Óbuda University

Booklet of PhD thesis

The use of extreme value statistics to develop new metrics for risk assessment in diabetes and trial data analysis

by

Mátyás Szigeti

Supervisor:

Tamás Ferenci

Doctoral Schoold of

Applied Informatics and

Applied Mathematics

Budapest, 2023

1 Introduction

Statistical analysis forms a fundamental part of contemporary medical research. Similarly to other disciplines, these methods are constantly evolving and facing new challenges. My dissertation provides two examples.

The primary part presents the development of new metrics to assess the risks of patients with type 1 diabetes based on a novel approach utilizing extreme value statistics (EVS). This is a lesser known branch of statistics focusing on the distributions of rare observations with unusually low or high values. This could find applications in medicine [1], where extremes naturally have an important role, as to some extent diseases are non-normal conditions leading to some biomarkers reaching abnormal levels.

Despite that, to my best knowledge, there were only two examples of the use of EVS methods in clinical medicine prior this study: an analysis on cholesterol levels [2] and a study on pneumonia and influenza deaths [3], both were published relatively recently.

In contrast, in other fields, where – similarly to medicine – such rare, extreme events have an overwhelming impact [4] there is widespread use of EVS, for example in architecture [5, 6, 7], weather and climate analysis [8, 9, 10] or in sports or finance statistics [11, 12, 13].

The lack of sufficient data is a serious barrier in the wider application of EVS in biomedical field, as in most cases, important biomarkers are those that require some sample taken which is typically followed by a complex and expensive process to analyse that, thus limiting the number of samples. Additionally, because of the rarity of the extreme events, they form just a fraction of the total sample, thus have much smaller effective sample size, therefore their analysis could be exceptionally difficult.

Advances in measurement technology make diabetology an exception. With the widespread availability of continuous glucose monitoring (CGM), high-frequency (typically 5 minutes sampling time) and longer-term (up to weeks or months, even in routine clinical practice) measurements became possible and gone into clinical practice relatively long time ago [14, 15].

The current practice of the assessment of continuous glucose monitoring results uses several indicators which enable the rapid evaluation of the CGM measurements collected for several days or weeks [16, 17, 18, 19, 20, 21]. These metrics, however, mostly focus on overall variability, not specifically on extremities which is not necessarily the same. A patient's variability can be very high, even if the blood glucose

level is never in the extreme range, or the opposite could also occur (although this is highly unlikely), with the patient spending a lot of time in extreme range with very low variability. Thus, these metrics are not really appropriate to capture this aspect and far less the associated (hidden) risks of high glucose levels (hyperglycaemia). Even those metrics that do account for extremities (such as time spent above range for hyperglycaemia) are usually very simply – and ad hoc – indicators mathematically speaking, which do not incorporate the statistical knowledge on the behavior of extremities.

In contrast to these metrics, EVS allows the estimation of the probability that a future measurement exceeds a certain threshold (which is the relevant factor for hyperglycaemia), even if such value was never observed in the sample. This can be used to calculate the probability that the patient's blood glucose will be above a threshold in a given time span (e.g., in 1 year) and directly give the expected time spent above the threshold in the interval too. Taken together, these raise the possibility that metrics based on EVS are more useful to accurately characterise the diabetes related risks. This approach is based on a more sophisticated statistical foundation, addressing the extremes directly [11, 3, 22, 23].

2 Objectives

The main objective of this study was to develop novel, more accurate, easily understandable metrics with more robust mathematical foundation, based on EVS approach that focuses on the extremes of glucose measurements instead of the traditional metrics used in current practice for the assessment of CGM data. In order to do this, one of the first steps was the examination the theoretical background, and the assessment and comparison of the two main EVS methods, namely the "the peak over threshold" and the "block maxima" approaches in terms of applicability to the problem.

