Correlation investigations between HbA1c and blood glucose indicators on type 1 diabetic Hungarian children

T. Ferenci*, A. Körner ** and L. Kovács*

* Óbuda University, John von Neumann Faculty of Informatics, Physiological Control Group, Budapest, Hungary
** Semmelweis University, 1st Department of Paediatrics, Budapest, Hungary.
ferenci.tamas@phd.uni-obuda.hu, korner.anna@med.semmelweis-univ.hu, kovacs.levente@nik.uni-obuda.hu

***Abstract*—Technological advancements in diabetes has improved treatment possibilities. The big amount of data obtained from continuous glucose measurement (CGM) devices gave new investigation perspectives of blood glucose variation and other parameters connected to diabetic treatment. The aim of the paper is to investigate the relationship between age, duration of diabetes and sex of the patient with blood glucose levels measured with continuous glucose monitor. Blood glucose level was characterized with the relative time spent in hyper- and hypoglycaemia, central tendency (mean and median) and dispersion (standard deviation and interquartile range). These parameters were measured on *n=58* Hungarian children suffering from type 1 diabetes mellitus. The children were poorly controlled, with non-blinded continuous glucose monitor performed to optimize therapy. The univariate association between the aforementioned parameters was investigated with scatterplots (with non-parametric smoothing), linear and Spearman-ρ correlation coefficients. Multivariate structure was investigated with canonical correlation analysis. It was found that age and duration of diabetes was not significantly associated with any descriptor of blood glucose, but hemoglobin A1c (HbA1c) was positively associated with time spent in hyperglycaemia (but not with hypoglycaemia), with central tendency and with dispersion. The association with dispersion is likely a spurious relationship induced by the central tendency (as a confounder). We observed that the associations of the descriptors with HbA1c are non-monotone, with a break in the positive correlation at 10%. We hypothesize that this is caused by the poor compliance of the patients, i.e. those in worse state increasingly "cheat" during CGM measurements, that is, they pay more attention to their blood glucose levels than what is usually done when no CGM measurement is present; therefore breaking the expected positive, monotone association.**

I. Introduction

The normal blood glucose concentration level in the human body varies in a narrow range (4 - 6 mmol/l). If for some reason the human body is unable to control the normal glucose-insulin interaction (e.g. the glucose concentration level is constantly out of the above mentioned range), diabetes is diagnosed.

The newest statistics of the World Health Organization (WHO) predict that the diabetes population will be doubled in the next 20 years (from 2010 to 2030) [1], which alarms diabetic societies to research more intensively this disease.

New technological improvements in diabetes represented specially by continuous glucose monitoring (CGM) techniques allows a more precise follow-up of the blood glucose changes in the human body. By their 5 minutes measurement time (288 measurements in one day) CGM possibilities have direct effect on short term monitoring, but questions raised also on long term effects. The most important indicator in this sense is represented by Hemoglobin A1c (HbA1c) that reflects the average glucose level of a patient over 120 days (the "life cycle" of hemoglobin) [2].

Correlation of HbA1c and blood glucose is well known, but CGM possibilities opened new horizons in this investigation [3]-[5]. The aim of the current paper is connected to this topic and analysis relationship between age, duration of diabetes and sex of the patient with blood glucose levels on type 1 diabetes Hungarian children.

The paper is structured as follows. Section II presents the background of our measurements and the considered parameters, but also the statistical apparatus to be used. Section III highlights the results obtained that are discussed in Section IV. The paper ends with conclusions.

II. Materials and Methods

A. Patient Data

Investigated Hungarian children (*n=58*), suffering from type 1 diabetes mellitus (DM) were selected as a convenience sample from the 1st Department of Paediatrics of Semmelweis University (Budapest, Hungary) database, with continuous glucose monitoring (CGM) measurement data between July 2009 and February 2011.

Patients were poorly controlled, with CGM performed to optimize therapy. CGM measurements were non-blinded to the patients.

CGM recordings were electronically processed and the following parameters were extracted from the time series of blood glucose (BG) measurements:

- Time spent in hyperglycemia and hypoglycemia relative to the whole length of the recording (hypoglycaemia was defined as BG < 3.9 mmol/l, hyperglycaemia was defined as BG > 10 mmol/l) in [%].
- Central tendency of BG levels (quantified with mean and median BG) in [mmol/l].

