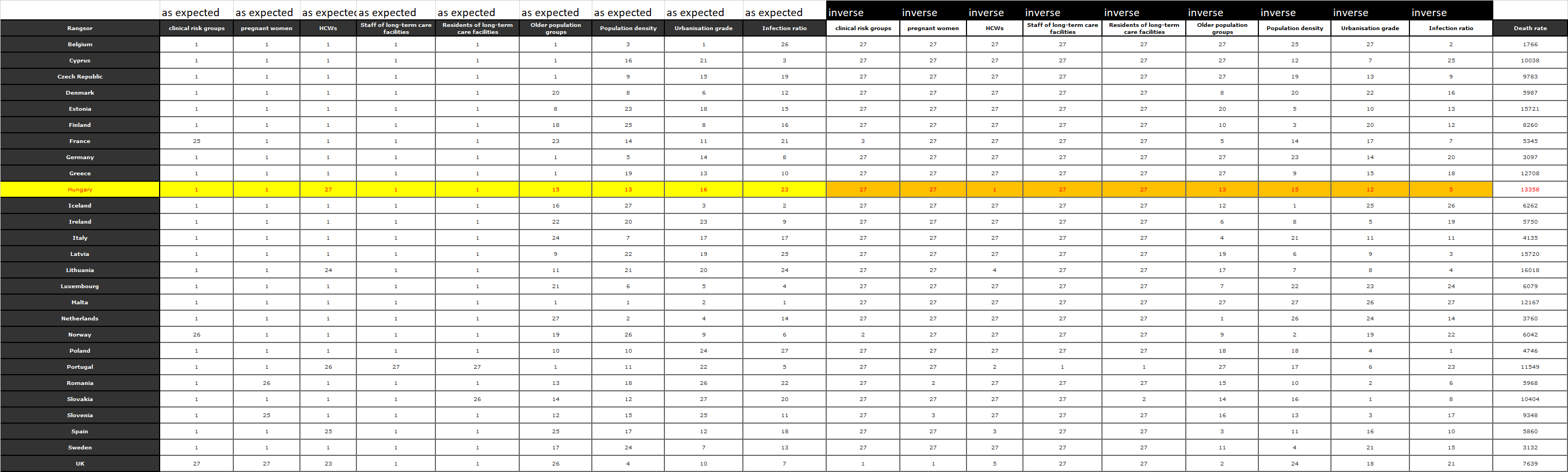


Figure Nr.12 – Case study of Hungary – Phase II – ranking values with doubled variables (source: own presentation)

