The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 23, Pages 292-299

ICRETS 2023: International Conference on Research in Engineering, Technology and Science

Applicability of a Gage R&R Study on a Home Blood Glucose Meter

Zoltan Varadi
Óbuda University

Gyozo Attila Szilagyi
Óbuda University

Abstract: Home glucose meters have become widely accessible and, in addition to self-monitoring applications, are now commonly used to support biohacking. Accu-Chek Instant is an easy-to-use and cheaply available device, and thus, it is recommended to use in a known lifestyle change program. Participants and users do not require special knowledge, though have due or undue concerns about changes in blood glucose levels. The study aims to find the minimal magnitude of change in blood glucose level, i.e. the effective resolution of the fore-mentioned device. In addition, we are seeking applicability of the standard Gage R&R study method which is well known in the automotive industry. As Gage R&R requires to incorporate several variance factors into the study, we reproduced typical home applications and daily routines of laymen. Our results on measurement uncertainty turned to be in range with the accuracy specified by the manufacturer, and the Gage R&R method helped us to formulate answer to home user concerns on a layman's language.

Keywords: Gage R&R, Measurement uncertainty, Blood glucose meter, Biohacking, Accu-Chek instant

Introduction

Modelling a human being with the analogy of a combustion engine provides a simplified energy balance calculation, with food and drinks being the fuel of such an engine. Like the operation of the combustion machine, biochemical processes in the human body can be characterized by numerous measurable parameters. Physiological parameters of human bodies are typically defined with a range considered to be healthy, and the same parameter having a value outside of that range may be an indication of an unhealthy condition; very similarly to the technical parameters of the engine or its components which are considered to function acceptable if they fall within the specified range; and are defective if they are outside the specification limits (Horváth, 2023). The fuel system of a human combustion engine contains complex and interrelated biochemical reactions, which transform food intake to usable energy to various parts of the body. The energy is then conveyed by blood circulation in various forms like ketone or glucose, and the energy balance is accompanied by several types of molecules of the hormone system take part in or control the use of energy. The energy is stored in various reserves in the body for immediate availability, short term storage, or long-term reserve (Lakatos, 2020). Long-term reserves are for example the subcutaneous or visceral fat deposits, which, above a certain percentage, turns to obesity. Excess fat reserves are only built up if there is excess energy in the body over a significant period, though it is an oversimplification to say that reducing energy intake will solve obesity problems.

Biohacking Program and Its Questions

By adjusting the ratio of macro- and micro-nutrients in the intake to the load of the human engine, biohacking targets to alter one's biological setup towards an optimum, often doing that without medical supervision. Those do-it-yourself programs use recent technology of smart tools such as smart armbands or body analyzer scales, and traditional tools like a measuring tape, to measure the activities of the people in study, and the body's reaction to the physical activity: e.g. with immediate feedback, adjust intensity to pulse rate during sports; monitor blood oxygen saturation; or count the number of steps taken during a day; or measure and monitor waist/hip ratio. The personalized programs are said to help prevent from elderly degradation of health, and are promoted with warnings of excessive healthcare costs and risks. The effect of biohacking is a concurrent topic widely studied in various forms. Biohacking has made biotechnology widely known, and the evolution of home use equipments made it financially accessible to those who would otherwise wouldn't need them for medical observations. Medical professionals are worried as users without proper knowledge and education are risking misusing the diagnostic tools and arrive at inadequate conclusions and undue concerns. On the other hand, many patients are overwhelmed and are keen on keeping their health in their own hand (Zheng, 2021). This study is not intending to take position on either side of the worries but takes a more technical approach on understanding the measurement uncertainty that might influence decisions of participant in DIY biohacking.

In his book, Lakatos (2020) introduces diverse approaches for maintaining and recovering healthy conditions. His portfolio includes considerations on food composition, on activities and muscle conditions, on breathing, and many others. Training programs and guided courses are set up to help people learn the basics and understand their body biology to adjust nutrition and activities. Participants are warned to understand their own physical conditions and set up some home measurement routines. Measured parameters are body dimensions and composition on a weekly basis; breathing, activity and sleep indicators daily; and blood glucose levels before-after every meal resulting in 5-15 or even more blood tests per day. The huge number of blood glucose values helps participants understand how certain foods influence their energy reserve. Closed groups on social media platforms are set up to facilitate communication with and between members of the training program, where participants raise their questions and share their experiences. Target audience of such a program is people over the hill who are or want to be active though, most of them live with obesity and its consequences.

