Methodology for the use of judgmental forecasting in public decision-making

Prague, 2022

T A
Č R
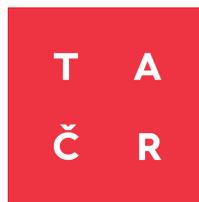
Author's collective

Mgr. Jan Kleňha, Czech Priorities, z.ú.

Ing. Vojtěch Brynych, Czech Priorities, z.ú.

Marek Havrda, PhD. Czech Priorities, z.ú.

PhDr. Ladislav Frühauf, Czech Priorities, z.ú.

Dedication

Objective of Methodology and Justification of Procedures

The main objective of this document is to present the procedures that public institutions should follow when facilitating forecasting tournaments or similar methods of judgmental forecasting in an understandable form. On the basis of the obtained predictions of future developments, public institutions in the Czech Republic can better plan their own activities, make decisions, or recommend the implementation of various measures. This document can also serve private institutions that are considering the use of forecasting tournaments for their own purposes, or non-profit organizations that find the methods of judgmental forecasting useful in making their own recommendations or that want to support applied research, development and implementation of these methods in public decision-making in the Czech Republic.

Procedures are described in this document according to a realistic time sequence (from the selection of the platform and recruitment of participants, to the creation of questions and administration, to the aggregation of outputs and the rewarding of participants), with emphasis on potentially problematic parts where particular care should be taken.

In addition to the actual process of facilitating forecasting tournaments, the methodology also addresses the steps that should precede the actual organization of the tournament, in particular the analysis of the suitability of using this tool in a given situation and the creation of a plan for how the predictions obtained will be subsequently used. The appropriate post-tournament steps are also described, which mainly concern the effective communication of the results both inside and outside the organization.

The procedures presented in this methodology have not been described in the Czech environment yet due to the considerable novelty of this method on an international scale. During the development of this methodology, these procedures were practically tested in the Czech environment and, without this information, there is a significant risk of incorrect tournament execution. This can lead to suboptimal or completely erroneous conclusions, which, due to the impact of public administration decisions, can have far-reaching consequences.

In view of the expected rapid development in this area, it is recommended to check whether superior tools or approaches that can improve these processes have been developed before utilizing this methodology.

Introduction

Judgemental forecasting

Judgemental forecasting is a method of informed estimation of future developments most often used in complex, multi-sectoral problems or in situations without data for accurate modeling of future development. This method, unlike statistical forecasting, is based on the formation of subjective judgements by individuals who share their estimates of future developments in the form of probabilistic estimates that can be more easily aggregated. Therefore, this approach is sometimes also called "probabilistic forecasting".

The principle of the so-called "wisdom of the crowds" is fundamental to judgmental forecasting. This principle was named by J. Surowiecki[1] and further extended by, for example, C. Sunstein[2] in the first decade of the 21st century, but this concept has been known since at least the early 20th century.[3] This principle is based on the observation that the aggregate prediction of a group of people is often more accurate than the judgment of an individual. Judgmental forecasting uses the aggregation of many opinions, inputs, and differing points of view to make an informed prediction of future events, outcomes, trends, or developments.[4] These estimates can be useful in various areas of public decision making, such as economic forecasting,[5] political elections,[6] and public policy.[7]

Group deliberation using judgmental forecasting can take many forms, with the ways in which participants are motivated and how they share information with each other significantly impacting outcomes. For example, the Delphi method[8] is a form of group deliberation in which judgmental forecasting can be applied. Delphi participants tend to be purposefully selected based on expertise, rewarded in a standardized way for participating in the project, and share information anonymously through a facilitator between rounds. However, the Delphi method has several limitations,[9] which can be addressed through innovative and increasingly-applied forms of deliberation, such as forecasting tournaments.

Because of its novelty, scientific robustness, and high application potential, this methodology focuses specifically on the application of forecasting tournaments in the Czech

---

[1] Surowiecki, J. The Wisdom of Crowds (Anchor, 2005).

[2] Sunstein, C. Infotopia: How Many Minds Produce Knowledge (Oxford University Press, USA, 2006).

[3] Galton, F. Vox populi. Nature 75, 450-451 (1907)

[4] Tetlock, P., Mellers, B., et al. (2014). Forecasting tournaments: Tools for increasing transparency and improving the quality of debate, Current Directions in Psychological Science, 23(4), 290-295.

[5] Budescu, D. V. & Chen, E. Identifying expertise to extract the wisdom of crowds. Manage. Sci. 61, 267-280 (2014).

[6] Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. Judgment and Decision Making, 6, 73-88.

[7] Morgan, M. G. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc. Natl Acad. Sci. USA 111, 7176-7184 (2014).

[8] Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. Management Science, 9 (3), 458-467.

[9] Kleňha, Jan, "Improving strategic foresight with the use of forecasting tournaments", International Relations-International Relations, (2022), internationalrelations-publishing.org/articles/2-4.

Republic. However, the general principles of the application of judgmental forecasting described in this methodology are applicable to other forms of deliberation.

Forecasting tournaments

The Forecasting Tournament is an innovative tool that, using financial and social incentives, can ensure effective collaboration of a group of tournament participants (hereinafter referred to as "forecasters") and dynamically obtain high-quality expert estimates and arguments, both for strategic planning and, for example, for urgent situations where there is not enough data or time for complex quantitative analysis.

In addition to aggregation of opinions and probabilistic estimates, a forecasting tournament can overcome the blind spots of conventional analytical methods, minimize the effects of cognitive biases in experts, and generally ensure greater nuance, accuracy and diversity in decision making. Forecasting tournaments thus have the potential to improve not only the quality of policy decision-making, but also, for example, public awareness of and therefore trust in politics.[10]

Forecasting tournaments are normally attended by at least a few dozen respondents, who publish their probabilistic predictions to given questions on an online platform. They can modify the published predictions according to their own judgment and supplement them with explanations in the form of comments, which allows them to communicate with each other in real time to inform and complement each other.[11] Participants with the most accurate predictions are evaluated ex-post and rewarded according to the success of their predictions. Tournaments can vary greatly in duration depending on the nature of the questions asked.[12]

Forecasting tournaments in public decision-making

Forecasting tournaments are useful in public decision-making, especially in complex, multiscale problems or in situations with high uncertainty and lack of data or historical experience. The best known examples of the use of forecasting tournaments for public decision support are The Good Judgment Project (USA)[13] and the current *Cosmic Bazaar* project (UK).[14] In the past year, forecasting tournaments in public decision making have

---

[10]Jason, D., Atanasov, P., Tetlock, P., Mellers, B. (2019). Are markets more accurate than polls?, Judgment and Decision Making 14, no. 2 (2019): 135-147.

[11] Mellers, Barbara, et. al. al, and "Identifying and Cultivating Superforecasters" *Perspectives on Psychological Science 2015*, Vol. 10(3); Mellers, Barbara, Philip Tetlock, et al, "Forecasting tournaments, epistemic humility and attitude depolarization." *Cognition 188* (2019): 19-26.

[12]Other relevant sources include:

Review of "Superforecasting: https: //bit.ly/superforecastingreviewAre Markets more accurate than Polls? : http://journal.sjdm.org/18/18919/jdm18919.pdf

Identifying and Cultivating Superforecasters: https://stanford.edu/~knutson/nfc/mellers15.pdf,

Evidence on Good Forecasting Practices: https://bit.ly/evidenceongoodforecasting

[13]Good Judgment Inc (2021, October 19). *See the future sooner with Superforecasting*. Good Judgment. https://goodjudgment.com/.