978-1-4673-6400-3/13/$31.00 ©2013 IEEE

- Dispersion of BG levels (quantified with standard deviation and interquartile range of BG) in [mmol/l].

In addition to these, the following parameters were extracted for each patient from the electronic records of the hospital information system:

- Sex [male/female],
- Age [years],
- Time since onset of diabetes [years],
- HbA1C [%]

at the time of the CGM measurement.

Simple descriptive statistics of these parameters for the whole sample are presented in Table I.

B. Statistical Analysis

As already discussed, our aim was to analyze the connection between age, time since onset of diabetes, HbA1c and the already introduced descriptors of the CGM measurements.

First, we performed this analysis pairwise (i.e. we analyzed the connection between every possible (altogether $3 \cdot 6 = 18$ pairs). As every variable was quantitative, measured on scale, this was essentially a question of correlation, which was investigated with graphical (scatterplot) and analytical (calculation of correlation coefficients) methods. To allow for the possible non-linear relationships, we calculated the Spearman-ρ correlation coefficient (a rank correlation coefficient that measures monotone connections generally, not only linear connections [6]) in addition to the traditional (Pearson-) linear correlation coefficient [7]. Also to facilitate the detection of non-linear connections, we plotted the LOWESS-estimators [8] for the non-parametric regression between the investigated variables on the scatterplots. 95% confidence intervals for the regression will be also shown.

To perform a more subtle analysis that incorporates the internal multivariate structure of the two data sets, we also performed a Canonical Correlation Analysis (CCA).

TABLE I.
MOST IMPORTANT DESCRIPTIVE STATISTICS OF THE INVESTIGATED PARAMETERS IN MEAN (MEDIAN) ± SD (IQR) [MIN-MAX] FORMAT

Parameter	Descriptive statistics
Sex	37 female, 21 male
Age	12,3 (12,5) ± 3,7 (5,8) [4 – 18]
Duration of DM	5,03 (4) ± 3,53 (5,75) [0 – 18]
HbA1c	8,6 (8,2) ± 1,5 (1,8) [5.9 – 12.1]
Ratio of hypergly.	32,0 (30,8) ± 17,0 (21,4) [0.37 – 74.3]
Ratio of hypogly.	1,10 (0,50) ± 1,40 (1,30) [0.00 – 4.66]
Mean BG	8,85 (8,85) ± 1,28 (1,47) [5.97 – 12.0]
Median BG	8,47 (8,40) ± 1,42 (1,88) [4.70 – 11.7]
BG standard dev.	2,76 (2,81) ± 0,59 (0,62) [1.18 – 4.09]
BG IQR	3,84 (3,90) ± 1,00 (1,05) [1.40 – 6.65]

CCA aims to form linear combinations (called canonical variables) of the variables in two sets of data (in this case: descriptors of the BG measurements on the one hand, and age, duration of DM and HbA1c on the other hand) such that the correlation of the formed linear composites will be maximal [9]. (Further combinations can be formed by imposing a restriction that every canonical variable is uncorrelated with the previous canonical variables.)

By inspecting the coefficients of these linear combinations, it is possible to draw conclusions on the common factors behind the two sets of data [10]. To practically achieve this, usually the correlation between the original variables and the canonical variables (the so-called loadings) are visualized (typically for the first two canonical correlations) by plotting each original variable as a point on a two-dimensional plane, with the loadings being the coordinates.

To test the significance of the canonical correlations, we calculated the classical *F*-approximation of the Wilks-Λ test statistic [11], which provides sequential test for each canonical correlation.

Drawback of this approach is its asymptotical nature (i.e. either a priori known multivariate normality is needed or large sample so that central limit theorem becomes effective). To mitigate this, we also calculated the significance for the Wilks-Λ test statistic based on permutation test [12].

C. Programs Used

Statistical analysis and visualization was performed under the R statistical environment, version 1.15.1 [13]. Canonical correlation analysis was performed using the CCA library [14] and CCP library [15].

The R script developed for this purpose is available at the corresponding author on request.

III. RESULTS

Scatterplots from the pairwise analyses are shown on :

- Figure 1 (hyper- and hypoglycaemia)
- Figure 2 (central tendency of BG)
- Figure 3 (dispersion of BG).

