THE EUROPEAN SOCIETY OF MEDICINE
Medical Research Archives, Volume 12 Issue 9

CASE REPORT

A long-term case study on marginal diabetes: demonstrating the need for effective calibration of home test devices

Franco Pavese[1,2]

[1] Independent Scientist
[2] Former Research Director in Metrology at CNR, Torino, Italy

OPEN ACCESS

PUBLISHED
30 September 2024

CITATION
Pavese, F., 2024. A long-term case study on marginal diabetes: demonstrating the need for effective calibration of home test devices. Medical Research Archives, [online] 12(9). https://doi.org/10.18103/mra.v12i9.5873

DOI
https://doi.org/10.18103/mra.v12i9.5873

ISSN
2375-1924

ABSTRACT

Accurate tracking of the trend in time of marginal diabetes level is not simple, as the precision of non-professional home testers is typically ±10–12 mg/dL—45% of the marginal range. A reliable assessment of the current calibration of home testers is difficult because it depends on the quality of the off-shell calibration and then on its subsequent validity in time. The paper reports test results specifically following a case of initial deterioration of the pancreas functionality, with the patient subsequently having to follow a strict diet with also assumption of Glucophage. The study carried out on its evolution made it possible to propose, with examples, a valid test procedure for optimising the performance of home testing without strictly requiring professional expertise, supplementing scarce/insufficient manufacturer information. The effect of the medical treatment was taken under observation for a total of ≈2300 morning glucose home tests over 8 years using several home testers models. The aim of the paper is not about the validity of the test method, but about getting evidence of the basic importance of the calibration of the testing devices, of its validity and sufficient accuracy, to improve the odds of a correct medical diagnosis—and patient confidence.

Keywords: marginal diabetes; long-term home testing; tester calibration; medical diagnosis; patient confidence

Introduction

Tracking the trend in time of *marginal* diabetes level (conventional capillary range 100-125 mg/dL) is not simple, as the precision of non-professional home testers is known to be ≈ ±10%, i.e. ±10–12 mg/dL —±45% of the full range: thus, an error of only 5 mg/dL already amounts to an uncertainty of 20% on the diagnostic basis.

According to the rules of measurement science, the fact that often the *reproducibility* of repeated tests is better than *precision* does not resolve the medical goal: making the diagnosis based on an estimate of the *accuracy* of the glucose level as indicated by the tester. That feature depends on the *calibration quality* (we are used to saying that measured values can be very repeatable but, in fact, be all *wrong*). Quality concerns either the value—initial and then its constancy in time—or its uncertainty.

For the accuracy estimation, this study reports the full health history from age 40 of a *patient—The author's* data were chosen to ensure full knowledge of how the tests were performed and that the rules of measurement science were strictly followed, being the Author's main professional expertise.

The basic features of the clinical method used for measuring the glucose level in capillary blood at home are only summarised here, as at the *initial* stage of this study they were already the object of a publication [1]. On the other hand, the issue of measurement accuracy of glucose level in diabetes home testers is amply discussed in the literature, e.g. see [2–6].

The paper is based on the study of a sufficiently extended set of results of the tests performed during the disease treatment for 8 years following a case of deterioration of the Author's pancreas, a period during which a strict diet in addition to the assumption of Glucophage was followed—the patient history before that period is reported only to indicate his glucose levels since age 40. The effect of the treatment was taken under sufficient control with a total of about 2300 glucose morning tests.

For this study, six different home testers from different manufacturers have been used so far for monitoring capillary blood. Occasional tests made in hospitals/laboratories on plasma blood are also reported. The effect of the departure of the indicated off-shell calibration from the *nominal* glucose value or the use of control solutions, when available, and how these data were used are reported and used for correcting the tester's readings to get *calibrated monthly means*.

The results are then discussed and suggestions are provided.

The method

As reported in [1], *glucose level readings* were always taken in the morning before any assumption of food or drink, normally between 8:00 and 9:00 a.m., by producing a small drop of blood on a fingertip (most often *the same finger*) of the cleaned hand with a new lancet, and the value immediately read on the tester and recorded. A delay after 9 o'clock was observed to possibly increase the reading at least by about 10 mg/dL. A limited increasing effect was occasionally observed in case of an out-of-diet previous dinner or of a short hill condition (a cold, a gastrointestinal disorder, …).