[14]The Economist. (2021, April 21). *How spooks are turning to superforecasting in the Cosmic Bazaar.* https://www.economist.com/science-and-technology/2021/04/15/how-spooks-are-turning-to-superforecasting-in-the-cosmic-bazaar

received considerable attention[15] especially in the context of predicting the global evolution of the Covid-19 pandemic.[16]

In the Czech environment, the first project testing the use of forecasting tournaments in public decision-making was the project TL04000315 - *Expert Prediction Team for Intensive Forecasting in Emergency Situations* (OPTIONS), which was the basis for this methodology.

---

[15]The Economist. (2021, April 21)
[16]Time. (2020, June 11). https://time.com/5848271/superforecasters-covid-19/

Basic terminology

Judgemental / Probabilistic forecasting

- A method of estimating future occurrences in the form of probabilistic estimates.

Crowd forecasting

- A method of deliberation of a group of participants using judgmental forecasting. Participants share their own judgements with each other, which are aggregated on an ongoing or intermittent basis.

Forecasting tournament

- A form of crowd forecasting that uses financial and social incentives to encourage active participation and information sharing to improve the accuracy of aggregate forecasts.

Prediction (Prediction / Forecast)

- Estimate of future developments. In the case of forecasting tournaments, the answer to the question asked, where this answer takes the form of a percentage estimate and associated textual commentary.

Pilot implementation

- The forecasting tournament, which we organized in 2021 in the Czech Republic as part of the OPTIONS research project (TL04000315) and which was the basis for this methodology.

Forecaster

- Participant in group forecasting (e.g. forecasting tournament).

Aggregate

- A summary of all predictions made on a given specific question. Describes the aggregate response of all participants to the question.

Calibration

- A person's ability to consistently form high quality probabilistic judgments.

Facilitation of forecasting tournaments

The following methodology is divided into three chapters according to the three basic stages of the correct procedure:

1) **Preparation**

2) **Tournament organisation**

3) **Use of outputs**

1. Preparation

1.1 Tool suitability analysis

Before designing a forecasting tournament, it is advisable to thoroughly analyze whether deliberation using judgmental forecasting is the most appropriate tool for supporting the pertinent decision. To this end, a list of assumptions that should be met is provided.

- **Decision-making institutions have a real interest in predictions**

  - Real interest should be manifested not only by understanding the forecasting itself and the usefulness of probabilistic predictions, but also by an interest in participating in the design of the questions, which is essential for producing useful outputs.

- **Unavailability of other methods that have been shown to deliver more robust results**

  - In some topics and situations (e.g., when a large amount of money and time is available for additional data collection or for the analysis itself), quantitative analysis (e.g., cost-benefit analysis) can provide as much or more quality information, but often it is appropriate to conduct research using deliberative tools. For these purposes, it is sufficient to have time on the order of weeks and resources to prepare for the tournament and to mobilize existing - or recruit new - participants. To increase the robustness of the outputs, it is possible and appropriate to combine judgmental forecasting with other quantitative analyses.

- **The answer to the problem is complex and robust prediction is needed**

  - The anticipated development of the situation on this issue is the result of several different factors, and expertise in multiple fields is necessary to fully understand them. At the same time, the supported decision is important and requires maximum robustness of prediction, not just the individual opinions of selected experts on the issue.

At this stage, it is also necessary to consider the two main risks of using this method. The first risk is the loss of interest of the organisation's management or partners in the

predictions made. This risk can be prevented by consistent transparency and communication throughout the project to build mutual trust. Uncertainty about acceptance by the media and the public can also be a concern, which can be mitigated by careful selection and definition of questions, as well as by presenting financial rewards as just one incentive, relatively minimal to the cost of actually influencing the outcome, in case someone is interested in influencing reality to win the tournament, which has been described as a risk factor in the past.[17]

The second risk is low quality of forecast outputs despite good recruitment, for example due to participants underestimating the time commitment. This risk can be reduced by slightly modifying the rules even during the tournament, for example by offering additional rewards. The presence of participants who would actively boycott the tournament or try to deceive it can be dealt with by removing the participant from the competition. Recruiting from the already-established group of forecasters who have demonstrated their ability and motivation to participate responsibly in the pilot tournament organized as part of the project that resulted in this methodology, can also serve to reduce this risk. The possible participation of individuals with access to relevant internal information to help them make better predictions is beneficial for this type of competition and their involvement is not to be feared because, if any of the tournament participants have access to non-public information, their predictions may move the aggregate towards a more correct outcome.

1.2 Possibilities of using prediction

For a more detailed understanding of what specific information can be obtained using a forecasting tournament, the following is an overview of foreign experiences with the application of tournaments in different situations and to different types of questions, which can serve as inspiration.

Forecasting differs from traditional methods of predicting the future in that it provides quantification and strong feedback as to whether or not the prediction was correct. However, the method obviously cannot answer all questions. Forecasting tournaments are best suited for use where the question and possible answers can be clearly defined. At the same time, a tournament can easily be combined with traditional forecasting methods such as scenario planning and quantitative analysis. These methods can also serve as background and supplementary information for forecasters in developing their estimates.

For example, if the goal is to determine future Sino-US relations, then the question "Will Sino-US relations be more strained next year than this year?" cannot be asked in a forecasting tournament because it is not clear how such a question can be evaluated. First, it is necessary to identify outcomes that describe the target situation and that can be quantified. Only in light of these trends can suitable questions for forecasters be developed. For the above example:

- "What will be the total volume of trade between China and the US?"

- "What percentage of U.S. O-1 visas will go to Chinese citizens?"

---

[17]Hanson, Robin, "Designing Real Terrorism Futures" Public Choice (2006),
https://www.jstor.org/stable/30026644

- "What percentage of U.S. citizens will report an 'unfavorable view of China' in the *Pew Global Attitudes Survey*?"

- "How many times will Japanese Air Force aircraft have to take off because of threats to Japanese territorial airspace by Chinese military aircraft?"

It is already possible to make clearly quantifiable predictions on these questions, which together give an overall picture of the situation for which quantified predictions cannot easily be made. [18]

1.2.1 Short-term forecasts (0-1 year)

Short-term forecasting can be used, for example, in crisis situations. During the covid-19 pandemic, the results of predictions were used to estimate deaths and new cases of disease.[19] Another possible use of short-term predictions, is for the results of scientific studies,[20] where predictions can improve the design of these studies before they begin.

1.2.2 Medium-term forecasts (1-3 years)

Medium-term forecasts can be used to predict public behavior or to estimate the impact of various measures. For example, the Cosmic Bazaar[21] tournament in the UK, in which civil servants participate, is used for this purpose. Another example is the Anticipation Hub,[22] where potential disasters are predicted in collaboration with development and non-profit organizations.

1.2.3 Long-term forecasts (3+ years)

Long-term predictions on specific issues tend to be highly imprecise due to the increasing complexity of possible influencing factors, but they can still be used to forecast geopolitical or technological trends. Examples are the Hypermind projects[23] or the US think-tank Center for Security and Emerging Technology,[24] which works with several US governmental departments.

1.3 Tournament Design

Before the tournament itself, it is necessary to think about its overall intention and then look for appropriate technical and organizational implementation. There are several questions

---

[18]This approach is developed by CSET Foretell and the above example is also taken from their documents. https://cset.georgetown.edu/wp-content/uploads/CSET-Future-Indices.pdf.