After these preliminary investigations, the applicability of the novel EVS based metrics was elaborated and demonstrated using the large sample of the REPLACE-BG trial containing 14.8 million measurements of 226 patients with type I diabetes. To my best knowledge, this is the first such application of an EVS based analysis, and it allowed the comprehensive comparison of the traditional and the novel metrics.

Using these results, patient level risk assessment was carried out by assessing the probability of reaching, and the time spent above clinically important levels of glucose. In addition to determining the correlation between the metrics, a further

objective was to identify and compare the group of patients with the highest risk according to the EVS analysis with the group with the highest risk according to the traditional metrics. The analysis was also planned to be extended to more comprehensive analyses with the use of non-stationary models, which made it possible to add patient level baseline clinical characteristics to the model in order to estimate their effects on the glucose level. Furthermore the analysis was extended to the validation of the data as well and the question of statistical independence was investigated with regards to the time-series nature of the CGM measurements.

Apart from the EVS related analysis presented in detail, the statistical aspects of another, advanced regression modelling of clinical trial data was presented. In this case, the statistical independence also played an important role and was the motivation for need of complex methods. In that part, these statistical aspects of primary analysis of the RECITAL trial were presented which was a contemporary, multi-centered, randomised, controlled trial.

3 Methods of investigation

The above mentioned two main approaches of EVS have many similarities but the most important is that they both use a secondary sample that is taken from the raw data which gets analysed and their names reflect the way this secondary sampling takes place.

Namely, the Peak Over Threshold (POT) approach uses a chosen cut-off level and takes only the observed values over that level as the secondary sample, while the block maxima (BM) approach splits the data to equal sized, non overlapping blocks of observations and takes the maximum (or minimum) of each block as the secondary sample to analyse.

Historically the BM approach was discovered and used first when the behaviour of extreme values was formally described, initially by Ronald Fisher and Leonard Henry Caleb Tippett in 1928 [24]. Their findings were later proven by Boris Vladimirovich Gnedenko in 1943 [25]. Together these form the so-called Fisher–Tippett–Gnedenko theorem which establishes that if there are constants with which the maximum of independent and identically-distributed random variables can be linearly transformed so that this renormalized variable converges to a non-degenerate distribution, then this distribution must be one of the following:

$$F(x) = \begin{cases} \exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-1/\xi}\right] & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}\right] & \text{if } \xi = 0 \end{cases}$$

Here $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ is the scale and $\xi \in \mathbb{R}$ is the shape parameter. Fortunately these distributions are closed under linear transformation, so instead of saying that the transformed maximum converges, we can say with the above formulation that the maximum converges to one of the above distributions without loss of generality, as location and scale parameters have to be estimated from the sample anyway.

The distribution presented above is called the Generalized Extreme Value (GEV) distribution. It covers three special cases based on the value of ξ, which are the following:

1. Frechet distribution ($\xi > 0$):

$$F_\alpha(x) = \exp\left\{-\left(\frac{x-u}{\sigma}\right)^{-\alpha}\right\}$$

2. Weibull distribution ($\xi < 0$):

$$F(x) = \exp\left\{-\left(-\left(\frac{x-u}{\sigma}\right)^{\alpha}\right)\right\}$$

3. Gumbel distribution ($\xi = 0$):

$$F(x) = \exp\left\{-\exp\left(\frac{x-u}{\sigma}\right)\right\}$$

While this formulation is suitable for analysis from the probability theory point of view, in statistical investigation the maxima has to be estimated. A sample is needed, so we can't simply take the maximum of the whole series, and that's where the BM approach first appeared, i.e., block maxima were used to capture the distribution of maxima.

The GEV distribution is described by three parameters: shape, scale and location. When we fit a model to the available empirical data (sample) the values of these parameters – and its uncertainty – has to be estimated. Several statistical methods can be used, but most commonly the maximum likelihood estimator (MLE), the L-moments and the Bayesian method are in used in practice[26, 27].