The Spring 2023 training program has started with 1144 participants, most of them living in Hungary. When subscribing, they were provided with a list of recommended tools; one of them was a glucometer, the recommended type is Accu-Chek Instant due to its simplicity and convenience in frequent use. The meter is packed with test stripes and a lancing device in a carry case; its outfit and quick start guide made it an ideal starting kit at an affordable price below €20 at that time. The group communication has not thoroughly been analyzed, though it is frequently seen that participants are concerned about their blood glucose measurement results. A typical question looks like: "My morning fasting level was 5.0, I haven't eaten anything yet, though it went up to 5.2 by now. What happened?" – cited from one question in the group discussions.

Glucometer Accuracy

Understanding measurement results of blood glucose levels helps diet and body shaping programs as well as medical treatments of diabetes. It is known that blood glucose levels are influenced by many factors including stress or activities, in addition to nutrition, and measurement conditions have significant effect on the readings, despite, diabetes patients must rely on the measurement results. In their article, Erbach et al. (2016) suggest that patients with diabetes as well as medical professionals in the treatment teams shall be well-informed about limitations and interferences. This may even be more crucial in case of DIY biohacking programs as they are lacking regular medical supervision of the participants' health conditions.

Various methods have been used to determine accuracy and reliability of glucometers. In their comparative study, Moore et al. (2021) have chosen young athletes on a ketogenic diet as sample and compared two types of field meters. With regards to the glucose testing, they found relatively small bias, and good agreement at fasting glucose level of 80 mg/dl (that is, 4.4 mmol/l). Their plot however shows wide variance, that is partly due to repeatability and partly due to differences in sample. Weak point of the comparative analysis is that they used a commercial field meter as a reference, not a gold standard laboratory method. The study concludes that those commercial meters are appropriate to detect nutritional ketosis conditions.

Accu-Chek instant accuracy has been studied in various ways. Both FDA and ISO have formulated standards that blood glucose meters shall meet. The ISO expectation for home use is that readings are 95 percent accurate within 15 percent of blood glucose equal to or above 100 mg/dl and are 95 percent accurate within 15 mg/dl for readings under 100mg/dl. As 1mg/l equals 0.0555mmol/l, the above means that an average healthy person's fasting glucose level of 5.6mmol/l needs to have a 0.83 mmol/l confidence interval at 95% confidence level. After two hours of eating a plate equivalent to 75g of carbohydrate, the glucose levels on an average healthy adult stands not above 7.8mmol/l, which is 2.2mmol/l higher than the fasting level. The expected confidence interval, which we consider as an upper bound for allowed measurement uncertainty, is consuming 37.5% of that change due to food, which is surprisingly far above the typically tolerated uncertainty budget of measurement systems used in automotive (VDA 5, 2021). The wider the range of uncertainty is, the more difficult it becomes for biohackers to fine-tune their diet with respect to effects of various food components. In addition to hygienic factors, i.e., washing hands before testing, testing conditions and tester conditions influence the reading. Typical sampling points are fingerprints, as the blood glucose level may change if sampled from other part of the body. The minimum amount of blood in the case of Accu-Chek Instant it is 0.6µl, according to its user manual. Environmental factors also contribute to the measurement reading, influencing its uncertainty (Smith, 2019), especially test strips are sensitive to handling related error sources.

Accuracy of the Accu-Chek Instant glucometer system is capable to measure blood glucose levels from 0.6mmol/l to 33.3mmol/l, with a readable resolution of 0.1mmol/l. The default target range is set to 3.9mmol/l to 8.9mmol/l (Roche, 2018). According to the documentation of test stripes (Roche, 2021), the system accuracy is within ±0.83mmol/l or 15% in the range of 2.2 to 26.7mmol/l, tested on 600 test stripes, which is in accordance with ISO inspection criteria (Freckmann et al., 2015). References of comparative analyses were gold standard laboratory methods. Detailed measurement uncertainties are given by the standard deviation of repeated measures (Table 1).