[19]Law, T. (2020, June 11). *"Superforecasters" Are Making Eerily Accurate Predictions About COVID-19. Our Leaders Could Learn From Their Approach*. Time. https://time.com/5848271/superforecasters-covid-19/

[20]Replication Markets. (n.d.), Replication Markets, Reliable Research Replicates. https://www.replicationmarkets.com/

[21]The Economist, *How spooks are turning to superforecasting in the Cosmic Bazaar*.

[22] Waldvogel, F., Rüth, A., Siahaan, K. (n.d.). *ANTICIPATION HUB*. Forecast-Based Financing. https://www.forecast-based-financing.org/anticipation-hub/

[23] https://docs.google.com/document/d/1fRg7twB2RLAc-Ey8NUj5qFUJCg-dp3yb/edit

[24] Page, M. (2021, March 23). *CSET Foretell Blog - Crowd Outperforms Projections from Historical Data in Early Results*. CSET Foretell Blog. https://www.cset-foretell.com/blog/crowd-performance-analysis

that need to be answered that will affect the overall complexity of the tournament preparation.

1.3.1 Platform

It is possible to use commercial or open-source platforms to organize a forecasting tournament. When choosing a platform, it is advisable to consider the following:

Cost - Commercial platforms may require payment of relatively high fees, but they will be tested and you can contact IT support if you have any problems. On the other hand, open-source platforms cost nothing out of pocket, but you need to have your own IT team ready to deal with any problems very quickly. When negotiating a collaboration, you need to check whether and how the data obtained on the platform can be exported and used after the tournament.

Prediction scoring - Platforms use different ways to evaluate the resulting predictions and then calculate a score. The two most common ways of calculating the score are Brier scoring[25] and logarithmic scoring.[26] The Brier score is calculated by averaging over all squared differences between predictions and the actual outcomes to produce a number between 0 and 1, where a lower number indicates a better prediction. The Brier score is "more moderate" than the logarithmic score, which gives much more weight to extreme predictions, so if a forecaster answers a question very confidently (for example, close to 100%) and is wrong, he will lose so many points that he will no longer have a chance to place in the tournament.

> In the pilot implementation of the forecasting tournament in the Czech Republic (within the OPTIONS project), we chose Brier scoring, as we considered it appropriate that our number of questions (24) would not overly disadvantage participants who made one bad confident prediction, which we assumed could happen even to participants with great potential but who have less experience with forecasting.

Localization and support - Few platforms offer Czech localization, but most platforms should be able to provide it at an additional cost. Some platforms also offer expert consultants who are experienced in tournament organization and question creation.

Prediction method - Some platforms allow you to input a prediction using an arbitrary probability distribution within a given range, others allow you to split the responses into "bins", where forecasters then predict the probabilities of each bin.

Example of prediction using an arbitrary distribution:

---

[25]Brier, Glenn W. "Verification of forecasts expressed in terms of probability." Monthly weather review 78, no. 1 (1950): 1-3.
[26]Bickel, E. J., "Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules", Decision Analysis, 2007, no. 4, 2, pp. 49-65.

10k   15k   20k
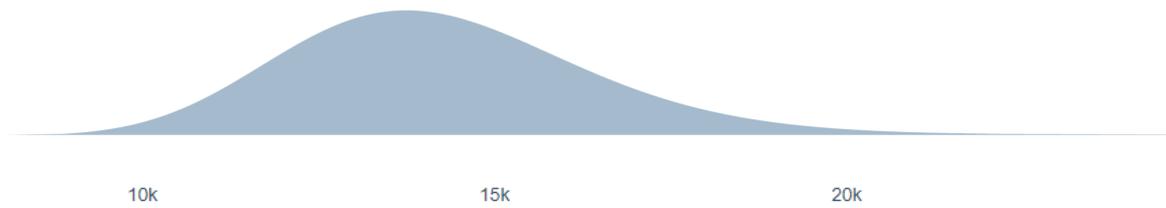
Foretold.io, 2021

Examples of two options for displaying prediction using compartments:
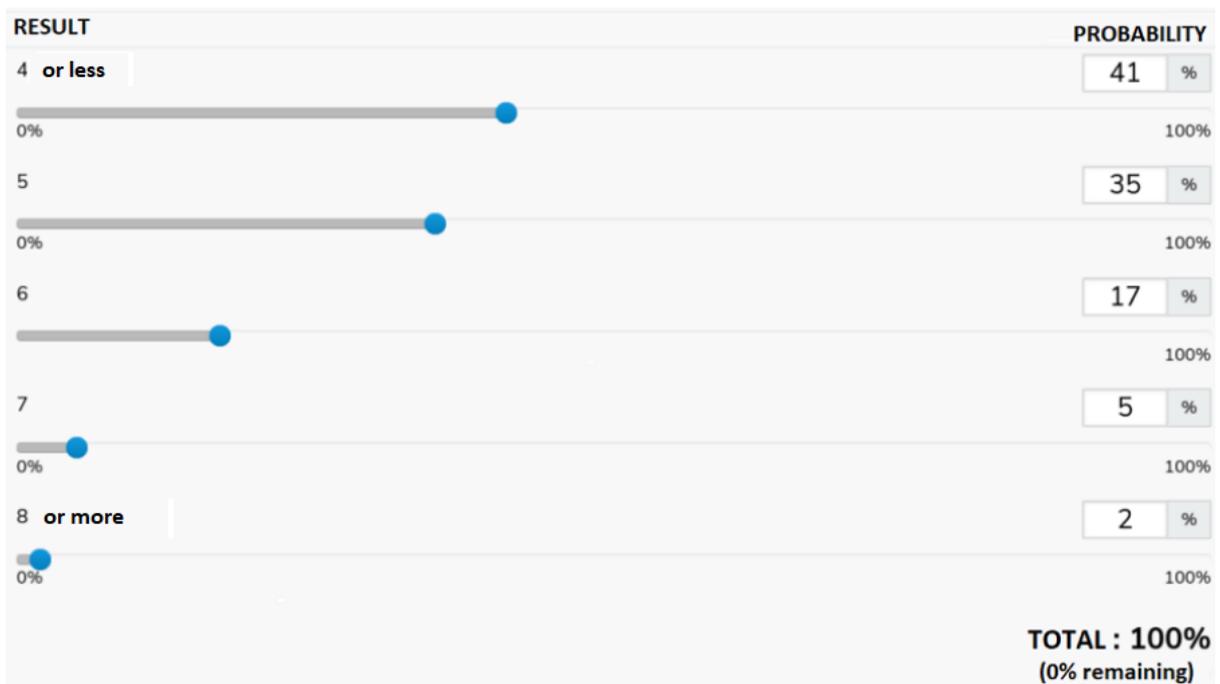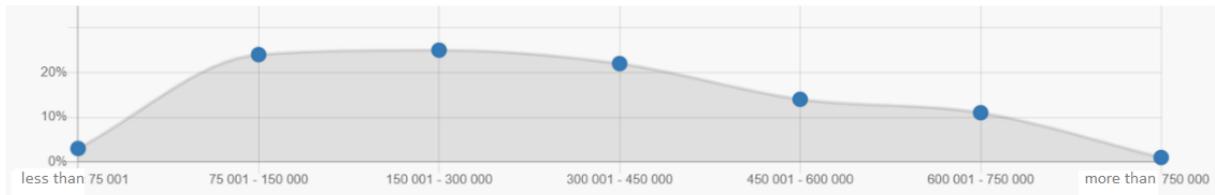
Cultivate Labs, 2021

1.3.2 Tournament length

The duration of the tournament depends on how urgently the end-user needs to obtain an aggregate judgment, and how far in the future the resolution of the questions being asked will be. If it is necessary to quickly obtain a prediction of an event that will occur after a considerable amount of time, participants can be incentivized to make multiple predictions by limiting the time they can make them. The resulting aggregate of predictions will be obtained relatively quickly, but its quality will improve significantly once a sufficient number of questions have been answered for quality forecasters to be selected. On the other hand,

in a situation where maximum prediction quality is important and it is not necessary to obtain informative input as soon as possible, it is advisable to leave the question open for a longer period of time (months or even years), though these will require additional incentives to continuously motivate participants provide input and adjust their judgments, even if the tournament evaluation is a long time away.