MLE is an often-used method due to its relatively good properties and its simplicity, and can be used for large data sets where other, more computationally intensive methods are not feasible. The L-moment method is based on the linear combinations of probability weighted moments, while the Bayesian approach utilizes the well-known Bayes theorem, to obtain the posterior distribution of the parameters using the information in the sample and a prior distribution.

Using the other main approach, the secondary sample of the Peak Over Threshold (POT) method asymptotically follows a so-called Generalised Pareto distribution (GPD) that was first introduced by James Pickands III half a century after Fisher and Tippett's work, in 1975 [28]. He has shown that that the behavior of these extreme values after the POT re-sampling follows the following probability law:

$$P(X - u < y | X > u) \approx 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi},$$

where ξ is the shape and σ is the scale parameter.

4 New scientific results

4.1 Thesis group 1

Thesis 1.1 I developed a novel approach that focuses on the maximums of Continuous Glucose Monitoring (CGM) measurements rather than the variability used by the traditional metrics of diabetology with the use of extreme value statistics (EVS). I have shown using a simulated dataset that the EVS with the peak over threshold (POT) approach can be used to characterise CGM curves and produce clinically relevant metrics to describe patient level risks by estimating the probability for a new peak to exceed a certain threshold, however, the choice of the threshold might be problematic.

Related publication: [29].

Thesis 1.2 I used, for the first time, the block maxima (BM) approach of EVS to characterise CGM curves. I have proven that it can also provide clinically relevant estimates that can be used as metrics to assess patient level risks and have many advantages over POT method. The estimated probability and the estimated time spent over certain, chosen thresholds can be calculated. These thresholds and estimates can be beyond the range of the CGM measurements. The analysis was conducted on real-life dataset.

Related publications: [29, 30].

Through this analysis, the concept of using EVS with BM approach to assess the quality of glucose control based on a CGM curve, thus the attributed risk to some extent, been successfully proven. However, it became clear that the BM approach has an important advantage over the POT approach and it can lead to clinically more relevant metrics. It can directly reflect the probabilities of reaching certain extremes and the amount of time spent above certain glucose levels for each patient even if it's beyond the observed range. POT enables us to calculate estimates for the probability of reaching certain extremes but does not take into account the extremes' occurrence in time, therefore it is less suitable for patient-level assessment.

4.2 Thesis group 2

Thesis 2.1 I applied the block maxima (BM) approach of EVS to a large sample of 226 patients from the REPLACE-BG clinical trial with CGM curves containing over 14.8 million observations. For the first time, EVS metrics were compared to widely used traditional metrics for patient level risk assessment of hyperglycaemia using real-life data. In general, a relatively weak or moderate correlation was found between the EVS and the traditional metrics.

Related publications: [29, 30, 31].

Thesis 2.2 The patients with the highest risk according to the new EVS metrics had only moderate scores according to the traditional metrics. A further investigation of these measurements have shown that these were heavily affected by saturation caused by the detection limit of the CGM sensor. Subsequent analysis shown that EVS metrics were more sensitive to simulated decrease of these saturation levels.

Related publication: [31].

Thesis 2.3 Similarly to regression type analyses, coefficients can be added to EVS models as well to investigate their effect on the modelled outcome. I investigated the effect of body mass index (BMI) on blood glucose maxima. A statistically significant effect was found with higher BMI being associated with higher values of hourly maxima of blood glucose levels.

Related publications: [31, 32].

The utilization of the data of REPLACE-BG which contains more than 14.8 million real-life CGM measurements was a huge boost in the effort to develop and prove the applicability of the EVS methods for such analysis. This analysis and its results were the first to my best knowledge where EVS was applied in diabetology, and on a particularly large dataset. The results of the new EVS metrics were compared with some of the most important and widely used traditional metrics used to assess CGM measurements and to provide patient level information. It was successfully demonstrated that EVS enables the characterisation of CGM measurements focusing on the more relevant extremes in terms of hyperglycaemia risk; this was used to create relevant and clinically easily interpretable patient-level summary metrics.