Table 1. Repeatability errors of Accu-Chek Instant (Roche, 2021)

Average level (mmol/)	2.6	4.7	6.6	7.6	12.0
Standard deviation (mmol/l)	0.1	0.1	0.2	0.2	0.3

Manufacturers' criteria as well as ISO or FDA standard criteria are used for pre-market evaluation of self-monitoring devices, though these conditions may not be applicable on field evaluations as there is limited number of available batches of test stripes and references. Ziegler et al. (2014) raise importance of determining accuracy in practical field application, though application circumstances are hard to standardize. Questions raised are: How do we define goodness of measurement? Is the measurement biased? What is the precision of measurement, i.e., is the repeatability error small? Is error influenced by reproducibility components? As bad decisions coming from measurement errors may have consequences on health conditions of people, it is very important to establish a concise and easy to implement method to evaluate glucometer accuracy.

In automotive industry, Gage R&R is a frequently used, mandatory tool to demonstrate that measurement processes are good enough for the respective control or inspection activities, and despite its simplicity and frequent use, Gage R&R is hardly used in health industries. This study is aiming to test how well the Gage R&R method can determine measurement uncertainty and what suggestions it may give to users of Accu-Chek Instant glucometer.

Method

The Gage R&R method (AIAG MSA, 2010) is suitable to determine applicability of a variable gage, like the Accu-Chek Instant self-monitoring glucometer. Amongst the various applications, the AIAG handbook recommends the ANOVA approach as it measures interactions of sources of measurement error, i.e., the part to operator interaction, and is flexible to incorporate other factors of measurement error into the model. The amount of calculation required to perform the analysis of variances (ANOVA) suggests using a computerized solution. Minitab Statistical Software v.21.4 has been chosen to perform the analysis, and all the charts below are generated with Minitab's Gage R&R function available under the menu of Stat → Quality Tools → Gage Study → Gage R&R Study (Crossed). We assume to set up an experimental design where a given sample's glucose level is considered constant throughout the repeated measurements, so repeatability is measurable. Methods and formulae, as well as the interpretation of statistics and graphs displayed are well described in Minitab's support (Minitab, 2023).

Setting up the Gage R&R Study

A crucial part of proper Gage R&R study is obtaining a sample that represents the realistic range of variation. In this case, we understood the concerns and methods of typical participants of the fore-mentioned biohacking program (Lakatos, 2020). The range of participants and measurement habits has been represented by 3 people aging between 45-61 years, a female and two males, with different but not unusual levels of obesity and some affectedness by IR. The typical range of measurable blood glucose level were simulated by multiple conditions of each people: morning fasting state; after a coffee; before meal and 90 minutes after meal. To keep the privacy of the medical conditions of sample, different people and conditions are not identified in the article below, every person and every condition of testing is identified as one sample, numbered from 1 to 10. Lancet sampling and testing was performed according to the user manual (Roche, 2018). Although the MSA manual (AIAG MSA, 2010) recommends using larger sample sizes for better confidence levels, automotive industry professionals used to stick with 10 samples as illustrated by the examples of the manual. So stayed we in our study at 10 samples.

Repeatability, in general, is understood as the same device is measuring the same workpiece, while circumstances are kept stable, however in this study it is practically impossible to repeat the same measurement on the same piece as test stripes are single use by design. Knowing that test stripes are well analyzed during manufacturing processes (Smith, 2019), we excluded the variance factor of test stripes by choosing to use the same batch of stripes throughout the whole experiment. Test stripe data are as follows: Lot id: 302037-04015630066810, Ref: 07819382053, best before date is May 15, 2024. In addition, we had to make sure that when a drop of blood is gained from fingerprint, it comes in sufficient volume to enable multiple tests within a short time. Knowing that measuring blood glucose level is very sensitive, we set up a little practice with the study participants to make sure we are able to do the 9 repeats within the shortest period of time. Processing one sample with 3 repeats in the 3 different testers is illustrated below (Figure 1.):