> As a pilot project, we divided the tournament into four rounds of ten days each. Each round started on Thursday morning and ended at the same time 10 days later (Monday). This was followed by a three-day break, and the next set of questions was posted again on Thursday morning. The main objective was to identify outstanding forecasters. The division of the tournament into several rounds, together with the obligation to answer at least 4 out of 6 questions in each round, resulted in high motivation of the participants and half of the forecasters entered met the tournament requirements. This provided sufficient data to identify outstanding forecasters. However, it is difficult to sustain such a high level of forecaster participation in the long term, so the tournament formatting requires some refinement.

1.3.3 Number of questions

There are no exact limits on the maximum number of questions (which themselves can vary widely in difficulty) in a tournament, but as the number of questions increases, the influence of chance in identifying the best forecasters decreases. Conversely, when using already-identified excellent forecasters, a smaller number of questions is advantageous, as this allows participants to pay more attention to individual questions.

> In the pilot tournament we asked 24 questions and participants were allowed to skip a maximum of 4 questions. We consider this to be a sufficient number to detect outstanding forecasters while maintaining motivation to participate. With a larger number of questions, there is a risk that a large number of forecasters will not complete the tournament due to excessive time requirements, unexpected emergencies, or other reasons.

1.3.4 Aggregate display

The aggregate of predictions, together with the aggregate of textual comments, is the main output which the tournament is trying to attain. It is a summary of all predictions on a single question and shows the opinion of the entire group of participating forecasters. The detailed method of calculating this aggregate is non-public on most platforms, but an example could use the average of the last ⅔ predictions, taking into account only the last prediction of each participant. A minimum of 10 predictions and a maximum of 150 predictions are used for the calculation. Any prediction less than 3 days old is included in the calculation. It may also include an algorithm that takes into account the previous results of each forecaster and

gives those with better scores more weight in the aggregate calculation. A graphical presentation of the aggregate may look like this:

5. How many cases of chickenpox in the Czech Republic will be registered by the State Health Office during the month of May 2021?

Number of participating forecasters: 211

The number of cases of chickenpox registered in the Czech Republic by the State Health Office in May 2021 according to preliminary data can be found here: www.szu.cz/publikace/data/2021. We will evaluate the question from the data published on the website of the State Health Office for May 2021, which will be available on the website of this institution on July 31, 2021 at 5:00 p.m.

Possible answer	Community prediction
less than 500	7.53 %
501 - 1000	24.11 %
1001 - 1500	33.88 %
1501 - 2000	24.57 %
more than 2000	9.91 %

Czech Priorities, 2021

Although some platforms give more weight to predictions from good forecasters and thus significantly refine the resulting aggregate, this approach is only appropriate when a sufficient number of questions are prepared to reveal capable forecasters.

1.3.5 Displaying the ranking score

Forecasting tournaments clearly quantify all predictions, making it possible to create accurate rankings of the ongoing success of individual forecasters. Public leaderboards motivate the best forecasters to compete against each other, but strongly discourage forecasters who underperformed and have low scores at the beginning of the tournament.[27] Similarly, public rankings can strongly discourage participation by senior staff or, for example, colleagues from the same institution or department who may fear for their reputations if others outperform them. This effect may significantly reduce the number of participants who complete the tournament. For this reason, when planning a tournament, it is essential to think about the tradeoff between finding the best forecasters versus having the widest possible participation.[28]

A possible solution is the publication of only the results of entire teams (which can be made up of members of individual departments), which can also motivate inter-team competition.

---

[27] Horowitz M., et. al. al, "Keeping Score: A New Approach to Geopolitical Forecasting", Perry World House (2021), available at https://global.upenn.edu/sites/default/files/perry-world-house/Keeping %20Score%20Forecasting%20White%20Paper.pdf
[28] Ibid.

This should be complemented by private leaderboards where the forecaster can see his own progress and improvement. The demotivation of a low position on the performance ladder can also be reduced by resetting it regularly or by introducing more rankings for different subject groups of questions. Forecaster motivation will also be increased if the ranking also takes into account forecaster activity/awards for best comments and not just accuracy, though not all platforms support this.

1.4 Participants

Quality forecasters can be found in almost any field of expertise. The data from the pilot implementation in the Czech Republic show that the main motivation to participate in the tournament was the opportunity to self-develop and test one's own skills. Participants can therefore be sought, for example, among students and at universities, in both the natural sciences and the humanities.

1.4.1 Recruitment

The great advantage of forecasting tournaments is precisely the ability to aggregate a large number of different inputs, so it is usually advisable to keep the recruitment relatively broad and get a large number of participants and therefore a wide range of views and opinions. However, if there is a need or end-user interest in participants who already possess certain skills or capabilities, the recruitment can be narrowed. For example, the Cosmic Bazaar project, organized starting in2020 in the British Civil Service, is only open to civil servants.

Only a small proportion of those interested in forecasting are known to be excellent forecasters without prior training, so it is advisable to teach the participants to work with probabilities before the tournament. The introductory training described in the previous chapters provides an example of such training. Even untrained participants can contribute valuable predictions by, for example, writing interesting comments that will influence other forecasters, but a short introductory training seems to be a very cost-effective way to increase the quality of the final outputs. A good forecaster can answer questions from many disciplines and alert subject matter experts to possible non-intuitive overlaps or cross-sector impacts, so it is advisable to recruit participants with no expertise in areas relevant to the questions being asked.

At the same time, however, it is necessary that at least some forecasters who have expertise in the core areas participate in the tournament. Indeed, the premise of judgmental forecasting is not to replace any expertise with the wisdom of crowds, but to be able to minimize the potential cognitive biases of experts on highly specific topics by effectively aggregating many judgments based on very diverse information sources, knowledge, and foundations.

The recruitment materials should mention the competitive aspect of the tournament and the size of prizes, the lack of the need to make a financial contribution (which is the case, for example, with prediction markets), the possibility of anonymity, the ability to withdraw from the tournament at any time, the time frame in which the tournament will take place, the time required for any mandatory entry training, and the expected intensity of participation or the expected time required to answer questions.

During the recruitment process, the subsequent organization of the entire tournament, and the presentation of the results, it is necessary to comply with all the principles of working with personal data according to the GDPR rules.