The validity of the CGM measurements were checked with a large number of available confirmatory blood glucose measurements with special attention to the upper detection limit which was proven to be an important limitation for the CGM sensor with measured values saturated at 400 mg/dl. An analysis was conducted simulating the effect of lower detection limits – thus lower saturation points – for the CGM measurements and in general the new EVS metrics were proven to be more sensitive to this effect compared to the traditional metrics.

Additionally, by utilising another type of analysis using the extreme value approach, is also possible to assess the impact of different individual clinical characteristics or treatments in a more precise and practical way using non-stationary models. This analysis have shown a statistically significant effect, with higher body mass index being associated with higher glucose maximums.

4.3 Thesis group 3

Thesis 3.1 I presented a mixed effects regression modelling strategy as a well-founded and more suitable solution than the possible alternatives to model a three-level, clustered, hierarchical data structure of a clinical trial. Through this, a reliable statistical analysis was conducted and published on a contemporary, multi-centered, randomised, controlled trial which was the first large scale study to assess the effectiveness of rituximab compared to cyclophosphamide for the treatment of interstitial lung disease associated with connective tissue disease.

Related publication: [33, 34].

This section focuses on the problems of statistical independence which was also investigated regarding the time-series data used for the EVS analysis but in a different setting. As in this clinical trial there were continuous outcomes with multiple measurements taken for each patient during the follow up and these could not be treated as independent measurement, ordinary regression models and statistical methods would not suitable for the analysis. A further twist is that these observations were also nested on a third level, by the patients' recruiting hospital (which might have even more important role as lung function outcomes were measured by spirometers), which made the use of other statistical methods commonly used for repeated measurements unadvised for this analysis. Further advantages of the presented approach is that it enables the use of all available data, including data with missing observations, and the handling unequal group sizes which is inevitable in real-life clinical settings. It also allows adding further covariates to the model. This adds more flexibility and robustness to the analysis which is exceptionally useful in a clinical trial setting where ideally the details of the analysis are planned and pre-specified upfront, with limited or no knowledge of the actual data. Besides it's advantages, the mathematical background and difference of mixed effects models compared to linear regression was also presented through an illustration, which provided a detailed explanation on how it is able to model these relationships when the assumptions of the independence of observations are not met.

5 Practical applicability of the results

Through the EVS and its block maxima approach, novel metrics were developed which are easily interpretable tools for patient level risk assessment of high

glucose levels as they can directly estimate the time spent above certain, clinically important thresholds, even if these were never attained in the sample. This is a mathematically better founded and more sophisticated approach than the – mostly simple and ad hoc – classical metrics, and, importantly, the patients identified as being of the highest risk were different according to the novel and the traditional metrics. It was noted that the detection limit of the sensors might have an important role as the measurements of the patients with the highest risk through the EVS metrics had been heavily affected by the upper detection limit of the sensor.

Furthermore, the analysis was extended with the assessment of the relationship between the body mass index and the hourly maximums of glucose measurements with the use of non-stationary models. This analysis have shown statstically significant association between higher body mass index and higher glucose maximums.

During this process, an error was discovered in the statistical software package used for the analysis which led to incorrect results in the calculation of one of the traditional blood glucose variability metrics. This had far-reaching consequences since this package is relatively widespread for the analysis of glycemic variability and this error affected the primary results of at least two published clinical trials. The author of the package has been contacted who confirmed these findings and fixed this error in a later version, and the authors of the two published trials have been also been notified.