The (asymmetric) correlation matrices between the two sets of variables are shown on:

- Table II (Pearson linear correlation coefficient)
- Table III. (Spearman-ρ correlation coefficient)

with *p*-values for the null hypothesis of no correlation.

Canonical correlations between the two sets of variables were ρ1=0.5593, ρ2=0.2572 and ρ3=0.1567. Unfortunately, p-value for the Wilks-Λ was 0.1445 (asymptotical), 0.1602 (permutation test) even for the first canonical correlation, so we were not able to detect any significant canonical correlation.

In the light of these, we used the already described graphical method purely descriptively, i.e. to perform a visualization of the data, the result of which is shown on Figure 4.

TABLE II.
ASYMMETRIC CORRELATION MATRIX OF THE INVESTIGATED PARAMETERS (LINEAR CORRELATION COEFFICIENT WITH SIGNIFICANCE)

		Age	DM length	HbA1C
Ratio of hypergly.	*r*	-0,05	-0,07	0,19
	p	0,6938	0,6215	0,1547
Ratio of hypoglyc.	*r*	0,00	0,04	-0,13
	p	0,9723	0,7613	0,3284
Mean BG	*r*	-0,05	-0,03	0,17
	p	0,7322	0,8308	0,2142
Median BG	*r*	-0,06	-0,01	0,11
	p	0,6624	0,9225	0,4166
BG std. deviation	*r*	0,08	-0,17	0,36
	p	0,5391	0,2090	0,0049
BG IQR	*r*	0,10	-0,08	0,42
	p	0,4776	0,5719	0,0012

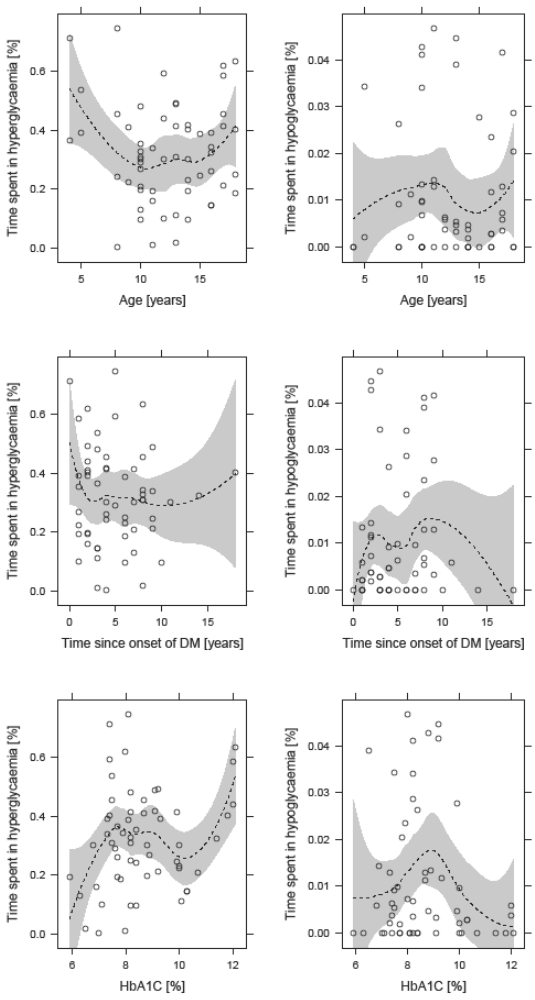

Figure 1. Scatterplots with LOWESS-estimators (dashed line) and 95% confidence intervals (shaded in grey) for the time spent in hyper- and hypoglycaemia vs. age, time since onset of DM and HbA1c.