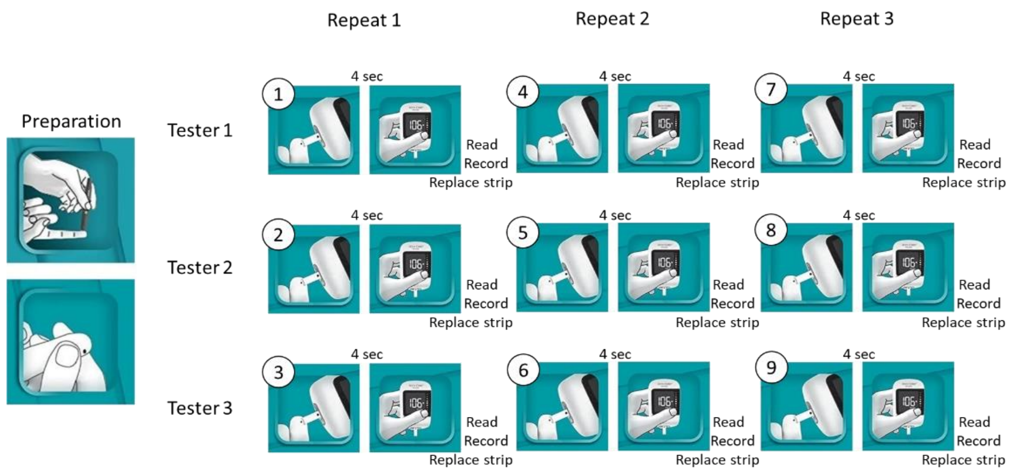

Figure 1. Illustration of how one sample was measured in 3 testers, with 3 repeats
(Own figure, with illustrations from Accu-Check quick start guide)

As bias is not tested in the Gage R&R study, there is no need to for a golden standard laboratory method as a reference. Reproducibility is often interpreted as appraiser variation (AIAG MSA, 2010), though in our case there may be other sources of variation. In self glucose monitoring, the operator him- or herself is the sample, we cannot evaluate the effect of operators. However, thousands of biohackers use the same type of glucometer, they are not using the very same equipment, and comparative gold standard laboratory techniques are hardly available. In the study we are examining if the three glucometers of the same Accu-Chek Instant type are identical with regards to the measurement results and uncertainty, thus the equipment being the reproducibility factor in the study. The glucometers used in the study are from various batches and purchases, and are identified in the table below (Table 2).

Table 2. Glucometers used in the study

Tester	Symbol in study	Type	Serial Number	GTIN
Tester 1	G	Accu-Check Instant	97308838832	04015630086627
Tester 2	K		97308493063	04015630086627
Tester 3	Z		98008221952	04015630086580

The total analysis eventually consisted of 90 readings of having 10 samples tested 3 times in three testers, with the same batch of test stripes throughout the whole study. The coverage factor (VDA 5, 2021) estimating the width of the uncertain range is chosen to be k=2, i.e., ±2σ represents 95% confidence interval, and this rounding is typically used in automotive industry. The process specification limits are of dispute in the case of blood glucose testing. In the study, we have chosen the default target range of Accu-Chek Instant as specification limits, the lower limit is set to 3.9mmol/l and the upper limit is set to 8.9mmol/l (Roche, 2018).

Results and Discussion

The 90 observations are plotted on a Gage Run Chart (Figure 2.) to inspect the range and any unusual patterns on the data. Time sequence for each sample follows the sequence as illustrated on Figure 1. No clear indications of time dependent factors are seen on the graph. Sample ranges from 5.5 to 9.0 mmol/l, that is comparable to the default target range, so we can conclude that the sample is a good representation of the realistic expected range. In some cases, there are unusually large variations observed, e.g. with the results of sample 3 at tester Z, but those do not seem to follow a recognizable pattern, and double checking records of memories from tester devices did not help identifying any variations of special causes.

