Investigating mortality data

Tamás Ferenci, Obuda University
ferenci.tamas@nik.uni-obuda.hu

EKIK Day 2026, 24 March, 2026

Table of contents

Possible metrics (and their problems)

- Reported case numbers
- Reported deaths
- Deaths according to death certificates

Excess mortality: a definition

- The difference between the actual mortality of a time period (*observed* mortality) and the mortality of the same period when forecasted from past data (*expected* mortality)
- Rationale: the mortality-modifying event is not present in the past data, so forecasting from them gives a "what-if" data (counterfactual) on what would have been the case without epidemic
- Thus the difference, i.e., the excess mortality represents the effects of the pandemic
- (The method is not specific to epidemics, in fact, it is often used for natural disasters in developing countries)

Excess mortality: a definition

- The difference between the actual mortality of a time period (*observed* mortality) and the mortality of the same period when forecasted from past data (*expected* mortality)
- Rationale: the mortality-modifying event is not present in the past data, so forecasting from them gives a "what-if" data (counterfactual) on what would have been the case without epidemic
- Thus the difference, i.e., the excess mortality represents the effects of the pandemic
- (The method is not specific to epidemics, in fact, it is often used for natural disasters in developing countries)

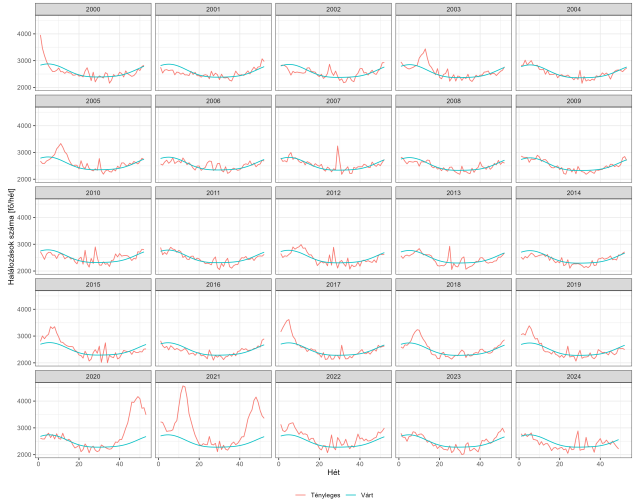
Excess mortality: a definition

- The difference between the actual mortality of a time period (*observed* mortality) and the mortality of the same period when forecasted from past data (*expected* mortality)
- Rationale: the mortality-modifying event is not present in the past data, so forecasting from them gives a "what-if" data (counterfactual) on what would have been the case without epidemic
- Thus the difference, i.e., the excess mortality represents the effects of the pandemic
- (The method is not specific to epidemics, in fact, it is often used for natural disasters in developing countries)

Excess mortality: a definition

- The difference between the actual mortality of a time period (*observed* mortality) and the mortality of the same period when forecasted from past data (*expected* mortality)
- Rationale: the mortality-modifying event is not present in the past data, so forecasting from them gives a "what-if" data (counterfactual) on what would have been the case without epidemic
- Thus the difference, i.e., the excess mortality represents the effects of the pandemic
- (The method is not specific to epidemics, in fact, it is often used for natural disasters in developing countries)