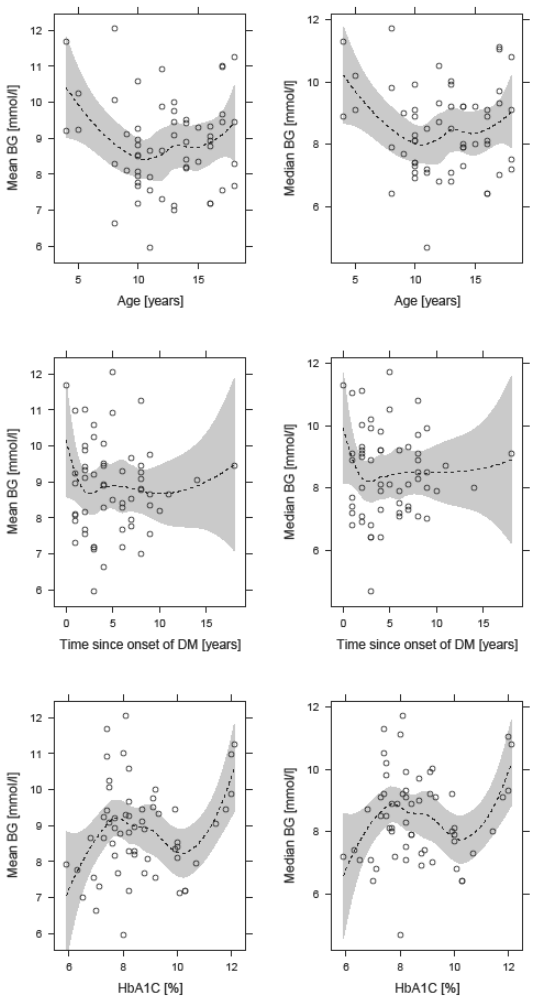

Figure 2. Scatterplots with LOWESS-estimators (dashed line) and 95% confidence intervals (shaded in grey) for the central tendency of BG (mean and median BG) vs. age, time since onset of DM and HbA1c.

TABLE III.
ASYMMETRIC CORRELATION MATRIX OF THE INVESTIGATED PARAMETERS (SPEARMAN-P CORRELATION COEFFICIENT WITH SIGNIFICANCE)

		Age	DM length	HbA1C
Ratio of hypergly.	ρ	0,02	-0,07	0,14
	p	0,9026	0,6152	0,3074
Ratio of hypoglyc.	ρ	0,02	0,08	-0,06
	p	0,8675	0,5272	0,6580
Mean BG	ρ	0,04	-0,03	0,11
	p	0,7935	0,8112	0,3951
Median BG	ρ	0,00	-0,02	0,04
	p	0,9915	0,9031	0,7561
BG std. deviation	ρ	0,05	-0,17	0,23
	p	0,7229	0,2125	0,0880
BG IQR	ρ	0,08	-0,06	0,32
	p	0,5643	0,6567	0,0138

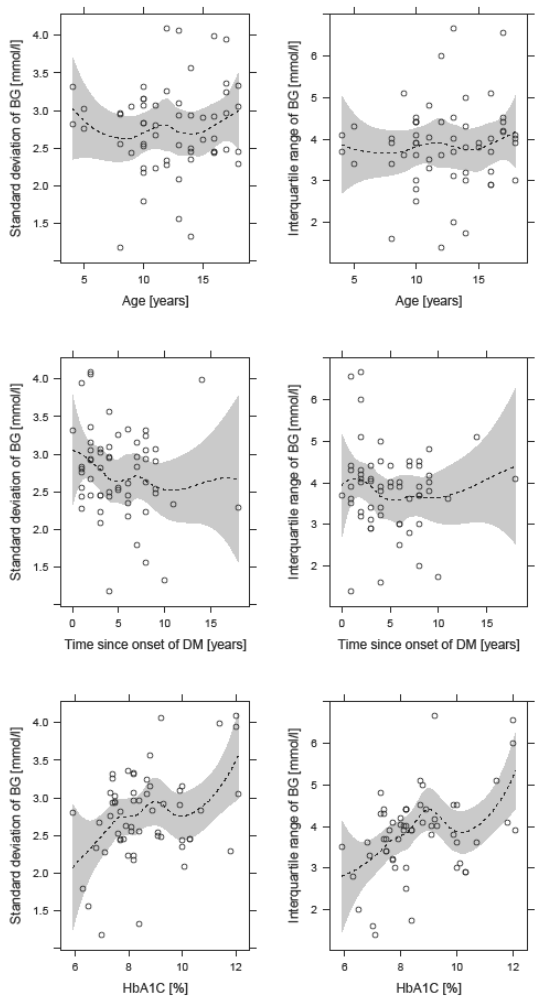

Figure 3. Scatterplots with LOWESS-estimators (dashed line) and 95% confidence intervals (shaded in grey) for the dispersion of BG (SD and IQR of BG) vs. age, time since onset of DM and HbA1c.