No effect was attributed to the size of the blood drop, anyway limited to a volume ratio variation of 1:2 max. No occasional "second drop" inconsistency was observed outside the normal variability range between strips. Nothing was ever reused.

Occasionally, measurements were repeated by using two different testers, the second being older but using strips still within their deadline date, to check the subsequent stability in the time of the older tester calibration.

The focus of this paper is not on the validity of the test method but on the possibility of getting evidence of the basic importance of the tester calibration. The details of the proposed method for the *corrected* computation of the glucose level are reported for convenience in the Appendix, with examples.

Readings elaboration

Tester readings were reported and elaborated on an Excel datasheet. When using testers indicating *plasma* blood, 1.11 divides the measured value in order to get the capillary blood value. At the end of every month, a *calibrated* (see later) mean value was computed as reported in Fig. 1 and recorded *with its standard deviation* (s.d.).

The strip-batch data reported by the manufacturer basically indicate the two boundaries of the *valid data range*. The *reference* (nominal) value of the batch was then considered to be the *middle* value of the valid range: for most of the testers, it was 100 mg/dL, except one which was instead 140 mg/dL. When a control solution for the nominal value was available (not for all testers), checks of the tester's current calibration were also performed. More information and comments about details for the first 3 testers until 2020 (test N° 1200 in Fig. 1) can be found in [1].

When these tester checks showed a deviation from the nominal value of the control solution, or when the centre of the *valid* interval R_C did not coincide with the *nominal* value R_N (e.g., 100 or 140 mg/dL in this study), the ratio $r = R_C/R_N$ is computed and use for readings' correction—see later.

Analysis of the data

Figure 1 reports the full set of data for the patient (including in Fig. 1(a) older data before 2016) and for all six testers used in the following subsequent periods: tester #1 before middle 2018; tester #2 2018-middle 2019 (s.d.: mean 11 mg/dL, range 5-18 mg/dL); tester #3 2019-20 (s.d.: mean 8 mg/dL); tester #4 2021 (s.d.: mean 5 mg/dL); tester #5 2022-23 (s.d.: mean 8 mg/dL, range 3-10 mg/dL); tester #6 2024.

In the Figure symbols indicate:
Red small dots are the original tester readings: these *uncorrected* tester readings can be amply above the

125 mg/dL upper limit of the marginal range, up to about 190 mg/dL (see Discussion Section).

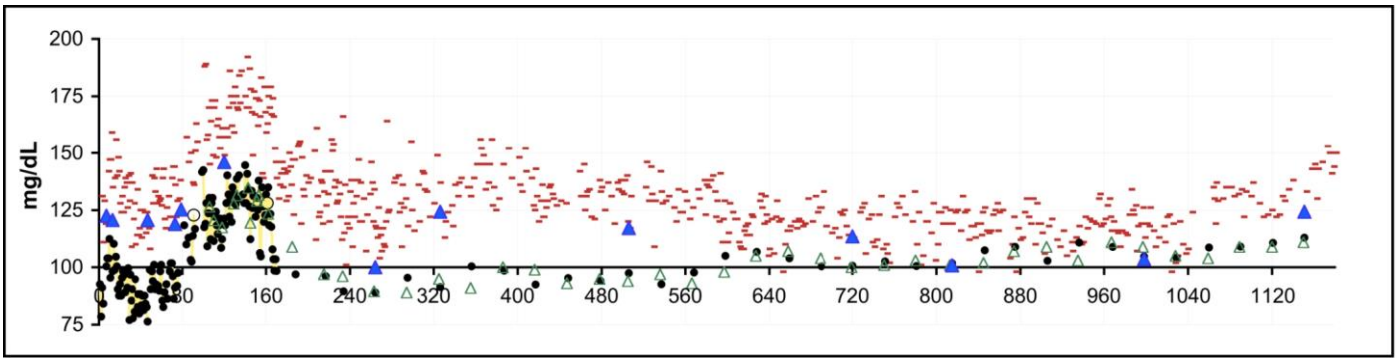
Black dots are the computed *calibrated monthly* values (Note: each black dot value is equal to the value of the *white triangle dot* of the previous month, for easy visualisation of the monthly change).