The last part of this work contains another, different biomedical application of statistical modelling of clinical trials. Because of the repeated measurements and nested setting of the presented trial, it required the use of a three level, hierarchical, mixed effects regression model. This method's background, the alternatives and the reasons behind this choice were presented. The key issue regarding this question was statistical independence which also appeared in the EVS analysis but in a very different context. The presented work was the actual main analysis of a clinical trial, so amongst meeting the requirements of the relevant guidelines, most of the decisions regarding the analysis had to be set early on, with no or limited knowledge of the data, meaning substantial difficulties compared to other applications of statistics. Thus the flexibility and robustness of reasonable regression models played an important role in this matter. Ultimately, this analysis provided a well-funded, scientific evidence on the effectiveness and safety of a novel treatment compared to standard care for a potentially life-threatening lung disease.

References

[1] Stephen J Roberts. "Extreme value statistics for novelty detection in biomedical data processing". In: *IEE Proceedings-Science, Measurement and Technology* 147.6 (2000), pp. 363–367.

[2] P. Bermudez and Zilda Mendes. "Extreme Value Theory in Medical Sciences: Modeling Total High Cholesterol Levels". In: *Journal of statistical theory and practice* 6 (Sept. 2012), pp. 468–491. DOI: 10.1080/15598608.2012.695673.

[3] Maud Thomas, Magali Lemaitre, Mark L Wilson, Cécile Viboud, Youri Yordanov, Hans Wackernagel, and Fabrice Carrat. "Applications of extreme value theory in public health". In: *PloS one* 11.7 (2016), e0159312.

[4] RR Kinnison. *Applied extreme-value statistics*. Tech. rep. Pacific Northwest Lab., Richland, WA (USA), 1983.

[5] Michel K Ochi. "Principles of extreme value statistics and their application". In: *Paper of the Society of Naval Architects and Marine Engineers, SNAME, 1981* (1981).

[6] Erwan Le Roux, Guillaume Evin, Nicolas Eckert, Juliette Blanchet, and Samuel Morin. "Non-stationary extreme value analysis of ground snow loads in the French Alps: a comparison with building standards". In: *Natural Hazards and Earth System Sciences* 20.11 (2020), pp. 2961–2977.

[7] Jonathan Auerbach and Phyllis Wan. "Forecasting the urban skyline with extreme value theory". In: *International Journal of Forecasting* 36.3 (2020), pp. 814–828.

[8] Richard W Katz. "Statistics of extremes in climate change". In: *Climatic change* 100.1 (2010), pp. 71–76.

[9] Georgia Lazoglou and Christina Anagnostopoulou. "An overview of statistical methods for studying the extreme rainfalls in Mediterranean". In: *Multidisciplinary Digital Publishing Institute Proceedings* 1.5 (2017), p. 681.

[10] Erin Towler, Balaji Rajagopalan, Eric Gilleland, R Scott Summers, David Yates, and Richard W Katz. "Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory". In: *Water Resources Research* 46.11 (2010).

[11] Manfred Gilli et al. "An application of extreme value theory for measuring financial risk". In: *Computational Economics* 27.2 (2006), pp. 207–228.

[12] MB Adam and Jonathan Angus Tawn. "Modelling record times in sport with extreme value methods". In: *Malaysian Journal of Mathematical Sciences* 10.1 (2016), pp. 1–21.

[13] Serguei Y Novak. *Extreme value methods with applications to finance*. CRC Press, 2011.

[14] SF Clarke and JR Foster. "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus". In: *British journal of biomedical science* 69.2 (2012), pp. 83–93.

[15] Bruce W Bode. "Clinical utility of the continuous glucose monitoring system". In: *Diabetes Technology & Therapeutics* 2.1, Supplement 1 (2000), pp. 35–41.

[16] F John Service, George D Molnar, John W Rosevear, Eugene Ackerman, Lael C Gatewood, and William F Taylor. "Mean Amplitude of Glycemic Excursions, a Measure of Diabetic Instability". In: *Diabetes* 19.9 (1970), pp. 644–655. ISSN: 0012-1797. DOI: 10.2337/diab.19.9.644. eprint: https://diabetes.diabetesjournals.org/content/19/9/644.full.pdf. URL: https://diabetes.diabetesjournals.org/content/19/9/644.