1.4.2 Selection

The group does not need to be representative of the population or scientific disciplines, but it needs to be sufficiently diverse in opinion to include minority views, specific experiences, knowledge, and different backgrounds. This can be achieved through broad recruitment at diverse institutions, across demographic groups, age categories, and regions. Topic-relevant social media groups or newsletters from universities and educational institutions may be appropriate channels for disseminating recruitment information. At the same time, creating your own website is recommended.

The potential overflow of underqualified candidates can be addressed by selecting participants according to demonstrated motivation, for example by making it mandatory for all forecasters to attend calibration training. It is important to organize the tournament in such a way that anonymity (or at least the possibility of remaining anonymous) is preserved for all participants, which should be emphasized in the recruitment materials.

In the pilot implementation, we targeted recruitment of thousands of potential participants by sharing in thematically relevant interest groups, sending out requests for inclusion in faculty newsletters of all Czech universities, and actively searching for and reaching out to organizations bringing together potential participants from all regions and demographics of the Czech Republic. During the recruitment process, we gave an estimated time commitment of 1 hour to answer each question.

The registration form was filled out by 838 participants, the mandatory initial training (lasting 1.5 hours) was completed by 534 participants, who then had accounts created on the platform. The entire tournament, lasting 8 weeks, was completed by 216 participants. The most common age bracket of participants was 28-34 years, while the youngest participant was 16 years old and the oldest was 65 years old. 68.2% of participants had at least a Master's degree and all areas of FORD[29] professional classification were represented with three exceptions. According to the final questionnaire, participants spent a total of 2,838 hours making predictions and made over 10,000 predictions and comments on them.

1.5 Rewards

Rewards can be set to motivate participants of different ages and socioeconomic backgrounds to participate, while signaling that it is appropriate to take the tournament seriously. Nevertheless, it can be expected that a significant proportion of quality forecasters will be experts in their fields and the main reason for them to participate will the opportunity for self-development or other abstract motives, not the financial reward. It is imperative to set the rewards in such a way that forecasters strive to have the best possible

---

[29] https://www.tacr.cz/dokums_raw/ck/FRASCATI_MANUAL.pdf

score, but at the same time work together and help each other. Rewarding the authors of the best comments, for example, can help with this. Social recognition from other participants, for example by being able to rate comments as helpful, also improves the motivation to write quality comments..

As part of the pilot implementation of the forecasting tournament, we chose to give cash prizes to the top thirty forecasters, as well as cash prizes to the twenty comments that were judged to be the best by our project team. The encouragement of comments was one of the factors that led to a collaborative atmosphere among the forecasters and the production of very high quality comments, which we then added to the final reports along with aggregate forecasts. All forecasters in the tournament were required to write one comment for each prediction made..

1.6 Calibration training

The effect of different types of training on individuals' ability to make accurate predictions has been the subject of several academic studies and research projects.[30,31,32] However, in order to ensure a basic ability to provide predictions, training should at a minimum teach the basics of working with probability,[33] the fundamentals of proper calibration of one's own judgments,[34] as well as explain to participants, through demonstrations, how the Brier score (which is not entirely intuitive) works, and also how to approach questions using so-called Fermi estimation,[35] which can be supplemented, for example, with a short demonstration of the proper process of judgmental forecasting. [36]

At the end of the training, it is advisable to insert a mandatory test to check the participants' understanding of the content. It should be possible to repeat this test several times in case of a poor result, but it should be a mandatory condition for entering the tournament. As the entire training session is certainly longer than 30 minutes (in the case of the pilot implementation it took participants on average 1.5 hours), it is therefore advisable to ensure

---

[30]Chang, W., Chen, E., Mellers, B., Tetlock, P., Developing expert political judgment: the impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments, Judgment and Decision Making, Vol. 11, No. 5, September 2016, pp. 509-526,
https://www.sas.upenn.edu/~baron/journal/16/16511/jdm16511.pdf
[31]Mellers, B., Ungar , L., Baron J., (2014), Psychological strategies for winning a geopolitical forecasting tournament. Psychological science, 25(5), https://escholarship.org/uc/item/4rg4n9vr
[32]Muelhauser, L. (2020, September 22). *Efforts to Improve the Accuracy of Our Judgments and Forecasts*. Open Philanthropy.
https://www.openphilanthropy.org/blog/efforts-improve-accuracy-our-judgments-and-forecasts
[33]Confidence Calibration in a Multi-Year Geopolitical Forecasting Competition: https://pubsonline.informs.org/doi/suppl/10.1287/mnsc.2016.2525/suppl_file/mnsc.2016.2525-sm.pdf
[34]Calibrate Your Judgement:
https://www.guidedtrack.com/programs/icg4cze/users/sign_in?return_to=%2Fprograms%2Ficg4cze%2Fembed%2FyOQgZsLqDNY8ty4pwawnE0RgIKeq11XdapF_xXK3T1w%2Fcomplete
[35]The use of potential problems in the STEM disciplines to support the development of twenty-first century competencies:
https://www.researchgate.net/publication/334817167_The_use_and_potential_of_Fermi_problems_in_the_STEM_disciplines_to_support_the_development_of_twenty-first_century_competencies
[36]Czech priorities [OPTIONS - Czech priorities]. (2021, February 1). *OPTIONS - How to make predictions?* [Video]. YouTube. https://www.youtube.com/watch?v=sqJnc2RRCWY

that the progress made in the training session is saved. Open-source platforms such as Guided Track exist for facilitating such interactive tests, but it is also possible to handle this part using professional commercial tools.

Czech Priorities, 2021

All the teaching material produced within the OPTIONS project is publicly available and ready for use on the website: https://www.guidedtrack.com/programs/7hr567y/run

Czech Priorities, 2021

2. Tournament organization

2.1 Creating prediction questions

When designing the questions, it is essential to ensure that each question is carefully and clearly defined so that contestants can only interpret it in one way. If a question is not formulated precisely, each participant could interpret it differently and the resulting aggregate of predictions would be misleading and unhelpful. In such a case, it would not be possible to compare individual predictions with each other or to determine the best forecaster. With the potential for question evaluation to be ambiguous, the willingness of rational participants to engage in research on the topic also decreases.

Along with a clear definition of the question, it is essential to identify the source against which the question will be evaluated. The most appropriate sources are typically official reports or regularly updated statistics published on the websites of respected national and international institutions. If the question does not allow the use of institutional statistics with high credibility, the question can be evaluated on the basis of information from reputable news sites. The uncertainty associated with the evaluation of a question if no source for evaluation is identified may significantly reduce the willingness of participants to make predictions on such a question.

Each question must be accompanied by additional information that clearly identifies how the question will be evaluated in light of all possible future scenarios. If an event occurs that was not contemplated in the question, the question must be canceled. Such cases should be minimized as, in addition to the loss to the sponsor of not receiving the necessary prediction, such errors lead to a loss of motivation for the participants. However, such events can occur regardless of the extent of preparation.

It is important to proactively communicate with the end user of the prediction when creating questions to ensure that the question asked is actually useful.

2.1.1 Requirements for quality questions

A good prediction question should satisfy the following points:

- It is clearly defined so that it can only be interpreted in one specific way.

- Each of the type terms used is clearly defined in the additional information for the question. For example: conflict, assault, breach of convention, etc.

- The question considers all possible scenarios including unlikely variants such as marginal cases, ties, postponed/canceled events, etc.

- Time zones and dates are specified. If the time zone is ambiguous, it is clarified in the specification information.