Blue dots indicate hospital professional tests on blood plasma corrected for capillary blood.

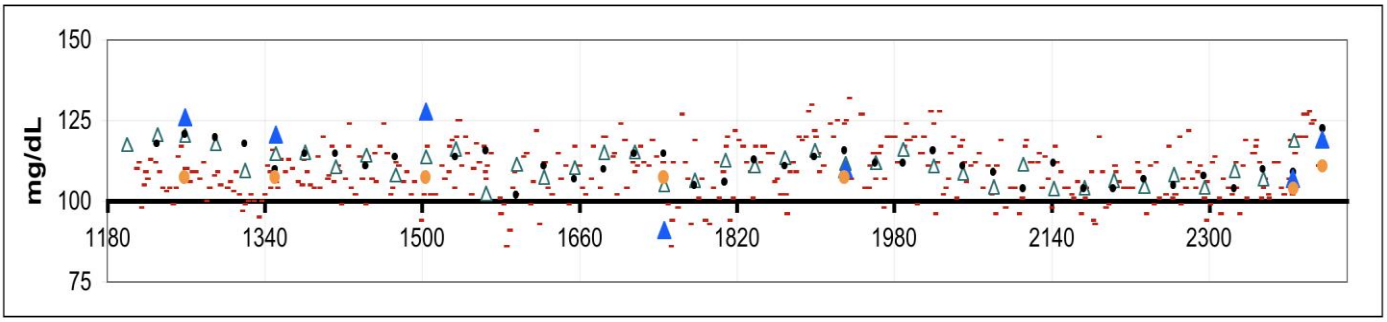
Green dots mark the start of January of each year: the first green dot at test N° 480 is at the end of year 2018 (while the *first black* dot below 100 mg/dL, test N°170, is April 1, 2018); the *last* green dot is at test N° 2280 at the end of year 2023.

Note: The author's data availability actually started at the age of 40, but the abscissa range omits those before 2009 (age 67) since the previous values were all below 100 mg/dL.

The rightist data in Fig. 1(a) are at age 67, the *age of retirement from work*, and the glucose level was still below 100 mg/L. Subsequently, there has been a level spike up to 110 mg/dL lasting two years (2011-12, change of lifestyle?). Then, until 2015, from age 69 to 73, glucose levels remained below 100 mg/dL again.



(a)



(b)

Figure 1. Database of marginal glycaemia measurements on whole *capillary* blood, performed on a single patient: 6 different non-professional testers for home testing were subsequently used (see text above).
On abscissa the progressive test days since 2009 (age 67): only since 2018 (Test N° 450) the scale reports individual days up to June 2024 (age 82).

In April 2016 at the age of 74.5, the Author suffered a health accident while travelling (virus disease due caused by a single accidental assumption of contaminated water in Senegal), with the consequence of an immediate rise of the glucose level to 120 mg/dL (*white larger dot*) and the subsequently increased up to 140 mg/dL (corrected value, the uncorrected being up to ≈190 mg/dL). Meanwhile, shortly after a medical consultation indicating damage to his pancreas functionality, the daily assumption of Glucophage (1000 mg single serving per day after dinner) and a strict diet started—the wider *yellow dot* marking it as the beginning of March 2018. Only after one year and a half since 2018, the treatment resulted in a significant drop of glucose level concentration in blood—after correction back—to within a range of 100–120 mg/dL and a standard deviation of ± 5 mg/dL.

The monthly means have been reported in Fig. 1(b) since 2018; however, the time scale has been enlarged to accommodate daily data, reflecting a higher resolution.

Actually, the testers used until 2020 were more or less all off calibration, as there is clear evidence when confronting the red and black clusters of data, while subsequent testers were progressively better qualified, also from the point of view of their *reproducibility*.

However, as shown in Fig. 1(a) for the year 2020 up to test N°1200, a bad tester/batch-of-strips was still encountered—its inconsistency being demonstrated by the values *after correction*, revealing a true increase of glucose but of only +10 mg/dL. Incidentally, that happened in the fall of the year 2020 (starting on test N° 850), i.e., a few months after the start of the COVID pandemic—however not having affected the patient living isolated since March 2020 at a mountain home.