[17] C.M. McDonnell, S.M. Donath, S.I. Vidmar, G.A. Werther, and F.J. Cameron. "A Novel Approach to Continuous Glucose Analysis Utilizing Glycemic Variation". In: *Diabetes Technology & Therapeutics* 7.2 (2005). PMID: 15857227, pp. 253–263. DOI: 10.1089/dia.2005.7.253. eprint: https://doi.org/10.1089/dia.2005.7.253. URL: https://doi.org/10.1089/dia.2005.7.253.

[18] T Battelino, T Danne, RM Bergenstal, SA Amiel, R Beck, T Biester, E Bosi, BA Buckingham, WT Cefalu, KL Close, et al. "Doyle FJ 3rd, Garg S, Grunberger G, Heller S, Heinemann L, Hirsch IB, Hovorka R, Jia W, Kordonouri O, Kovatchev B, Kowalski A, Laffel L, Levine B, Mayorov A, Mathieu C, Murphy HR, Nimri R, Nørgaard K, Parkin CG, Renard E, Rodbard D, Saboo B, Schatz D, Stoner K, Urakami T, Weinzimer SA, Phillip M. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range". In: *Diabetes Care* 42.8 (2019), p. 1593.

[19] Lalo Magni, Davide M Raimondo, Chiara Dalla Man, Marc Breton, Stephen Patek, Giuseppe De Nicolao, Claudio Cobelli, and Boris P Kovatchev. "Evaluating the efficacy of closed-loop glucose regulation via control-variability grid analysis". In: *Journal of diabetes science and technology* 2.4 (2008), pp. 630–635.

[20] David Rodbard. “Evaluating quality of glycemic control: graphical displays of hypo-and hyperglycemia, time in target range, and mean glucose”. In: *Journal of diabetes science and technology* 9.1 (2014), pp. 56–62.

[21] Michelle Nguyen, Julia Han, Elias K Spanakis, Boris P Kovatchev, and David C Klonoff. “A review of continuous glucose monitoring-based composite metrics for glycemic control”. In: *Diabetes technology & therapeutics* 22.8 (2020), pp. 613–622.

[22] Daniel Cooley, Douglas Nychka, and Philippe Naveau. “Bayesian spatial modeling of extreme precipitation return levels”. In: *Journal of the American Statistical Association* 102.479 (2007), pp. 824–840.

[23] Lee Fawcett and David Walshaw. “Estimating return levels from serially dependent extremes”. In: *Environmetrics* 23.3 (2012), pp. 272–283.

[24] R. A. Fisher and L. H. C. Tippett. “Limiting forms of the frequency distribution of the largest or smallest member of a sample”. en. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 24.2 (Apr. 1928), pp. 180–190. ISSN: 1469-8064, 0305-0041. DOI: 10.1017/S0305004100015681. (Visited on 02/09/2020).

[25] B. Gnedenko. “Sur La Distribution Limite Du Terme Maximum D’Une Serie Aleatoire”. In: *Annals of Mathematics* 44.3 (1943), pp. 423–453. ISSN: 0003486X. URL: http://www.jstor.org/stable/1968974.

[26] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events: for insurance and finance*. Vol. 33. Springer Science & Business Media, 2013.

[27] Laurens De Haan, Ana Ferreira, and Ana Ferreira. *Extreme value theory: an introduction*. Vol. 21. Springer, 2006.

[28] James Pickands III et al. “Statistical inference using extreme order statistics”. In: *the Annals of Statistics* 3.1 (1975), pp. 119–131.

Own publications related to the theses

[29] Mátyás Szigeti, Tamás Ferenci, and Levente Kovács. “The use of peak over threshold methods to characterise blood glucose curves”. In: *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. 2020, pp. 000199–000204. DOI: 10.1109/SACI49304.2020.9118838.