- The question and its context are sufficiently interesting to motivate forecasters to make predictions, and explanations are provided for why the question is important or which institution may value the results of the predictions.

- It is accompanied by information on the source from which it will be evaluated. At the same time, the criteria for its evaluation are unambiguous based on evolving data sources.

- The wording of the question is tasteful; for example, the death of a particular person or group should not be predicted. In the case of public interest, the question could be rephrased, for example, "When will X no longer be a member of the Court?" and so on.

- The question is not defamatory and respects privacy (for example, it does not interfere with the personal lives of people who are not publicly known).

- Before the question is asked, an expert who has long been involved in the subject matter of the question is consulted. This point is especially true for questions whose results are given in a scale.

- It is not possible to influence the outcome of a given question by tournament participants, for example by voting in a poll on the topic or by technical research such as Google trends or automatically generated polls.

It is necessary to account for cases when there will be errors in the questions and it will be necessary to cancel them. For such situations, it is necessary to have alternative questions prepared. There may also be a situation where the event described in the question does not happen at all, for example: which runner will win the 100m sprint at the 2020 Summer Olympics?

> **Wrong wording of the question:**
>
> *How many people will be vaccinated against Covid-19 in the Czech Republic by 30 June 2021?*
>
> **Correct wording of the question (example from the pilot implementation):**
>
> *How many complete vaccinations against Covid-19 will be registered by the Ministry of Health of the Czech Republic as of 30 June 2021?*
>
> *A fully vaccinated individual is defined as one who has received all of the prescribed doses of Covid-19 vaccine for one person (the vaccinated individual has received as many doses of vaccine as required by the EMA authorization for that vaccine - https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/treatments-vaccines-covid-19-authorised-medicines). Data on the number of vaccinees as of 30 June 2021 will be taken from the Department of Health website: https://onemocneni-aktualne.mzcr.cz/vakcinace-cr on 1 July 2021 at 17:00 CET. If a part of the Czech population has been vaccinated with a vaccine type that is not approved by the European Medicines Agency (EMA) at the time of the assessment of the question on 1 July 2021, those vaccinated with this vaccine not approved by the EMA will not be counted in the total number of full immunizations against Covid-19. In the event that a coronavirus mutation against which the current vaccines do not immunize is discovered by the question evaluation date, those immunized*

Reflecting on the above example, it can be noted that the first formulation of the question may already seem straightforward, but on closer examination several questions arise: What does "vaccinated" mean? Does it mean the first dose or the second? What about single-dose vaccines? Do people who have been vaccinated abroad count? etc. Therefore, the question needs to be carefully thought through, rephrased and clarified.

The new version of the question and the clarifying information specify exactly what "being vaccinated" means. The revised question does not ask how many people will be vaccinated, but how many vaccinated people will be registered by the Ministry of Health, which clearly defines the conditions for evaluation. The reference to the European Medicines Agency again addresses the issue of single-dose vaccines. At the same time, additional information has been provided which seeks to include all possible additional circumstances and to give the issue the necessary context.

2.1.2 Types of questions

Distribution of types of prediction questions:

Binary question - a question that simply asks whether (and with what probability) a given event will or will not happen. For example:

- Will Peter Sagan win the Tour de France?

Multinomial with one outcome - a question with multiple answers, but only one of them is correct. For example:

- "Who will win the 2019 presidential election in Argentina?" with possible answers "Alberto Fernández", "Mauricio Macri", "Roberto Lavagna", "Someone else".

Multinomial with multiple possible outcomes - a question with multiple answers that can occur simultaneously. For example:

- "Which of the following films will win at least one Academy Award?" with answers "The Trial of the Chicago 7", "Nomadland", "Minari", "One Night in Miami", "Promising Young Woman".

Multinomial with ranked categorical evaluation - this is a multiple choice question where the order of the answers matters because the range of possible answers is divided among several "compartments". For example:

- "When will Iran launch another ballistic missile?" with answers "Before March 1, 2016", "Between March 1 and April 30, 2016 (inclusive)", "Between May 1 and June 30, 2016 (inclusive)", "Not before July 1, 2016".

- "What will be the final exchange rate of the US dollar against the renminbi on 1 January 2016?" with the answers "Less than 6.30", "Between 6.30 and 6.35 (inclusive)", "More than 6.35 and less than 6.40" and "6.40 or more".

- "How many seats will the Justice and Development Party win in Turkey's snap elections?" with the answers "Absolute majority", "Relative majority" and "Something else".

Questions with number or date ranges - some forecasting platforms allow you to ask questions where the forecaster can fill in a number or date (or their distribution). For example:

- How much will the winning bidder pay to go into space with Jeff Bezos?

- When will the FDA grant Novavax SARS-CoV-2 vaccine approval for emergency use?

In general, forecasters prefer binary questions to multinomial questions because they are cognitively simpler. If multinomial questions are used, it is recommended not to use too many options, a maximum of five.

2.2 Administration

The administration of the forecasting tournament itself can be summarized in the following steps:

1. Posting a question on the platform

2. Active monitoring of comments in the first days of the tournament

3. Clarifying, modifying or canceling questions, if necessary

4. Evaluating the question after the predicted event occurs

After each question is posted, comments should be actively monitored to verify that participants understand the question correctly. If participants find ambiguities in the question, additional clarifications can be issued. Most platforms provide the opportunity for participants to publicly "report" an ambiguity in a question to the tournament administrators, but clarifications can be provided without prior reporting. If participants find a more fundamental problem with a question, it is recommended (especially in the early stages of the tournament) that the question be canceled and reworded.

In case it is appropriate to cancel a question altogether, it is necessary to have several replacement questions prepared in advance. The first few days after the questions are asked can be very time consuming for the administrators, due to the high volume of comments which must be read, which in turn require the administrator to track down information and provide clarification. During the tournament itself, 1-2 hours per day by one team member may be enough, unless it is necessary to provide clarification.

The final termination of the tournament should occur immediately after the occurrence of the event that is required to evaluate the question. The administrator fills in the correct answer on the forecasting platform and thus completes the prediction of the question

(unless the prediction itself has already been completed), which on most platforms will automatically calculate the scores for the participants, send a notification email, and adjust the ranking accordingly according to the overall score.

2.3 Cost of tournaments

The financial complexity of facilitating a forecasting tournament is highly dependent on the goals of the organizer, which determines the necessary scope of recruitment, duration of the tournament, number of questions, and other basic parameters. In general, however, a slightly higher financial intensity can be expected than using traditional qualitative and quantitative methods, mainly due to the novelty of this approach.

For the initial facilitation of a tournament on a scale similar to the pilot implementation mentioned above (1.5-year project, creation of the Czech language version of the platform, hundreds of participants, 8-week tournament, three-month evaluation and over 200,000 CZK in financial rewards), in addition to the costs of financial rewards and platform provision, it is necessary to count on the capacity of at least 3 full-time medium-senior experts for at least 10 months, and therefore a budget in the order of millions of CZK.

The use of these methodological procedures and the possible use of an already established group of proven forecasters, calibration training, the Czech language version of the platform and other shortcuts can significantly reduce the financial intensity. Furthermore, it can be expected that, as the application of judgmental forecasting methods becomes more common in the public and private sectors, competition will grow on the part of platform providers. Additionally, as the experience of facilitators grows, it will also be possible to reduce costs by sharing expertise and standardizing established and proven procedures.