Excess mortality: an illustration

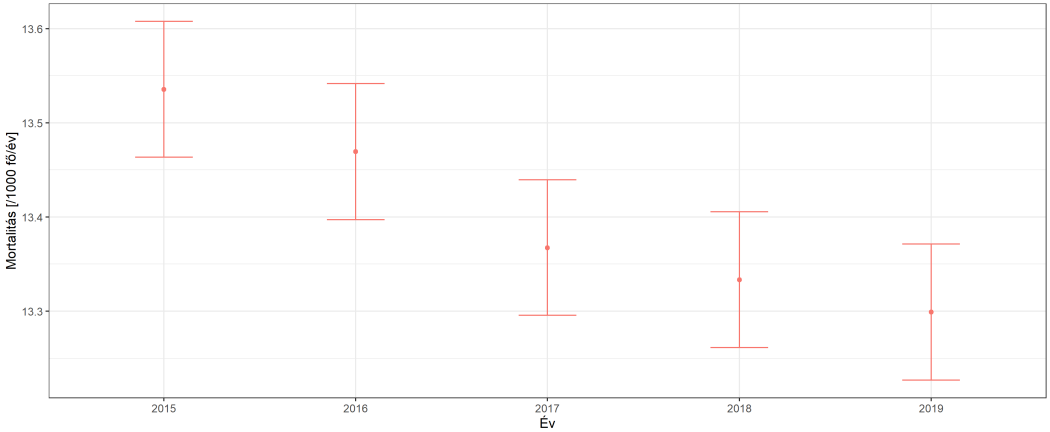


Excess mortality: pros and cons

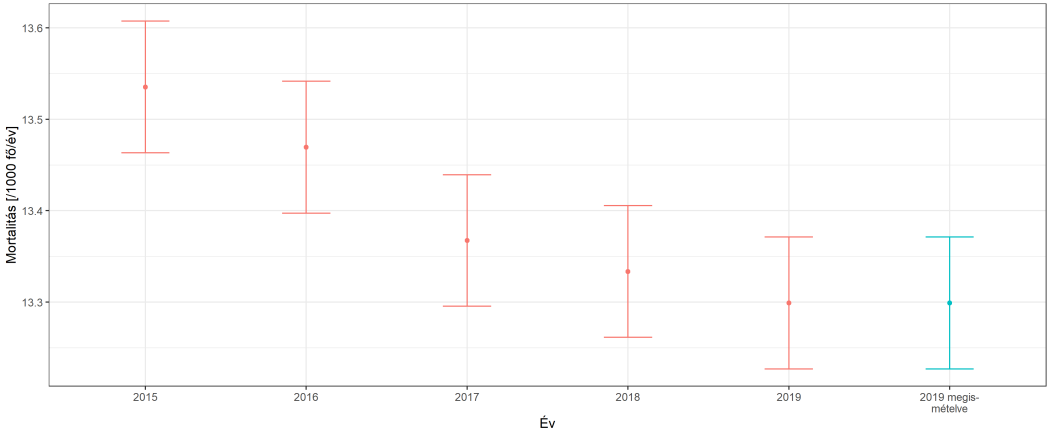
- Pro
  - Now, it is *really* insensitive to testing activity
  - Now, it is *really* insensitive to cause-of-death classification
- Con
  - Very slow (usually published with 4-6 weeks of delay, even that is sometimes revised in the next few weeks)
  - Gross indicator: "effects of the pandemic" actually includes *all* direct and indirect effects – both positive and negative – of the epidemic and its handling
  - Requires an estimation of the baseline, subject to forecasting error