After that previous experience, more recent testers with plasma indication were used, found to require *smaller calibration corrections*—closer to their *reproducibility*—or even none. The quality improvement is evident from Fig. 1(b).

On the other hand, the standard deviations of the tester readings, reported at the beginning of this section, did not correlate with the out-of-calibration condition of the older testers, which was limited to only a larger reading irreproducibility, 11 mg/dL instead of 5–8 mg/dL to—anyway meaning an uncertainty increase from ≈25% to ≈45% of the marginal glycaemia range.

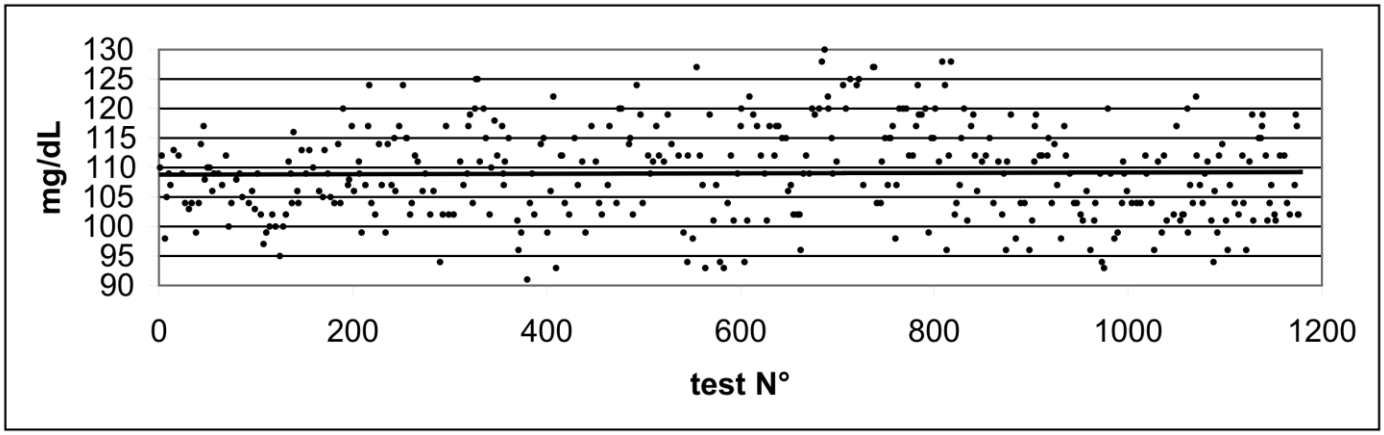
Note: after 2020, previous testers were still occasionally also used (data not reported in the Figure), and their *corrected* values still always resulted in agreement with those of the current tester.

The trend in time of hospital measurements provided on blood plasma by a few *hospitals/laboratories* showed variations larger than the mean trend of home tests, with frequent significant differences with respect to them.