As far as the central tendency is concerned, roughly the same can be said about age and duration of DM, with the exception that both exhibit an initial decreasing phase (no matter whether mean or median is used to characterize central tendency), which seems to be more significant for age. Here HbA1c again shows a clear positive association both for mean and median. The paradoxical non-monote region around 10% shows up again.

The dispersion of BG shows no association with age and duration of DM (neither for standard deviation, nor for interquartile range). However, with HbA1c there is a clear positive association (both for standard deviation and interquartile range). The paradoxical temporary trend-change around 10% can be observed again.

Results of analytical investigations (i.e. the correlation coefficients) corroborate the above findings. Linear correlation coefficients are insignificant with the exception of BG standard deviation and BG interquartile range vs. HbA1c. This comes as no surprise given the fact that even in instances were association was observed on the scatterplots, it was mostly non-linear, save for this two.

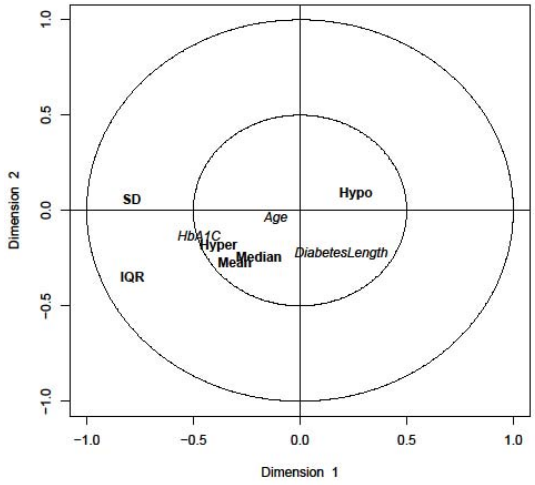

Figure 4. Visualization of the results of CCA: loadings for different investigated variables.

The application of Spearman-ρ correlation coefficients yields similar results (albeit the significance is worse even for these two cases, which can be explained by lower power of the more general estimator and the relatively large number of outliers).

Note, that now we performed 18 hypothesis testings in parallel (for each correlation coefficient), so results can not be compared to the traditional threshold of 5% due to the multiple comparisons situation [16]. However, significant correlations – especially for the linear correlation coefficient – had a magnitude smaller *p*-values, so the labeling of these as significant association is justified, even in the light of the multiple comparisons situation.

Canonical correlation analysis confirms the association of HbA1c with SD and IQR of BG measurements (i.e. the dispersion of BG), but also reveals a closer connection to the ratio spent in hyperglycaemia and the indicators of the central tendency of BG. It is very interesting to note that this phenomenon was already seen – to some extent – on the scatterplots, but was not appreciable from the correlation coefficients.

The age and duration of DM shows some association with the time spent in hyper- and hypoglycaemia, and also with the indicators of the central tendency of BG, but not with the indicators of the dispersion of BG measurements.

From these results, the positive association between the indicators of the central tendency of blood glucose and HbA1c is long known, even quantitatively [17]-[19]. The positive association between the indicators of the dispersion and HbA1c, however, might be surprising, given the physiological mechanism governing the forming of HbA1c (which is related to the level of BG not its dispersion).

This can be easily explained by noting that the correlation coefficient between mean BG and BG standard deviation is *0.56*, suggesting that the association between dispersion and HbA1C is simply a spurious relationship induced by the central tendency of BG (as a confounder).

The explanation for the seemingly paradoxical behavior of the association of HbA1C with the various descriptors of BG is possibly the poor compliance of the investigated children. The motivation to "cheat" during the CGM-measurements (i.e. to pay closer attention to BG levels than it is usually done by the children) is higher for those having worse metabolic state (as evidenced by HbA1C higher than 10%). Also, there is not only a motivation, but also a possibility for this, given that the measurements were non-blinded for the children. Hence, we hypothesize that children above 10% HbA1c start to increasingly "cheat" during CGM measurement, breaking the association between HbA1c and the mean (or median) glucose during CGM.

IV. CONCLUSION

Our results confirmed the well-known positive association of HbA1c with the indicators of the central tendency of blood glucose (mean, median), and demonstrated that hyperglycaemia (but not hypoglycaemia) is also positively associated with HbA1C. The positive association of HbA1C with the indicators of the dispersion of blood glucose is likely a spurious relationship induced by the mean blood glucose (as a confounder).