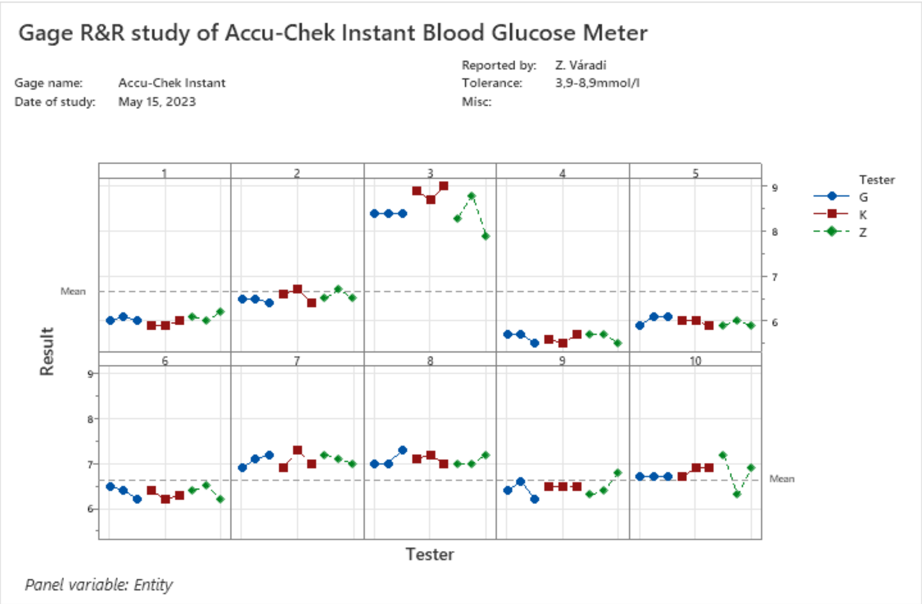

Figure 2. Gage run chart
(own graph)

The ANOVA table (Tables 3.) indicate the significance of factors in the study. First check is at the significance of the interaction term: if the interaction is not significant, Minitab removes that term from the model and repeats the calculation again, as degrees of freedom (DF) and the sums of squares of deviations (SS) change. In our study, interaction term is removed from the model, and the ANOVA results with the main effects only is summarized in (Table 4.) below.

Table 3. ANOVA tabulated results with interactions

Source	DF	SS	MS	F	P
Sample	9	53.8333	5.98148	177.863	0.000
Tester	2	0.0436	0.02178	0.648	0.535
Sample * Tester	18	0.6053	0.03363	1.173	0.312
Repeatability	60	1.7200	0.02867		
Total	89	56.2022			

Looking at the significance of main effects, we can conclude that not all of the samples are identical at 95% confidence level, i.e., the measurement system is capable to differentiate between glucose levels on the examined range. Practically speaking, this is a positive outcome as it means the measurement uncertainty is not hiding detectable changes, thus, the measurement system is good, though its goodness is yet unquantified. The other main effect, the difference between the three testers does not turn significant, i.e., different production batches and wear conditions in this period did not have any detectable effect on the measurement result.

Table 4. ANOVA tabulated results without interactions, α=0,05 to remove interactions

Source	DF	SS	MS	F	P
Sample	9	53.8333	5.98148	200.640	0.000
Tester	2	0.0436	0.02178	0.731	0.485
Repeatability	78	2.3253	0.02981		
Total	89	56.2022			

In addition to the statistical analysis, there needs to be an answer to the acceptability, i.e., the goodness of that measurement system, which consists of Accu-Chek Instant glucometers, lancets, test stripes, the sampled humans, and measurement conditions. We chose to use the automotive practices (AIAG MSA, 2010) saying if GRR%<10%, the measurement system is generally considered acceptable; if GRR%>30% then the measurement system is generally considered unacceptable; and in-between there may be some conditions to consider when deciding about acceptability of the measurement system. GRR percentages study (Table 5.) of the indicate that the total gage variation represented by the ±2σ range accounts for 13.81% of the target range of 3.9 to 8.9 mmol/l, and that percentage is considered conditionally acceptable.

Table 5. Gage R&R results

Source	StdDev (σ)	Study Var (4*σ)	%Study Var	GRR% of the spec range
Total Gage R&R	0.172661	0.69065	20.77	13.81
- Repeatability	0.172661	0.69065	20.77	13.81
- Reproducibility	0.000000	0.00000	0.00	0.00
- Tester	0.000000	0.00000	0.00	0.00
- Part-To-Part	0.813201	3.25281	97.82	65.06
Total Variation	0.831329	3.32532	100.00	66.51

An additional metric of acceptability of a measurement system is the number of distinct categories (NDC), of which values above 5 are in general considered acceptable. For the measurement system analysis performed, NDC=6, suggesting the system is acceptable.