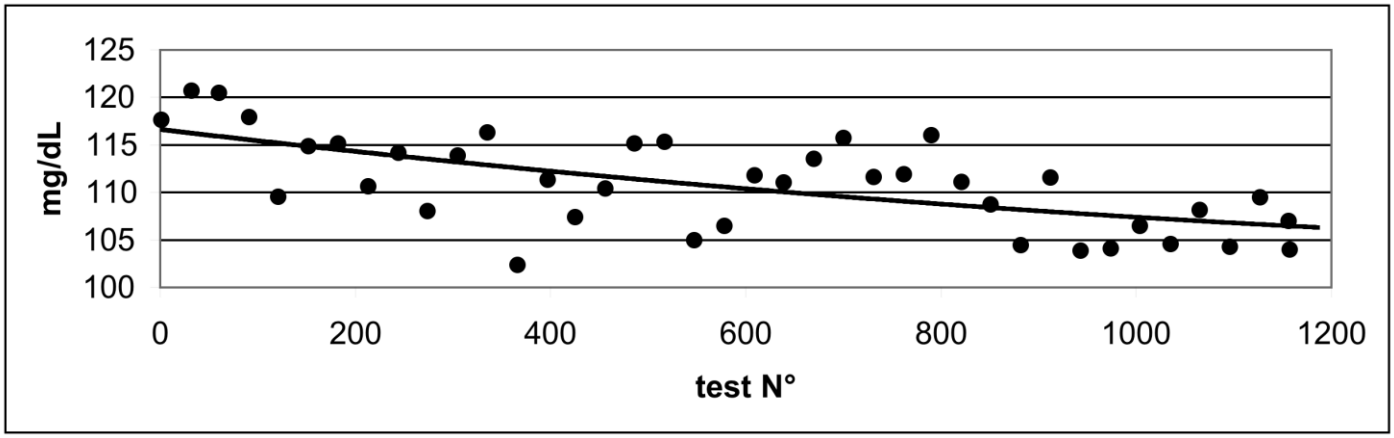
Discussion

The comparison of corrected and uncorrected data in Fig. 1 indicates beyond any doubt that the method used in this study to correct the original tester readings is accurate and consistent. The corrected values fall within a narrower range of marginal diabetes—except during the very initial part of the disease—while the uncorrected data show a much larger dispersion of values.

To provide more evidence to the above statements, Fig. 2 reports the period from January 2021 on, when two testers were used, still with unstable calibration in time and correction factors in the range of ±14%.



(a)



(b)

Figure 2: (a) Original tester readings; (b) *Corrected* monthly means.

The s.d. improved from about 8.1 mg/dL to about 4.9 mg/dL. Note that, while in Fig. 2(a) measurements show an almost random trend around a fixed value of ≈110 mg/dL, the *calibrated* monthly values in Fig. 2(b) show a significant trend with a slightly quadratic fit and a decrease of -10 mg/dL: in fact, the yearly *corrected* means are 115 mg/dL in 2021, 111 mg/dL in 2022, 109 mg/dL in 2023 and 107 mg/dL in 2024 (partial).

In the Appendix, detailed information is supplied on the suggested tests and calibration methods used in the present study.

The kind of correction detailed in the Appendix should be indicated in the instructions to users, embedded in the tester software, or computed by the tester software.

An unresolved requirement to get valid measurements comes from "control solutions". As already stressed in [1], the needed full information for such a "reference material" should be provided by the manufacturer, according to the rules internationally agreed upon for such important materials: in particular, the indication of the nominal concentration value and its associated uncertainty, presently not found for the supplied ones, is a must.

The present giant market of billions of dollars of diabetic testers for home checks shows a great variety of *quality* of the supplied instruments. That is not acceptable and stricter and specific ISO standards should also be established for them.

On the other hand, for the vast majority of users, it is not possible to assume sufficient individual technical knowledge and capability to apply specific procedures. Consequently, the correct procedure should better be embedded into the tester software, it is not a problem today as that feature is the only and most extensive progress noted in modern testers.

Conclusions

As the disease evolved smoothly in time, it was possible in this study to get full evidence of the importance of the *current* calibration of the test devices for making a correct medical diagnosis possible. It is now possible to report a valid procedure supplementing the scarce/insufficient information often provided by tester manufacturers.

The vast majority of the patients using home testing have no technical expertise, so the manufacturers are making an effort to embed into the software of the measuring devices what is a more and more complex procedure—but often less and less clear and clearly explained.

However, the author's experience from half a dozen testers is that the information about the meaning of the results, or the use of the control solutions—from the tester software or literature/information sheets—is becoming more and more confusing and insufficient, often voiding numerical information, probably considered less familiar to the user. At present, nothing better is available—neither, possibly, a technical supplement for medical doctors and experts in the field—to make understanding and confidence sufficient about the *quality* of the measured values and confidence on them, on which the medical treatment then depends.

As an example, qualitative indications like "low", "medium" and "high" glucose levels are often used instead of the quantitative numerical ranges, also concerning the glucose level in the control solutions. As shown in the Example Section, the issue is quite complicated and the risk of mistakes in getting the data and in their subsequent mistreatment looks high.
On the other hand, data of professional origin (Hospitals, Laboratories) seem unable to provide a trend firmly reassuring about the course of the illness, also because, generally, that dataset cannot be as dense as home tests are.

Conversely, the use of continuous information today allowed by informatics with the use of permanently implanted or installed sensors, may simply provide an unnecessarily large amount of data (on the style of the popular Big Data), or provide useful alarms, but without ensuring in themselves that the *process* is correctly monitored and, even more, that the measurements are then treated with the correct procedure—namely concerning the issue of sensor's calibration and it stability in time.