The relationship between HbA1c and the indicators of the central tendency of blood glucose are non-linear, which is likely caused by the varying compliance of the patients.

We found no association between the duration of the diabetes and the age of the patient with any indicator of the blood glucose levels.

However, other modeling techniques will be investigated similar to [20], [21], [22].

ACKNOWLEDGMENT

Levente Kovács is supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

[1] J. E. Shaw, R. A. Sicree and P. Z. Zimmet, "Global Estimates of the Prevalence of Diabetes for 2010 and 2030",. *Diab Res Clin Pract*, vol. 87, pp. 4-14, 2010.

[2] D. LeRoith, S. I. Taylor and J. M. Olefsky, *Diabetes Mellitus. A fundamental and clinical te*xt, 3rd ed, Lippincott Williams & Wilkins, 2004.

[3] J. K. Nielsen, C. H. Gravholt, C. B. Djurhuus, D. Brandt, J. Becker, L. Heinemnn and J. S. Christiansen, " Continuous Subcutaneous Glucose Monitoring Shows a Close Correlation between Mean Glucose and Time Spent in Hyperglycemia and Hemoglobin A1c", *J Diab Sci & Tech*, vol. 1 (6), 857-863, 2007.

[4] R. Mazze, "The Future of Self-Monitored Blood Glucose:Mean Blood Glucose Versus Glycosylated Hemoglobin", *Diab Tech & Therap*, vol. 10 (1), pp. S-93 – S-101, 2008.

[5] K. Sikaris, "The Correlation of Hemoglobin A1c to Blood Glucose", *J Diab Sci & Tech*, vol. 3 (3), pp. 429-438, 2009.

[6] M. Hollander and D. A. Wolfe, *Nonparametric statistical methods*, New York, Wiley, 1973.

[7] P. Armitage, G. Berry and J. N. S. Matthews, *Statistical Methods in Medical Research*, 4th ed, Blackwell, Malden, 2001.

[8] W. S. Cleveland, E. Grosse and W. M. Shyu, "Local regression models," in *Statistical Models in S*, J. M. Chambers and T. J. Hastie, Eds. Wadsworth & Brooks/Cole, 1992.

[9] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., Prentice Hall, 2007.

[10] A. Sherry and R. K. Henson, "Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer," *J Personality Assessment*, vol. 84, pp. 37-48, 2005.

[11] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley-Interscience, 3rd ed., 2003.

[12] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1994.

[13] R Core Team, R, *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.

[14] I. Gonzalez and S. Dejean, *CCA: Canonical correlation analysis*, R package version 1.2, 2012.

[15] U. Menzel, *CCP: Significance Tests for Canonical Correlation Analysis (CCA)*, R package version 1.1, 2012.

[16] R. G. Miller, *Simultaneous Statistical Inference*, 2nd ed, Springer, New York, 1981.

[17] D. M. Nathan, J. Kuenen, R. Borg, H. Zheng, D. Schoenfeld and R. J. Heine, "Translating the A1C Assay Into Estimated Average Glucose Values", *Diabetes Car*, vol. 31, pp. 1473-1478.

[18] K. Makris and L. Spanou, "Is There a Relationship between Mean Blood Glucose and Glycated Hemoglobin?", *Journal of Diabetes Science and Technology*, vol. 5, pp.1572-1583.

[19] D. B. Sacks, "Correlation between Hemoglobin A1c (HbA1c) and Average Blood Glucose: Can HbA1c Be Reported as Estimated Blood Glucose Concentration?", *Journal of Diabetes Science and Technology*, vol. 1, pp. 801-803.

[20] I. B. Wilson and P. D. Cleary, "Linking Clinical Variables With Health-Related Quality of Life. A Conceptual Model of Patient Outcomes", *Journal of the American Medical Association*, vol.273(1), pp. 59-65, 1995.

[21] M. Badida, R. Kralikova, and E. Lumnitzer, "Modeling and the use of simulation methods for the design of lighting systems", *Acta Polytechnica Hungarica*, vol. 8(2), pp. 91–102, 2011.

[22] K. C. Johnston, A. F. Connors Jr, D. P. Wagner, W. A. Knaus, X. Q. Wang and E. Clarke Haley Jr, "A Predictive Risk Model for Outcomes of Ischemic Stroke", *Stroke*, vol. 31, pp. 448-455.