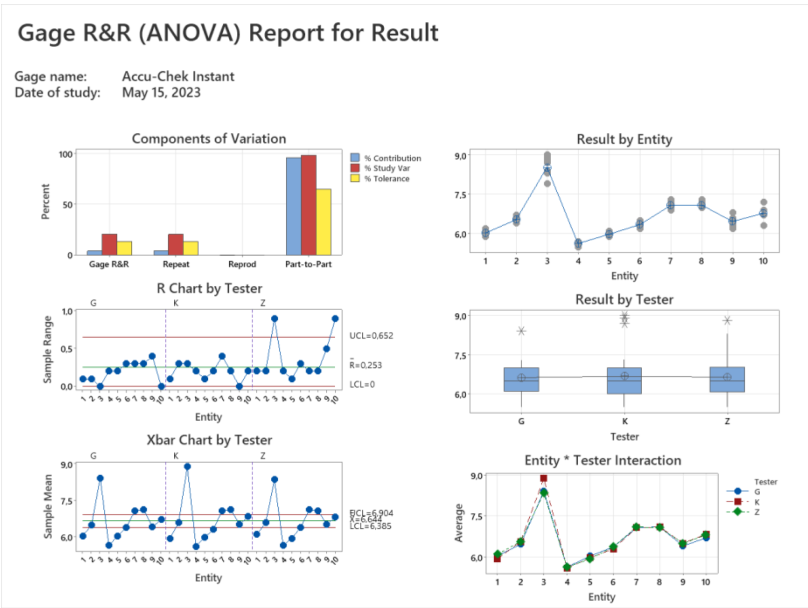

Figure 3. Gage R&R report details
(own graph)

The standard deviation of the total gage turns $\sigma_{GageR\&R}$=0.17 mmol/l, and the average of our 90 measurements are 6.64mmol/l. If we compare that $\sigma_{GageR\&R}$ to those standard deviations given by the manufacturer in Table 1., we can conclude that the Gage R&R method provided very similar results.

Before finally judging the goodness of the measurement system, the assessors have to make sure that the system is stable during the measurement process, and during the experiment. Stability is understood in terms of statistical process control (SPC), meaning that no special cause variation shall influence the process in study.

Neither of the above metrics would by itself indicate stability or signs of special case variations; and so are they unable to identify details of improvement opportunities.

The SPC charts, the R and Xbar charts (Figure 3.) help us identify any unusual pattern or signs of instabilities. Firstly, we need to understand the R chart where any dot represents the range of values of the three repeated measurements on the same sampled blood with a respective glucometer. We see that samples 3 and 10 have extremely large ranges at tester “Z”, i.e., the minimum and the maximum of repeated measurements from the same drop of blood at sample 3 and 10 are atypically far from each other. Investigating the process and the data, no special reason of that unusual range was identified. Assessors suspect that the conditions of tester “Z” may be an unstudied factor: We know that the affected tester “Z” is the oldest equipment of the three, having more than a year of regular use, whereas the other two glucometers were fresh purchases and have been used only for a week or a month, respectively. It is suspected that by aging or wear the measurement uncertainty might change, and it needs to be examined in a further study.

Conclusion

In contrast to the manufacturer’s analysis methods of measurement errors, which are performed on hundreds of tests during and after the manufacturing process, and are difficult to apply on field practices, a routine Gage R&R study brings very similar results on the uncertainty expressed by the total Gage standard deviation. The calculated σ_{Gage} turned to be 0.17 mmol/l, which matches the standard deviation stated by the manufacturer, thus, our study with real field application “in the kitchen” was capable to reproduce the same level of accuracy.

The ANOVA method is very flexible on expanding the study with additional factors of measurement result, and there is no special restriction on samples and study apart from having a sample that represents the range in study. The MSA manual and automotive practices define acceptance criteria for measurement systems, which may serve as guidelines to establish similarly structured acceptance criteria for measurement processes with self-monitoring devices applied on the field.