Table of contents

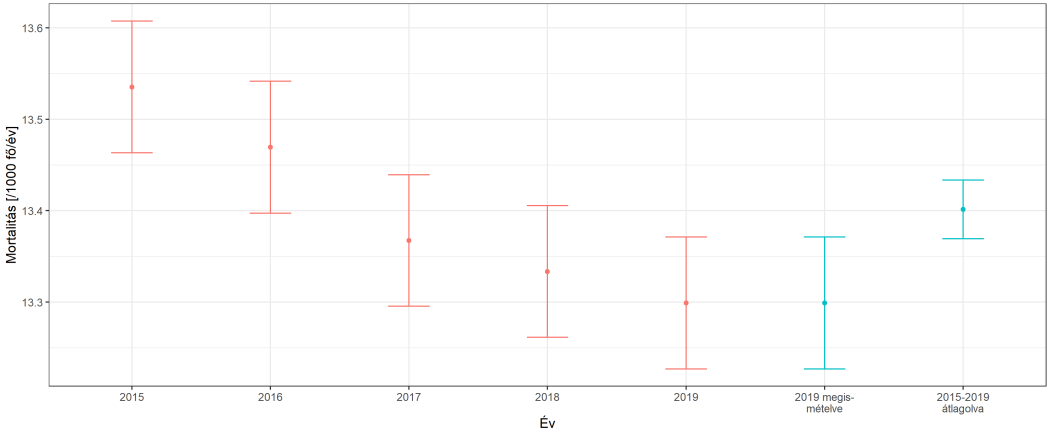
An example country



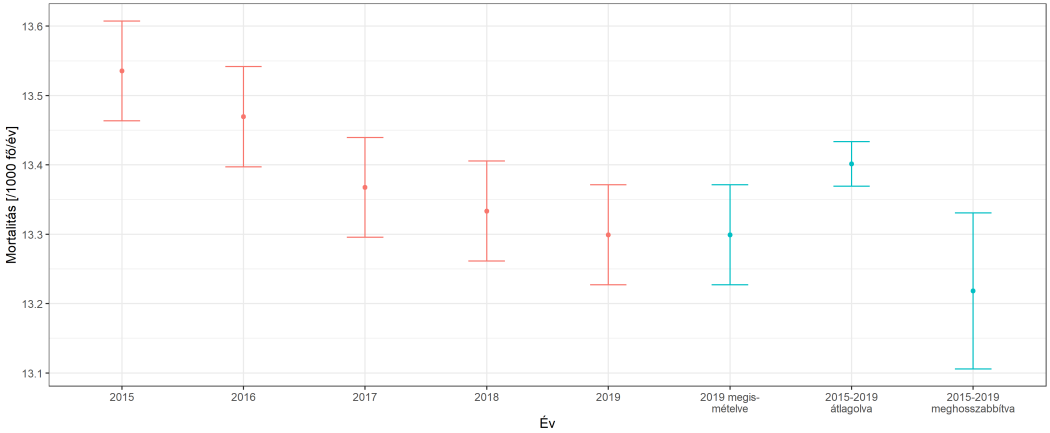
Last year carried forward



The average of the last few years



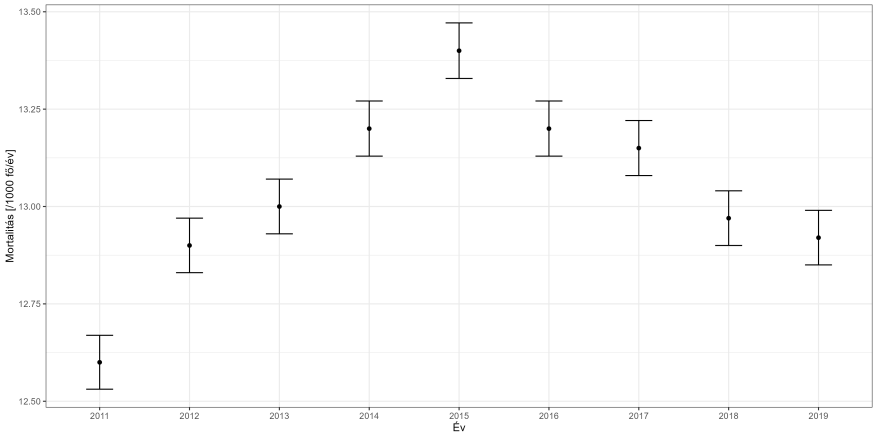
Linear extrapolation based on the last few years