In case of inappropriateness of in-house facilitation of a forecasting tournament (due to lack of internal capacity,time pressure, etc.), it is possible to outsource the forecasting tournament or other method of judgmental forecasting. In this case, it is advisable to use the rules mentioned in this methodology (such as expected work process, expected question types, expected timeframe) in the formulation of such a procurement.

Particularly for institutions of larger scale and importance, the most cost-effective option in the long term is to create its own internal unit to systematically deal with judgmental forecasting and the facilitation of forecasting tournaments. This step is recommended, for example, by the University of Pennsylvania research center Perry World House.[37] Because of the cross-sectoral nature of judgmental forecasting, the existence of such a unit can also increase efficiency in finding solutions to cross-sectoral problems and for transferring information across government bodies.

The impact of establishing such a unit and the use of judgmental forecasting in planning is very difficult to quantify and monetize, precisely because of the complexity of the problems to which these tools are usually deployed. But the high importance of strategic decisions that are usually informed by judgmental forecasting suggests the positive impacts and high cost-effectiveness of these methods.

---

[37]"Keeping Score: A New Approach to Geopolitical Forecasting," Perry World House, University of Pennsylvania, 2021.

3. Use of outputs

The practical applicability of the produced predictions is a crucial aspect of the forecasting tournament method, which should be kept in mind when applying it and especially when formulating the questions.

If the forecasting tournament is facilitated by the institution for its own use, it is essential to first come to a consensus within the organization - including leadership or management - on what questions it is interested in knowing the answers to and how exactly the resulting predictions and comments will be used in decision-making. Experience from abroad shows that this step, however crucial, is often omitted and the result is then not translated into an adequate response, for example because of a lack of interest on the part of the management of the institution or because of a lack of readiness to make changes in the decision-making processes. When formulating the questions, it is therefore advisable to have several alternative plans already prepared and written to adjust one's own activities and related decision-making according to the resulting prediction.

3.1 Internal use of outputs

If the questions have been asked correctly, the resulting aggregate of probabilistic predictions is very understandable and easy to present. However, in addition to the probabilistic estimates, the comments are also of significant value and typically need to be condensed into a multi-page summary for presentation to the board, management, or other decision makers within the institution. In the UK civil service, the so-called key drivers are used for this purpose and are identified as the most common components that individual forecasters mention in their decision-making and for or against which they define themselves.

In the case of predicting the number of people infected with COVID-19, the most common key indicators arguing for an increase in the number of infections may be, for example, the decline of social distancing or the spread of more infectious mutations of the virus, while indicators arguing against an increase may be, for example, high vaccination rates, the availability of respirators or more people outdoors in the summer months. These indicators need to be identified by an analytics team on the part of the tournament organizer; in the future, automated AI-based algorithms could take on this role.

When reporting outputs, it is also advisable to use consistent terminology in expressing probability and uncertainty. The UK civil service uses the so-called PHIA Yardstick for this purpose[38] , which divides the probability scale into 7 buckets, each with a corresponding verbal expression. When presenting the outputs, it is possible to use the Czech terminology that we used in the initial calibration training:

---

[38]"The Professional Development Framework," Professional Head of Intelligence Assessment, 2019.

almost no chance	very unlikely	unlikely	roughly even chance	likely	very likely	almost certain
remote	highly improbable	improbable	roughly even odds	probable	highly probable	nearly certain
01-05%	05-20%	20-45%	45-55%	55-80%	80-95%	95-99%

Czech Priorities, 2021

3.2 Cooperation with other institutions

If the forecasting tournament is facilitated by an organization other than the institution that is to be the end-user of the outputs (for example, a research or non-profit organization or a procurement contractor), it is essential to communicate with the end-user at all times.

In this case, early contact should be made with entities that may be interested in the resulting prediction (unless such institutions are already co-organisers or sponsors of the tournament). By providing additional information regarding the question, these entities can significantly influence the question's wording and thus its usefulness for their needs. Furthermore, it can be highly motivating for tournament participants to be informed in the additional information to the question that the resulting aggregate of the entire forecaster pool, including the best comments, will be provided to the institution.

During the pilot implementation, we worked with organizations at different levels of intensity:

- National Institute of Mental Health
- National Monitoring Center for Drugs and Addictions
- National Cyber Security Authority
- Interdisciplinary Epidemic Situations Group (MESES)
- Faculty of Social Sciences, Charles University
- Association for International Affairs (AMO)
- Association of Primary School Principals
- Czech Tourism
- Jobs.cz

In the first phase of cooperation with institutions, it is possible to send the subject this methodology, or a brief 1-3 A4 page summary, which describes the principles of judgmental forecasting and the ways in which the subject can ask a question (including an example of how such a question is formed and what it must contain) in general terms. Given the difficulty of formulating good questions, it is necessary to work intensively with the institution at this stage.

In the second phase of cooperation, it is a good idea to provide the institution with a set of already defined questions that may be of fundamental interest to them. This second phase is fundamentally more demanding and requires an understanding of the environment and strategic considerations under which the institution operates. We know from experience that not all questions defined in this way will be of interest to the institution and that questions which do appeal to the institution often need to be fundamentally redefined or

adapted due to resource constraints. Despite the significantly higher time commitment involved in creating 'tailor-made questions' for the institution, this method is significantly more efficient in terms of providing the desired information.

After the deadline for the evaluation of the question has passed, the final report on the question must be provided to the institution. It is recommended that the report be formulated with the specific needs of the institution in mind; some institutions may be particularly interested in raw data while others, for example, may want a summary of all the sources cited. It is advisable to have the final report prepared by an analyst on the organizer's side. Ideally, this analyst should monitor the tournament during throughout and reflect potentially interesting moments that occurred during the tournament in the final report.

The final report should include information such as:

1) The exact wording of the question asked

2) How many forecasters were involved in the prediction

3) The resulting aggregate prediction (and/or the prediction of a group of already identified outstanding forecasters)

4) How this prediction developed over time

5) How the resulting data can be interpreted

6) The best comments added to each prediction

Here is an example of such a report:

Forecasting tournament outputs

March – April 2021, www.predikce.org

„Will a Czech teaching hospital or university at any time in April, May or June 2021 become the target of a successful ransomware attack, which will be reported by the media?"

Ransomware is a type of malicious program that blocks a computer system or encrypts the data written on it, and then demands a ransom from the victim for restored access. Some forms of ransomware encrypt files on the hard drive (cryptoviral blackmail) others just lock the system and a threatening message tries to force the user to pay. The list of universities includes all universities in the Czech Republic - state, public and private. The list of teaching hospitals includes all teaching hospitals in the Czech Republic: teaching hospital bulovka, teaching hospital brno, teaching hospital u sv. Anny in Brno, University Hospital Hradec Králové, University Hospital Královské vinohrady, University Hospital in Motol, University Hospital Olomouc, University Hospital Ostrava, University Hospital Pilsen,

Czech Priorities, 2021

Beyond this information, it is possible to distill the so-called "key indicators" based on the comments, the key themes most frequently mentioned by forecasters in their comments.

3.3 Further work with the results

Beyond the transmission of the results to the relevant institutions, it is also advisable to compare the results with the prediction of prominent experts,media personalities active in the field, or other single-person sources. If there is sufficient space, once the aggregated forecast has been produced, it can be supplemented with the opinion of analysts in the field to find out whether they agree or disagree with the estimate, as well as giving the reasons for their opinion.