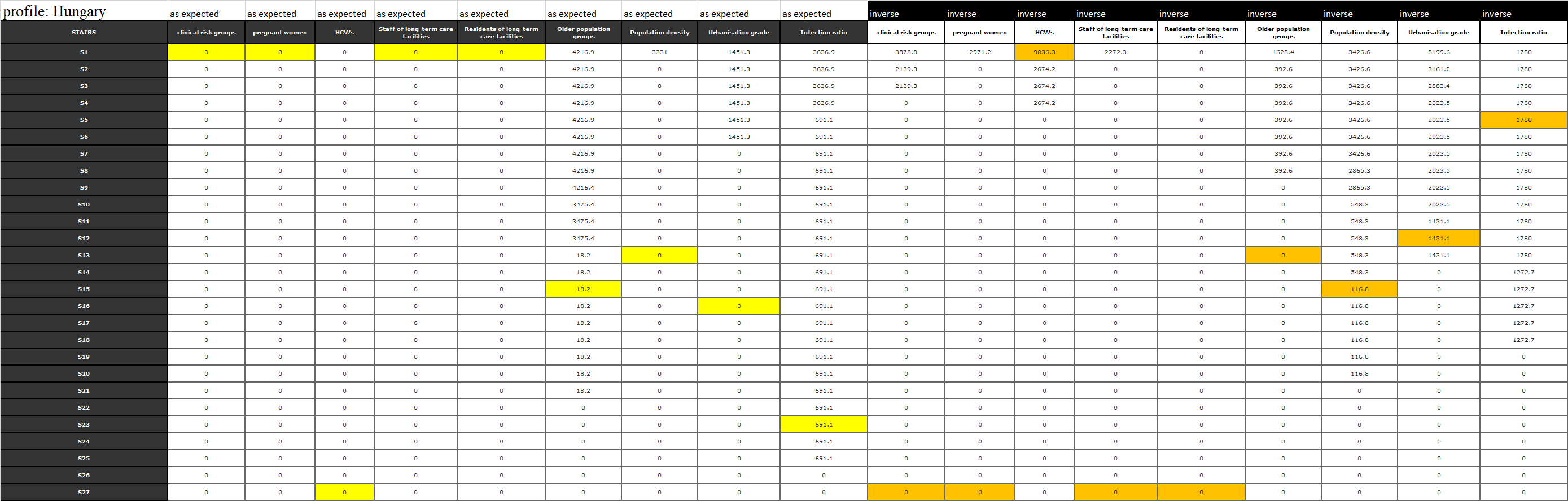


Figure Nr.13 – Case study of Hungary – Phase III – simulations (source: own presentation – where the yellow- and orange-highlighted cells are steps in a staircase function derived from Figure Nr. 12 – and the values of the highlighted cells are the same compared to the values for Hungary in Figure Nr.10).

# Discussion

The relatively low number of the objects – of the countries can be seen as a kind of risk concerning each result layer. On the other hand, the unexpected relationships can be realistic. As already declared: the statistical impact of the vaccination depends on a lot of other circumstances. It is therefore possible to vaccinate in a less appropriate time and or in case of not really affected groups, persons. The unexpected impacts concerning the vaccination could be derived through three different model types.

The preventive bubbles according to the Liebig principle can however be seen in case of particular countries. Therefore, it is rational in a parallel way, that the expected preventive impact could not be generalized, but the existing preventive impact can be realized in certain situations where the circumstances are ideal compared to each other. As everybody can see, the complexity of the problem does not make rational to speak about in a naïve way.

The Liebig bubbles are valid concerning the expected prevention impacts for countries where the substitution values of the expected variables are involved into the estimation process and parallel, the substitution values of the unexpected impacts are weaker than the substitution values of the expected variables. So, the similarity analyses and the doubled staircase functions make possible to identify Liebig-like effects in case of countries. A classic regression model where each variable should always have the calculated average impact is not capable of simulating the Liebig-law.

# Conclusions

The conclusions have to take in account both the derivations based on the big-data + the artificial intelligence-oriented engines and the critical aspects. In ideal case, the conclusions are operative commands: what should be done or what should have been done. In this chapter, it is also possible to speak about analogies like H1N1-situation vs. corona-virus-situation. Each analytical process should be automated. In case of appropriate data assets, the analytical steps should be used immediately. The analytical steps should include the interpretation rules.

Chapter – Conclusions: The results let confirm the suspicion that the vaccination as such could not be expected in an ideal way. The argumentation is very simple: the statistical evidence of the expected preventive impacts could not be explored. The existence of the Liebig-principle in a country-specific way makes the suspicion stronger. There will be further investigations necessary and therefore more data should still be completed about countries where the expected preventive vaccination impact could not be confirmed in order to explore what should have been done in a more ideal way.

The results make also possible to evaluate the activities of the countries. The simulation potential ensures that we can speak about robust or rational behaviours or confused activity patterns. Germany demonstrated a robust pattern. In case of Hungary, the situation is confused but there are no real possibilities to be better. The experiences of the so-called good countries can be used as a kind of best practice. The adaptability of these best practices can be checked in form of simulation where the environmental factors will also be changed (like population density).

Chapter – Conclusions: The presented steps and analytical logic can be used for similar cases. It means even in the case of the corona-virus-situation. The similar logic can be interpreted not only for the modelling of death rates. It is also possible the simulate other phenomena. The modelling will therefore deliver not only suspicions about the statistical existence of the preventive impacts. This type of modelling will deliver information about each relationship between X- and Y-variables. This modelling type only needs an OAM and the staircase functions based on online similarity analyses.

# References

…see in the text stream…

# Annexes

The collected data, their sources, each calculation, and the involved visualizations effects can be studied in the background XLS-file: <https://miau.my-x.hu/miau/quilt/2020/projekt_h1n1/OAM1_h1n1.xlsx>

The article has a kind of werkfilm where a lot of specific interpretations can be read concerning the H1N1-analyses. Therefore, parallel, the werkfilm can be seen as a kind of annex.

<https://miau.my-x.hu/miau/quilt/2020/quilt2/launching2020III25/part0.html>

The storyboard of the werkfilm is also an annex of this specific article:

<https://miau.my-x.hu/miau/quilt/2020/quilt2/launching2020IV08/quilt_2_0.docx>