The simple definition of a GRR% limit may help practitioners and biohackers to judge if their measurement process is good enough for their application. The weakest part of determining GRR% is the sample range: despite our efforts to cover the target range of the glucometer, very low levels of blood glucose are not represented in our sampling, despite, the default target range is well represented. Accu-Chek Instant, a cheap and simple to use self-monitoring device, is capable to perform well in biohacking programs where the main target is to check on a regular basis what effect a certain food brings to the user, or is the blood glucose level low enough to maintain nutritional ketosis.

Recommendations

Repeatedly occurring of questions from participants of biohacking course has motivated the author to perform a Gage R&R analysis on Accu-Chek Instant. A suggestion of the study for the course trainers is to emphasize measurement uncertainty better and support it with numbers. Based on the study, we can suggest setting a limit, e.g., the half width of the 95% confidence interval, that is $2\sigma = 0.34$mmol/l. However, an average biohacker is not necessarily familiar with those statistical terms, and preparation of participants shall more happen on the layman’s language. The instructor may, for example, warn not to worry if a single measurement is not more than 0.3mmol/l off the expected or baseline value.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

I would like to say thank you for my kind colleagues who dared to adjust their diet to the experiment and contributed with their own blood to the measurements.

* This article was presented as an oral presentation at the International Conference on Research in Engineering, Technology and Science (www.icrets.net) held in Budapest/Hungary on July 06-09, 2023.

References

Automotive Industry Action Group. (2010). *Measurement systems analysis. Reference manual* (4th ed.). Chrysler Group LLC, Ford Motor Company, General Motors Corporation

Erbach, M., Freckmann, G., Hinzmann, R., Kulzer, B., Ziegler, R., Heinemann, L., & Schnell, O. (2016). Interferences and limitations in blood glucose self-testing: An overview of the current knowledge. *Journal of Diabetes Science and Technology*. *10*(5), 1161-1168.

Freckmann, G., Schmid, C., Baumstark, A., Rutschmann, M., Haug, C., & Heinemann, L. (2015). Analytical performance requirements for systems for self-monitoring of blood glucose with focus on system accuracy. *Journal of Diabetes Science and Technology*. *9*(4), 885–894.

Horváth, G. (2023). Examination of the physiological effects of functional foods pilot study. *American Journal of Research, Education and Development, 1*, 58-69.

Lakatos, P. (2020). *Kortalanul!* (1st ed.). Budapest, Hungary: Jaffa Kiadó és Kereskedelmi Kft.

Minitab. (2023). *Minitab® 21 Support: Methods and formulas for method of analysis in Crossed Gage R&R Study*. State College, PA: Minitab. Retrieved from https://support.minitab.com/en-us/minitab/21/help-and-how-to/quality-and-process-improvement/measurement-system-analysis/how-to/gage-study/crossed-gage-r-r-study/methods-and-formulas/method-of-analysis/

Roche. (2018). *Accu-chek instant: User's manual. Ref: 07947836003* Indianapolis, IN: Roche Diabetes Care

Roche. (2021). *Accu-chek instant: Tests. Ref:07819382* Indianapolis, IN: Roche Diabetes Care

Smith, T. (2019). How accurate are blood glucose meters? *Beyond Type 2*. Retrieved from https://www.accu-chek.com/print/7961

Verband der Automobilindustrie. (2021). *Volume 5 measurement and inspection processes. Capability, planning and Mğmanagement* (3rd ed.). Verband der Automobilindustrie.

Zheng, J. (2021). Ethical implications of biohacking as activism: Democratızed health care, danger, or what? *Aresty Rutgers Undergraduate Research Journal, 1*(3).

Ziegler, R., Freckmann, G. Heinemann, L. (2014). Genauigkeit von Blutzuckermessgeräten in der Praxis evaluieren: Ist das möglich?. *Diabetes, Stoffwechsel und Herz, 23.*

Author Information

Zoltán Váradi
Óbuda University,
Budapest, Hungary
Contact e-mail: *varadi.zoltan@kgk.uni-obuda.hu*

Győző Attila Szilágyi
Óbuda University,
Budapest, Hungary

To cite this article:

Varadi, Z., Szilagyi, G. A. (2023). Applicability of a gage R&R study on a home blood glucose meter. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 23,* 292-299.