BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

OPPORTUNITIES OF MICROSOFT EXCEL IN VISUALIZING ELECTRONIC HEALTH RECORDS AND CALCULATING MEDICAL DATA CORRELATION IN CLINICAL PRACTICE

Turdimuratov Baxtiyor Kurbonovich
Tashkent State Medical University, Termez Branch
Assistant at the Department of Social and Humanitarian Sciences
baxtiyorturdimuratov6691@gmail.com
Shamatov Faxriddin Uralovich
Tashkent State Medical University, Termez Branch
Assistant at the Department of Social and Humanitarian Sciences
faxriddinshamatov@gmail.com
Ibragimova Dildora Uktamjon kizi
Tashkent State Medical University, Termez Branch
Assistant at the Department of Social and Humanitarian Sciences
idildora9966@gmail.com

Istamova Durdona Rustam kizi

2nd-year student, Faculty of Medicine

istamovadurdona32@gmail.com

Abdullayev Yusufbek Fayzulla ugli

2nd-year student, Faculty of Medicine

yusufbek9203@gmail.com

https://doi.org/10.5281/zenodo.20321035

Abstract: This study investigates the practical capabilities of Microsoft Excel for primary data processing, visualization, and the determination of correlation links between parametric indicators from large-scale medical datasets obtained from Electronic Health Records (EHR) and Hospital Information Systems (HIS). Using clinical variables (patient age and systolic blood pressure) as a model, this paper demonstrates the algorithm for calculating the Pearson correlation coefficient (r) and creating correlation matrices within the Excel environment, as well as the methodology for visualizing these datasets using scatter plots and trendlines.

Keywords: medical informatics, electronic health record, Microsoft Excel, Pearson correlation, data visualization, scatter plot, clinical statistics.

INTRODUCTION

Modern medicine and pharmacy are entering the era of "Big Data" at an accelerating pace. Hospital Information Systems (HIS) and Electronic Health Records (EHR) generate vast amounts of parametric data every day, including laboratory test results, instrumental examination conclusions, and clinical metrics regarding the therapeutic efficacy of pharmaceuticals. However,

managing this data in its raw, unprocessed state poses significant challenges for clinicians and researchers when trying to establish accurate diagnoses or determine optimal treatment strategies.

A common challenge in clinical practice is the rapid and precise evaluation of the relationship strength between distinct medical factors and disease symptoms—such as the correlation between blood glucose levels and body mass index, or the link between patient age and the progression of arterial hypertension. In such scenarios, healthcare professionals are often expected to master complex, highly specialized, and expensive statistical software packages (e.g., SPSS, R-Studio), which can slow down the daily workflow of a practicing physician.

In reality, **Microsoft Excel**, a software application universally available on virtually every workplace desktop and possessing a relatively gentle learning curve, offers exceptional functionality for the primary analysis and correlation modeling of clinical data, serving as a highly pragmatic and accessible instrumental tool.

The objective of this study: To systematically elucidate the methodology for grouping parametric indicators from electronic health records, determining their linear correlation within the Microsoft Excel environment, and effectively visualizing the obtained results using clinical data models.

METHODS

The research methodology relies on the built-in statistical functions and graphic data-processing capabilities of Microsoft Excel. To demonstrate the practical execution of this methodology, a clinical dataset consisting of 20 conditional patients from a cardiology department was modeled, capturing their age (Variable X) and their corresponding systolic blood pressure metrics (Variable Y, measured in mmHg).

1. Correlation Analysis Methodology

To determine the strength and direction of the linear relationship between the two datasets, the **Pearson correlation coefficient** (r) was applied. In Microsoft Excel, this coefficient is computed using the dedicated =CORREL(array1, array2) function (or =КОРРЕЛ(массив1; массив2) in regional versions).

Mathematically, the coefficient is defined by the following formula:

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Where:

- The value of the correlation coefficient fluctuates within a strict boundary from -1 to $+1$.
- $r > 0$ indicates a positive (direct) linear relationship.
- $r < 0$ indicates a negative (inverse) linear relationship.
- $r = 0$ implies the total absence of a linear correlation. The statistical strength of the correlation was interpreted using the Chaddock scale (where $0.7 < |r| < 0.9$ represents a high correlation).

2. Visualization and Regression Modeling

To visualize the structured clinical data, a **Scatter Plot** was generated via the *"Insert"* $\rightarrow$ *"Charts"* menu in Excel. A linear trendline was superimposed onto the chart layout, and both the linear regression equation ($Y = kX + b$) and the coefficient of determination (R^2) were displayed directly on the plot area.

RESULTS

Once the clinical database was compiled into the Excel spreadsheet, descriptive statistics were initially run to capture the general characteristics of the variables. Subsequently, the exact correlation between the variables was analyzed.

The computational calculations executed via Excel's =CORREL() function yielded a Pearson correlation coefficient of **$r = 0.78$** between patient age and systolic blood pressure. According to the criteria of the Chaddock scale, this empirical metric establishes a **strong positive linear relationship** between the examined parameters. This indicates a statistically regular pattern where an increase in patient age corresponds with an elevation in systolic blood pressure.