[30] Mátyás Szigeti, Tamás Ferenci, and Levente Kovács. "The use of block maxima method of extreme value statistics to characterise blood glucose curves". In: *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*. 2020, pp. 433–438. DOI: 10.1109/SoSE50414.2020.9130427.

[31] Mátyás Szigeti, Tamás Ferenci, and Levente Kovács. "The Use of Extreme Value Statistics to Characterize Blood Glucose Curves and Patient Level Risk Assessment of Patients With Type I Diabetes". In: *Journal of Diabetes Science and Technology* 17.2 (2023). PMID: 34814774, pp. 400–408. DOI: 10.1177/19322968211059547.

[32] Tamás Ferenci, Mátyás Szigeti, and Levente Kovács. "Using non-stationary extreme value analysis to characterize blood glucose curves". In: *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE. 2022, pp. 000171–000176. DOI: 10.1109/SAMI54271.2022.9780743.

[33] Toby M Maher, Veronica A Tudor, Peter Saunders, Michael A Gibbons, Sophie V Fletcher, Christopher P Denton, Rachel K Hoyles, Helen Parfrey, Elisabetta A Renzoni, Maria Kokosi, Matyas Szigeti, et al. "Rituximab versus intravenous cyclophosphamide in patients with connective tissue disease-associated interstitial lung disease in the UK (RECITAL): a double-blind, double-dummy, randomised, controlled, phase 2b trial". In: *The Lancet Respiratory Medicine* 11.1 (2023), pp. 45–54. ISSN: 2213-2600. DOI: https://doi.org/10.1016/S2213-2600(22)00359-9.

[34] Vicky Tsipouri, Peter Saunders, Greg J. Keir, Deborah Ashby, Sophie V. Fletcher, Michael Gibbons, Matyas Szigeti, Helen Parfrey, Elizabeth A. Renzoni, and Chris P. Denton. "Poster: Rituximab versus cyclophosphamide for the treatment of connective tissue disease associated interstitial lung disease (RECITAL): a randomised controlled trial". In: *4th International Clinical Trials Methodology Conference (ICTMC) and the 38th Annual Meeting of the Society for Clinical Trials*. Trials, May 2017, p. 130. DOI: 10.1186/s13063-017-1902-y.

Own publications not related to the theses

[35] Manjit S Gohel, Jocelyn Mora, Matyas Szigeti, David M Epstein, Francine Heatley, Andrew Bradbury, Richard Bulbulia, Nicky Cullum, Isaac Nyamekye,

Keith R Poskitt, et al. “Long-term clinical and cost-effectiveness of early endovenous ablation in venous ulceration: a randomized clinical trial”. In: *JAMA surgery* 155.12 (2020), pp. 1113–1121. DOI: 10.1001/jamasurg.2020.3845.

[36] Ping-Tee Tan, Suzie Cro, Eleanor Van Vogt, Matyas Szigeti, and Victoria R Cornelius. “A review of the use of controlled multiple imputation in randomised controlled trials with missing outcome data”. In: *BMC medical research methodology* 21.1 (2021), pp. 1–17. DOI: 10.1186/s12874-021-01261-6.

[37] Neil R Poulter, Christos Savopoulos, Aisha Anjum, Martha Apostolopoulou, Neil Chapman, Mary Cross, Emanuela Falaschetti, Spiros Fotiadis, Rebecca M James, Ilias Kanellos, Matyas Szigeti, et al. “Randomized crossover trial of the impact of morning or evening dosing of antihypertensive agents on 24-hour ambulatory blood pressure: the HARMONY trial”. In: *Hypertension* 72.4 (2018), pp. 870–873. DOI: 10.1161/HYPERTENSIONAHA.118.11101. URL: https://www.ahajournals.org/doi/abs/10.1161/HYPERTENSIONAHA.118.11101.