Some conclusions

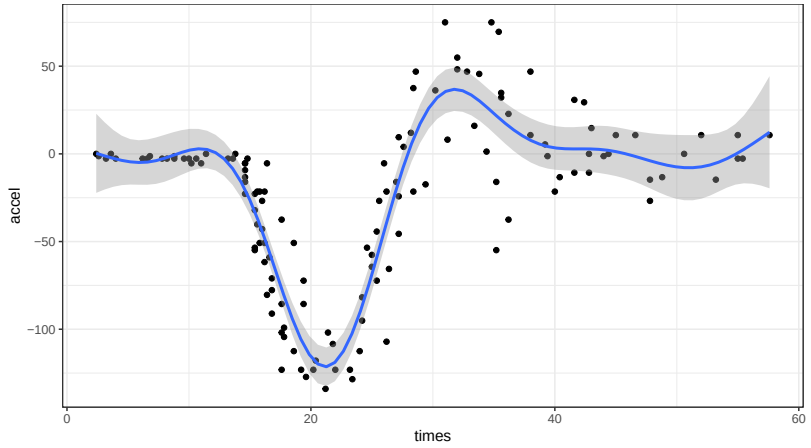
- The method matters...
- ...especially given that the subtraction will magnify errors (let's say the observed value is 1,000 deaths, then an expected of 990 vs. 900, a 10% difference, gives rise to an excess mortality of 10 vs. 100 respectively, a 100% difference!)
- Variance-bias trade-off

Even linear extrapolation is not straightforward: where to begin?

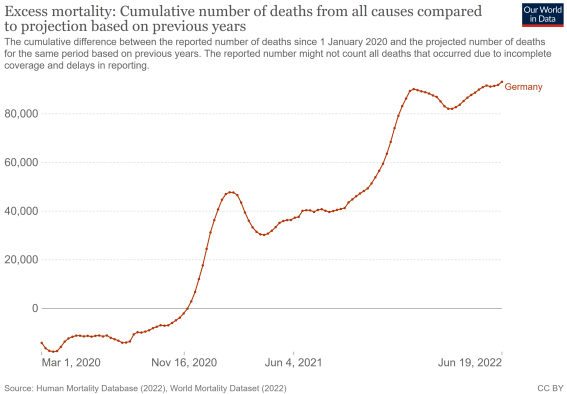


The best (?) solution therefore: using splines

A logical idea, but...

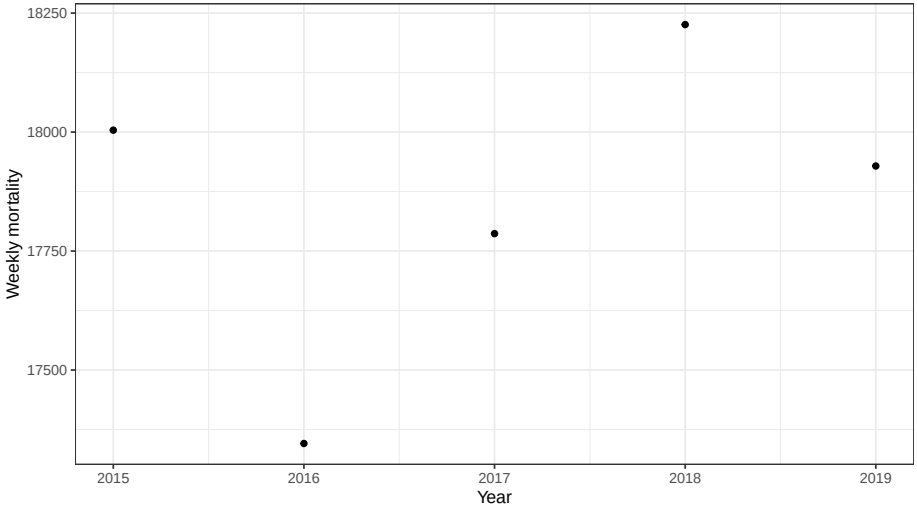


"The German Puzzle"