For large-scale electronic health records containing multi-dimensional parameters (e.g., age, systolic blood pressure, heart rate, cholesterol levels, glucose), a comprehensive correlation matrix was generated simultaneously using the *"Data Analysis" $\rightarrow$ "Correlation"* toolpak in Excel (Table 1).

Table 1. Inter-correlation matrix of clinical parameters (Excel output)

Clinical Parameters	Patient Age	Systolic Bp	Cholesterol Level
Patient Age	1		
Systolic BP	**0.78**	1	
Cholesterol Level	0.62	0.55	1

The resulting matrix clearly indicates that serum cholesterol levels also maintain a moderate positive correlation with both patient age ($r = 0.62$) and systolic blood pressure ($r = 0.55$).

To visualize this linear relationship model, the constructed scatter plot and its mathematical trendline revealed a coefficient of determination of **$R^2 = 0.6084$**. This implies that approximately 60.8% of the variance observed in systolic blood pressure is directly attributable to patient age, while the remaining 39.2% of the variance is influenced by alternative internal or external clinical confounding variables (such as acute stress, genetic predisposition, dietary habits, and lifestyle factors).

DISCUSSION

In clinical trials and daily medical practice, archiving information solely within static digital tables is insufficient. By constructing dynamic correlation matrices within the Excel environment, a clinician can easily extract and discover "hidden clinical regularities" hidden within massive datasets. For instance, identifying a negative correlation ($r < 0$) between a specific pharmaceutical dosage and patient recovery time mathematically validates the therapeutic efficacy of that drug.

Furthermore, data visualization via scatter plots not only elevates the scientific rigor of clinical reports and academic papers but also enables clinicians to evaluate patient progress visually. The spatial proximity of data points to the regression trendline verifies the accuracy of the linear model, whereas extreme "outliers" (data points deviating sharply from the trendline) allow for the rapid identification of clinical anomalies or high-risk patient subgroups requiring immediate medical intervention.

Consequently, training future pharmacy and medical students to perform clinical data analytics using practical, everyday instrumental tools like Microsoft Excel—rather than relying solely on abstract theory—is a strategic necessity for modern medical education.

CONCLUSION

1. Microsoft Excel is a highly functional and universal instrumental application for clinicians, facilitating primary statistical processing, grouping, and correlation modeling of clinical variables derived from electronic health records.

2. The deployment of the =CORREL() function and multi-parametric correlation matrices significantly accelerates the process of grouping complex clinical traits and objectively evaluating their diagnostic relevance via standardized scales.

3. The application of scatter plots and automated trendlines allows for the effective visualization of clinical data dynamics, enabling healthcare researchers to forecast diagnostic outcomes using the coefficient of determination (R^2).

REFERENCES

1. Ganiyev, A. K., & Mamatqulova, N. X. (2024). *Textbook on Information Technologies in Medicine and Clinical Statistics*. Tashkent.

2. Fletcher, R. H., Fletcher, S. W., & Fletcher, G. S. (2014). *Clinical Epidemiology: The Essentials*. Lippincott Williams & Wilkins.

3. Microsoft Corporation. (2023). *Advanced Statistical Functions in Excel*. Retrieved from Microsoft Support.

4. Altman, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall.

5. Kurbonovich, T. B. et al. (2025). Digital Technologies in Medicine. Telemedicine. *IMRAS*, 8(12), 39-41.

6. Kurbonovich, T. B. et al. (2026). Problems in Modern Culture: Socio-Spiritual Analysis in the Context of Globalization and Digital Transformation. *Global Science Review*, 18(1), 183-188.

7. Panji o‘g‘li, C. O. et al. (2026). Regenerative Properties and Modern Histological Analysis of Connective Tissue in the Morphofunctional System of the Organism. *American Journal of Applied Medical Science*, 4(2), 230-235.

8. Turdimuratov, B. K. (2022). *Teaching Medical Sciences Using Innovative Methods and ICT*. Tashkent: Uzbekistan Medical Publishing House.

9. Kurbonovich, T. B., & Bahodirovich, B. B. (2026). Step-by-step acquisition of practical skills in studying information technologies in medicine. *Global Science Review*, 17(1), 203–209.

10. Kurbonovich, T. B., & Nurhayat, M. (2026). Compilation and steps of the medical situational issues algorithm. *American Journal of Applied Medical Science*, 4(2), 59–63.

11. Turdimurodov, B. K., et al. The essence of electronic textbooks in medical education. *European Journal of Humanities and Educational Advancements*, 3(4), 48–50.

12. Kurbonovich T. B. et al. USE OF ICT AND SOFTWARE IN TEACHING PRACTICAL MEDICAL EXERCISES //AMERICAN JOURNAL OF APPLIED MEDICAL SCIENCE. – 2025. – Т. 3. – №. 11. – С. 1-6.