The *marginal* glycaemia range (often called "range 2"), important for the correct diagnosis and then prevention of its evolution to full diabetes, should require more effort from the tester manufacturers and from ISO to reduce the instrumental errors.

Final Note: The Author understands that the dialect of measurement science used in this paper (namely in the Appendix) may be not easy or fully familiar to non-specialists, and feels the need to alert them about this fact. The *precision* evaluation of the testers/strips' quality is quite a complicated issue: a multidisciplinary approach could be useful.

Conflicts of Interest Statement

The Author does not report any conflict of interest concerning the paper contents.

Funding statement

The Author did nor receive any king of fund for the preparation of this paper.

References

1. F. Pavese, A testing/metrological look at the accuracy of glucose strip measurements in home care for marginal diabetes, for mitigating diabetic kidney disease, Commentary, Journal of Nephrology 2022 https://doi.org/10.1007/s40620-021-01224-6, pp. 1–5. With Supplementary Information.
2. L. Heinemann, V. Lodwig, G. Freckmann, Accuracy in blood glucose measurement: what will a tightening of requirements yield? J Diabetes Sci Technol (2012) 6:435–443
3. S. Finfer, J. Wemerman, J.-C. Preiser, T. Cass, T. Desalve, Clinical review: consensus recommendations on measurement of blood glucose and reporting glycaemic control in critically ill adults. Crit Care (2013) 17:229
4. L. Ekhlaspour, D. Mondesir, N. Lautsch, C. Balliro, M. Hillard, K. Magyar, L. Goergen Radocchia, A. Esmaeili, M. Sinha, S.J. Russell, Comparative Accuracy of 17 Point-of-Care Glucose Meters. J Diabetes Sci Technol. 2017 May; 11(3): 558–566. https://doi.org/10.1177/1932296816672237
5. F. King, D. Ahn, V. Hsiao, T. Porco, D.C. Klonoff, A Review of Blood Glucose Monitor Accuracy. Diabetes Technology & Therapeutics (2018) Vol. 20, No. 12. https://doi.org/10.1089/dia.2018.0232
6. M. Tauschmann, G. Forlenza, K. Hood, R. Cardona-Hernandez, E. Giani, C. Hendrieckx, D.J. DeSalvo, L.M. Laffel, B. Saboo, B.J. Wheeler, D.N. Laptev, I. Yarhere, L.A. DiMeglio, M. Tauschmann, et al., ISPAD Clinical Practice Consensus Guidelines 2022: Diabetes technologies: Glucose monitoring. Pediatr Diabetes. 2022 Dec;23(8):1390-1405. https://doi.org/10.1111/pedi.13451

Appendix

Calibration procedure for the correction of tester readings in the marginal diabetes range.

Capillary blood will only be considered here. To obtain up-to-date calibrated readings from the current tester readings, the procedure to follow is provided and illustrated below.

One must make a distinction between the tester and the strip lot.

(A) **Tester.** The manufacturer, with exceptions, does not explicitly indicate if the off-shell tester is calibrated, i.e. if it is exactly centred on the *nominal* value of its range(s) when tested with a control solution, or if there is an offset—nor indicates the precision of that value, e.g., 5% or 10% or else. It happens that no indication is provided about that value (e.g. it may be 140 mg/dL instead of 100 mg/dL or else—here reference is to the marginal diabetes range, not necessarily coinciding with the so-called "intermediate" range).

For the range considered in this paper, let us assume that the correct *nominal* value is intended to be R_N =100 mg/dL exactly.

The user should buy the control solution (CS_{tester}) for that range: that may not be a trivial task due to incomplete information supplied by the manufacturer, e.g., of "medium" range or "range 2". One can understand if the CS_{tester} is the correct one only from the results of the measurements with it, see later.

The user should perform the test on a drop of CS_{tester} as indicated by the tester information sheet. For each measurement, the user will use a strip of the current batch available. It is recommended to perform at least 3 tests by using 3+ strips. Then one compares the *tester mean value* $R_{tester,mean}$ of the results with the *nominal* one of the range: a standard deviation can also be computed, s.d.(R_{tester}).