[38] Mátyás Szigeti, Levente Kovács, and Tamás Ferenci. “Stability of relative and absolute metrics: empirical evidence from pulmonology”. In: *2019 IEEE 17TH World Symposium on Applied Machine Intelligence and Informatics (SAMI 2019)*. 2019, pp. 235–238. DOI: 10.1109/SAMI.2019.8782769.

[39] Lívia Priyanka Elek, Matyas Szigeti, Berta Erdélyi-Hamza, Mátyás Szigeti, Konstantinos N Fountoulakis, and Xénia Gonda. “What you see is what you get? Association of belief in conspiracy theories and mental health during COVID-19.” In: *Neuropsychopharmacologia Hungarica* 24 (2022), pp. 42–55. ISSN: 1419-8711. URL: https://europepmc.org/article/med/35451591.

[40] Lívia Priyanka Elek, Mátyás Szigeti, Berta Erdélyi-Hamza, Nikolett Beáta Vadon, Konstantinos N. Fountoulakis, Daria Smirnova, and Xénia Gonda. “Association of lifestyle changes during the pandemic are associated depression and its distinct symptom clusters – consideration for prevention and intervention”. In: *Neuroscience Applied* (2023). ISSN: 2772-4085. DOI: 10.1016/j.nsa.2022.100818. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9789424/.

[41] Nikolett Beata Vadon, Livia Priyanka Elek, Matyas Szigeti, Berta Erdelyi-Hamza, Daria Smirnova, Konstantinos N Fountoulakis, and Xenia Gonda. “Association between Lifestyle-and Circadian Rhythm-Related Changes, and

Different Depression Symptom Clusters during COVID-19." In: *Psychiatria Danubina* 34.Suppl 8 (Sept. 2022), pp. 81–89. ISSN: 0353-5053. URL: https://europepmc.org/article/med/36170708.

[42] Péter Torzsa, László Kalabay, Dalma Dorottya Csatlós, Csenge Hargittay, Bernadett Márkus, András Mohos, Mátyás Szigeti, Tamás Ferenci, Verschoor Marjolein, Rozsnyai Zsofia, Gussekloo Jacobijn, K. E. Poortvliet Rosalinde, and Streit Sven. "A nagyon idős és esendő állapotú betegek antihipertenzív kezelési gyakorlata az alapellátásban". In: *Lege Artis Medicinae* 30 (2020), pp. 111–121. ISSN: 0866-4811. DOI: 10.33616/lam.30.011. URL: http://real-j.mtak.hu/14007/7/LAM_2020_03.pdf#page=30.

[43] Péter Torzsa, László Kalabay, Dalma Dorottya Csatlós, Csenge Hargittay, Bernadett Márkus, András Mohos, Mátyás Szigeti, Tamás Ferenci, Marjolein Verschoor, Zsófia Rozsnyai, Gussekloo Jacobijn, Rosalinde KE Poortvliet, and Sven Streit. "Antihipertenzív kezelés: A családorvosok inkább kezelik az esendő pácienseket". In: *Medical Tribune* 18 (2020), pp. 39–41. ISSN: 1589-1283.

[44] Tímea Vissi, Regina Szabó, Blanka Bágyi, Adél Göntér, Fanni Akkir, Mátyás Szigeti, Gabriella Erzsébet Papp, Éva Feketéné Szabó, and Anna Kelemen. *Cerebrális parézissel élő gyermekek számára készült diagnózis specifikus életminőség felmérő kérdőív (CPQOL) magyar nyelven történő alkalmazása*. 2018.

[45] Anita Zadori, Zsuzsanna Kis, Tibor Toth, Matyas Szigeti, Andras Temesvari, Geza Fontos, Noémi Nyolczas, and Peter Andreka. "Long-Term Efficacy and Safety of Left Atrial Appendage Closure Procedures". In: *International Heart Journal* 64.2 (2023), pp. 188–195. DOI: 10.1536/ihj.22-639.