It may also be appropriate to introduce calibration training or similar educational tools for policy makers and analysts in public administration so they can better understand the resulting predictions. In addition, it is advisable to design an appropriate communication

system in the run-up to the tournament to effectively convey information to policy makers without creating false confidence. In particular, the aim should be to correctly present the probabilities and their justifications, while also highlighting the accuracy that past forecasting tournament applications have achieved as well as their limitations.

Bibliography

Bickel, E. J., "Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules", Decision Analysis, 2007, č. 4, 2, str. 49-65.

Brier, Glenn W. "Verification of forecasts expressed in terms of probability." Monthly weather review 78, no. 1 (1950): 1-3.

Budescu, D. V. & Chen, E. Identifying expertise to extract the wisdom of crowds. Manage. Sci. 61, 267-280 (2014).

Chang, W., Chen, E., Mellers, B., Tetlock, P., Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments, Judgment and Decision Making, Vol. 11, No. 5, September 2016, pp. 509–526, https://www.sas.upenn.edu/~baron/journal/16/16511/jdm16511.pdf

Chang, Welton, Eva Chen, Barbara Mellers, and Philip Tetlock. "Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments." Judgment & Decision Making 11, 5 (2016).

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. Management Science, 9 (3), 458-467.

Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. Judgment and Decision Making, 6, 73–88.

Galton, F. Vox populi. Nature 75, 450-451 (1907).

Good Judgment Inc. (2021, October 19). *See the future sooner with Superforecasting*. Good Judgment. https://goodjudgment.com/.

Hanson, Robin, "Designing Real Terrorism Futures" Public Choice (2006), https://www.jstor.org/stable/30026644

Horowitz M., et. al, "Keeping Score: A New Approach to Geopolitical Forecasting", Perry World House (2021), available at https://global.upenn.edu/sites/default/files/perry-world-house/Keeping%20Score%20Forecasting%20White%20Paper.pdf (accessed 2. 11. 2021).

Jason, D., Atanasov, P., Tetlock, P., Mellers, B. (2019). Are markets more accurate than polls?, Judgment and Decision Making 14, no. 2 (2019): 135-147.

Karger, Ezra and Monrad, Joshua and Mellers, Barbara and Tetlock, Philip, "Reciprocal Scoring: A Method for Forecasting Unanswerable Questions" (31. 10. 2021), pre-print version, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3954498

Kleňha, Jan, "Improving strategic foresight with the use of forecasting tournaments and its implications for the study of international relations", Stosunki międzynarodowe - Inter. Relations (21.1.2022), available at https://internationalrelations-publishing.org/articles/2-4.

Law, T. (2020, June 11). *"Superforecasters" Are Making Eerily Accurate Predictions About COVID-19. Our Leaders Could Learn From Their Approach*. Time. https://time.com/5848271/superforecasters-covid-19/

Mellers, B., Ungar , L., Baron J., (2014), Psychological strategies for winning a geopolitical forecasting tournament. Psychological science, 25(5), https://escholarship.org/uc/item/4rg4n9vr

Mellers, Barbara, et. al, and "Identifying and Cultivating Superforecasters" *Perspectives on Psychological Science 2015*, Vol. 10(3); Mellers, Barbara, Philip Tetlock, et al, "Forecasting tournaments, epistemic humility and attitude depolarization." *Cognition 188* (2019): 19-26.

Morgan, M. G. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc. Natl Acad. Sci. USA 111, 7176-7184 (2014).

Muelhauser, L. (2020, September 22). *Efforts to Improve the Accuracy of Our Judgments and Forecasts*. Open Philanthropy. https://www.openphilanthropy.org/blog/efforts-improve-accuracy-our-judgments-and-forecasts

Page, M. (2021, March 23). *CSET Foretell Blog - Crowd Outperforms Projections from Historical Data in Early Results*. CSET Foretell Blog. https://www.cset-foretell.com/blog/crowd-performance-analysis

Page, Michael and Barker, Alex, "Forecasting Conflict in the South China Sea" CSET-Foretell (30. 10. 2020), available at https://www.cset-foretell.com/blog/forecasts-south-china-sea.

Page, Michael, "Crowd Outperforms Projections from Historical Data in Early Results" CSET-Foretell (23. 3. 2021), available at https://www.cset-foretell.com /blog/crowd-performance-analysis.

Scoblic, J. Peter, and Philip E. Tetlock. "A better crystal ball: The right way to think about the future." Foreign Affairs 99 (2020): 10, https://www.foreignaffairs.com/articles/united-states/2020-10-13/better-crystal-ball (accessed 2. 11. 2021).

Siegel, Adam. "Tracking the Outcome of Strategic Questions with Crowd Forecasting", Cultivate Labs blog (2021), https://www.cultivatelabs.com/posts/tracking-the-outcome-of-strategic-questions-with- crowd-forecasting (accessed 2. 11. 2021).

Social Science Prediction Platform - An interview with Stefano DellaVigna, UC Berkeley Social Science Matrix (15. 9. 2020), available at https://live-ssmatrix.pantheon. berkeley.edu/research-article/social-science-prediction-platform/.

Sunstein, C. Infotopia: How Many Minds Produce Knowledge (Oxford University Press, USA, 2006).

Surowiecki, J. The Wisdom of Crowds (Anchor, 2005).

Tetlock, P., Mellers, B., et al. (2014). Forecasting tournaments: Tools for increasing transparency and improving the quality of debate, Current Directions in Psychological Science, 23(4), 290-295.

Waldvogel F., Rüth, A., Siahaan, K. (n.d.). *ANTICIPATION HUB*. Forecast-Based Financing. https://www.forecast-based-financing.org/anticipation-hub/

Other inspiration

Anticipation Hub, 2021, available from: https://www.forecast-based-financing.org/anticipation -hub/.

CSET Foretell, 2021, dostupné z: https://www.cset-foretell.com/issue-campaigns.

Cultivate Labs case studies, 2021: dostupné z: https://www.cultivatelabs.com/case-studies.

Drew Armstrong, Fiona Rutherford, 2021: *CDC recruits outsiders to lead a new center on disease forecasting*, in: fortune.com, available from: https://fortune.com/ 2021/08/18/cdc-center-forecasting-and-outbreak-analytics-public-health/.

European Covid-19 Forecast Hub, 2021, dostupné z: https://covid19forecasthub.eu /reports.html.

Foretold, 2020, UK Policy Question, available from: https://www.foretold.io/c/47ff5c49- 9c20-4f3d-bd57-1897c35cd42d.

Good Judgment Open, 2021, available from: https://www.gjopen.com/.

Hypermind, 2019, available from: https://www.hypermind.com/en/case-studies/.

IARPA ACE, 2015, Aggregative Contingent Estimation, available from: https://www.iarpa .gov/index.php/research-programs/ace.

Matt Scott Frank, 2021: *May 2021 Election Violence Risk Briefing*, in: medium.com, available from: https://medium.com/the-die-is-forecast/may-2021-election-violence-risk-brief ing-cd57130fb8bb.

Metaculus, 2021, dostupné z: https://www.metaculus.com/tournament/vdh/ více informací: Gaia Dempsey, 2021, *Metaculus Announces Forecasting Causes*, in: metaculus.medium.com, available from: https://metaculus.medium.com/metaculus-announces- forecasting-causes-776473f42e38.

Replication Markets, 2020, available from: https://www.replicationmarkets.com/.