In general, a difference will be obtained: let us assume as an example $R_{tester,mean}$ = (110 ± 4) mg/dL instead of the nominal value 100 mg/dL. Notice that also the strips have their imprecision (see later) concurring with the registered imprecision: we will illustrate that later in (B), but it is clear now that the results have a "circular" influence between apparatus and strips, which may sometimes be considered as a "second order" effect.

In the first approximation, the above off-calibration (d_{tester} = (+10 ± 4) mg/dL) must be recorded as it will be used in the following. The off-nominal deviation of the *control solution* and its uncertainty are assumed to be irrelevant, at least at a non-professional level: otherwise, they would require a supplementary adjustment in all the above parameters.

(B) **Strip batch.** Each batch of strips has its imprecision, different from batch to batch, independent of tester calibration. They have *their own control solution*, (CS_s) for each of their reported nominal correct ranges of use.

The tester must previously have undergone the calibration illustrated in (A) so that the measurements on the strips are corrected by $-d_{tester,cal}$.

A drop of CS_{strip} is measured with at least 3 strips, using 3+ strips, and the mean value $R_{strip,mean}$ and its s.d.(R_{strip}) are computed for the tests (A). As said, each measured value (or the mean value) on the strip must be corrected for the calibration offset of the tester, $-d_{tester,cal}$.

In general, there will be a difference, as before: let us assume as an example $R_{strip,mean} - d_{tester,cal}$ =(116 ± 10) mg/dL instead of the nominal value 100 mg/dL. Thus, a strip-batch offset d_{strip} = (+16 ± 10) mg/dL is recorded.

All the tester readings on (capillary) blood must then be corrected by $d_{strip,corr}$ = (–16 ± 10) mg/dL.

The off-nominal deviation of the *strip control solution* and its uncertainty are assumed to be irrelevant, at least at a non-professional level: otherwise, they would require a supplementary adjustment in all the above parameters.

The above is correct only in the first approximation. In fact, the results in (B) are *also* affected by the *tester offset and uncertainty*: however, the above procedure, of correcting the measured values of the strip control solution for the tester offset, basically limits the effect of tester off calibration.

Instead, the uncertainty of $u(d_{tester})$ must be added *in quadrature* to $u(d_{strip})$, so that the uncertainty affecting the correction $d_{strip,corr}$ = –16 mg/dL is actually: $u(d_{tester}) = \sqrt{2}(4^2 + 10^2) = \pm$ 11 mg/dL, 44% of the marginal glycaemia range. That may be the order of magnitude of the correction that one has to perform on all tester readings on all capillary blood measurements, and of its large uncertainty.

An example: comparing two home-type testers in March 2024 during an interval of 15 days.

Tester #1

Mean of 11 readings in the same period of tester #2: *110* mg/dL, s.d. *7* mg/dL (= 6% readings variability).

Corrections (no information of strips batch and tester uncertainty from the manufacturers): tester #1 calibration 1.016, strips 1.014; tot 1.031.

Corrected value: ***104.1*** ± *7* mg/dL (*7*%).

Difference of calibrated value from reading: –*5.7* mg/dL (–5.2%), **23%** of the marginal glycaemia range— Notice how an apparently small difference corresponds to a large uncertainty of the result with respect to the restricted marginal range.

Tester #2

Mean of 4 readings made in 15 days: *111* mg/dL, s.d. 2.6 mg/dL; tester #2 – tester #1 = 1 mg/dL.

Corrections (no information of strips batch and tester uncertainty from the manufacturers): strips (same batch) 1.014, mean (tester #2 – tester #1) ratio: 1.071; tot 1.086. No calibration is available for tester #2.

Corrected value in the same period: ***102.2*** ± 3 mg/dL; tester #2 – tester #1 = –1.9 mg/dL (2%), with respect to the just-calibrated tester #1.

Difference of calibrated value from reading: –8.8 mg/dL (–9%), **35%** of the marginal glycaemia range.

For this kind of calibration (the minimum type required), the role of the systematic error effect remains unknown, at least over a short period. However, the procedure may look too complex for the average patient concerning their home tests: the tester should embed it, only asking the user for the correct sequence of tests—it might even look too costly to the patients, though they are advised to perform it only 2-3 times per year. Standard deviation should also be made available, possibly named, e.g., “dispersion of results”, as a *quality index* of the results.