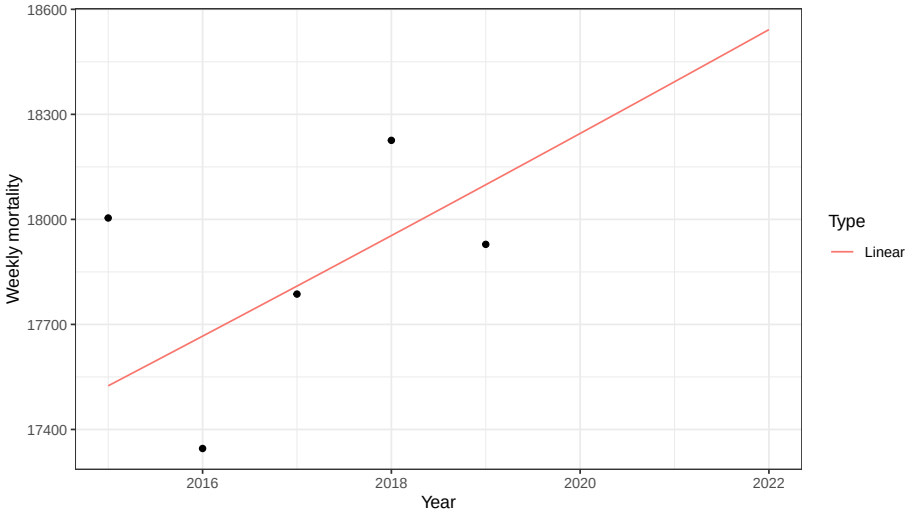


Compared to that, the estimate of the WHO presented on May 5, 2022 was 195,000!

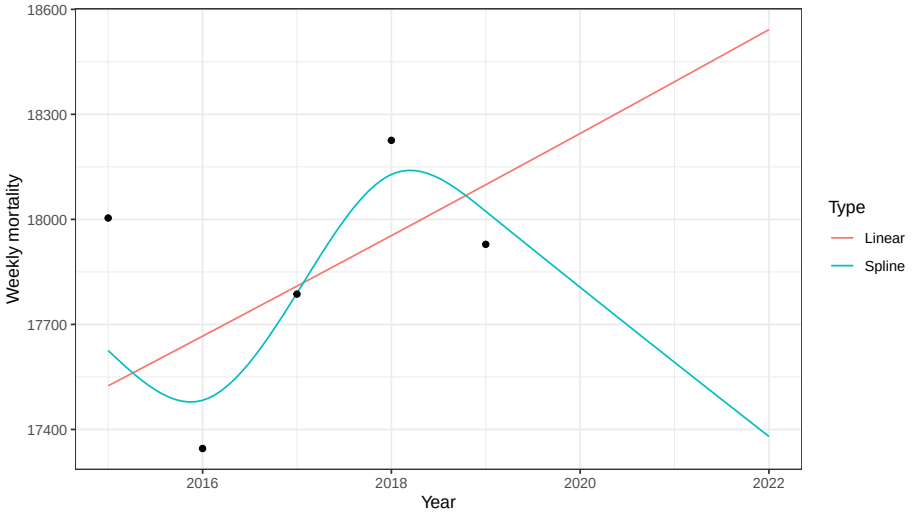
„The German Puzzle”



„The German Puzzle”



„The German Puzzle”



Our work

- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Our work

- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Our work

- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Our work

- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Our work

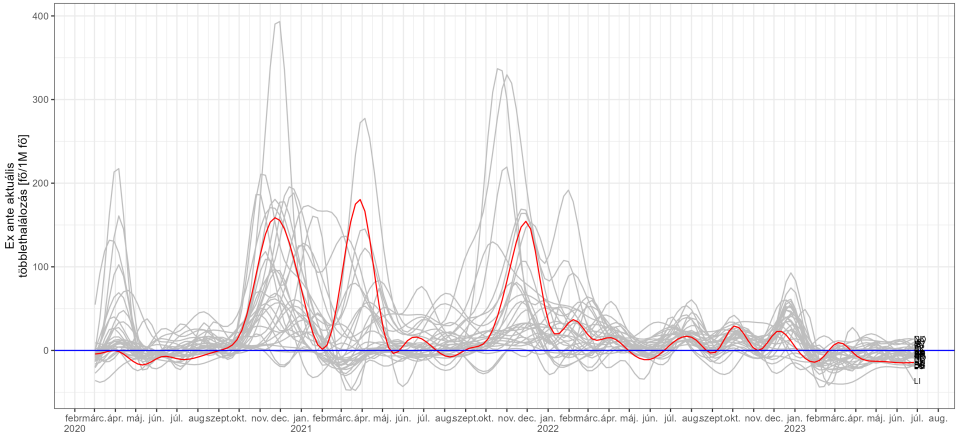
- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Our work

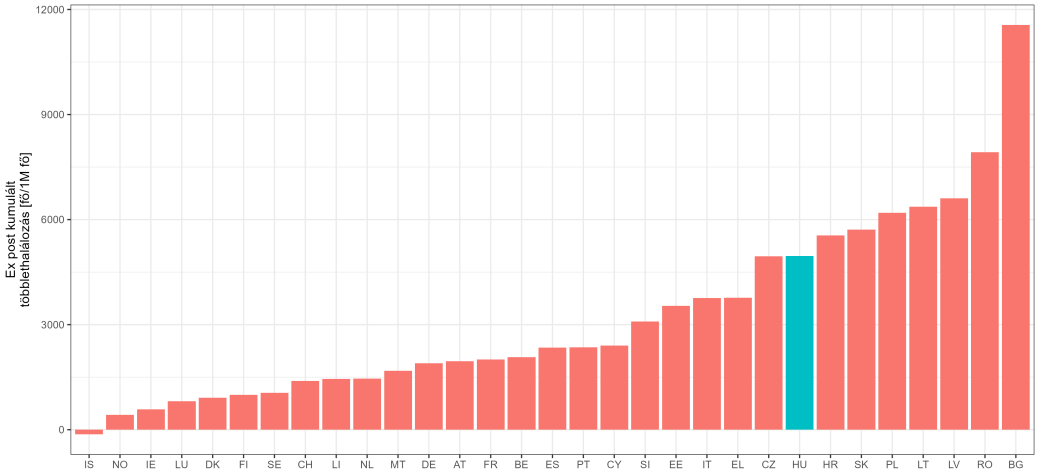
- We carried out an empirical study (extensive simulation using synthetic datasets)
- To clarify whether the usage of splines was the culprit *per se*
- What properties – both of the data and of the method – gave rise to this problem?
- Based on this, what is the best method (and parametrization) to use?
- Published results: T Ferenci. Comparing methods to predict baseline mortality for excess mortality calculations. BMC Med Res Methodol. 2023 Oct 18;23(1):239. doi: 10.1186/s12874-023-02061-w.
- Full source code (in R) available at `https://github.com/ferenci-tamas/MortalityPrediction`

Table of contents

Ex ante results during the epidemic



Ex post snapshot at the end of the epidemic



Ongoing monitoring

- The whole process described above has been implemented in one, single informatics pipeline (using R)
- Available at `github.com/ferenci-tamas/ExcessMortEUR/`
- Capable of continuous, automatically updated monitoring via CI/CD

Table of contents

Rationale

- Causes of deaths, even with their limitations, are relevant for public health analysis
- They are publicly available through the WHO
- So here the problem is not really monitoring, but rather the interpretation and dissemination of the results
- But we developed a portal to facilitate this!

Rationale

- Causes of deaths, even with their limitations, are relevant for public health analysis
- They are publicly available through the WHO
- So here the problem is not really monitoring, but rather the interpretation and dissemination of the results
- But we developed a portal to facilitate this!

Rationale

- Causes of deaths, even with their limitations, are relevant for public health analysis
- They are publicly available through the WHO
- So here the problem is not really monitoring, but rather the interpretation and dissemination of the results
- But we developed a portal to facilitate this!

Rationale

- Causes of deaths, even with their limitations, are relevant for public health analysis
- They are publicly available through the WHO
- So here the problem is not really monitoring, but rather the interpretation and dissemination of the results
- But we developed a portal to facilitate this!

Ongoing monitoring

Available at `https://research.physcon.uni-obuda.hu/OkspecifikusMortalitasiAdatbazis/`

Thanks for your attention!