On the Commonly-Used Incorrect Visual Representation of Accuracy and Precision

Shervin Shirmohammadi, Luca Mari, and Dario Petri

We performed a Google image search using the search phrase "accuracy and precision" and, after removing unresponsive and duplicate web pages from the search results, found that 78 of the top 100 results use the bullseye chart to visually explain this concept. Unfortunately, we also found that 52 of those 78 results, i.e., an incredible two thirds, use a visual similar to what is shown in Fig. 1(i) or Fig. 1(ii), both of which are at best highly misleading and factually incorrect according to relevant standards and guidelines. In this figure, the black dots represent the values obtained by replicated measurements of the same measurand (the quantity intended to be measured), and the visual aims to present how close the measured values are to each other and to an agreed-upon reference value, represented by the red bullseye. In the search results, visuals similar to Fig. 1(i) and Fig. 1(ii) appeared 25 and 27 times, respectively. In the top 20 search results, which is where the great majority of users look for answers, the earliest correct visual appeared only at position 11, followed by 13, 15, and 19 in the rankings. The great majority of the 52 incorrect visuals were in non-peer-reviewed documents, reinforcing the notion that one should not believe everything one sees on the internet, although shockingly, a few were in papers published in peer-reviewed scientific venues, including one in Elsevier's *Journal of Clinical Epidemiology* and another one in the SCMR's *Journal of Cardiovascular Magnetic Resonance*, both well-cited journals. The actual numbers are probably higher, as many scientific articles are behind paywalls and inaccessible to Google search.

There is of course a degree of uncertainty in this search experiment: repeating the search in a different geographic

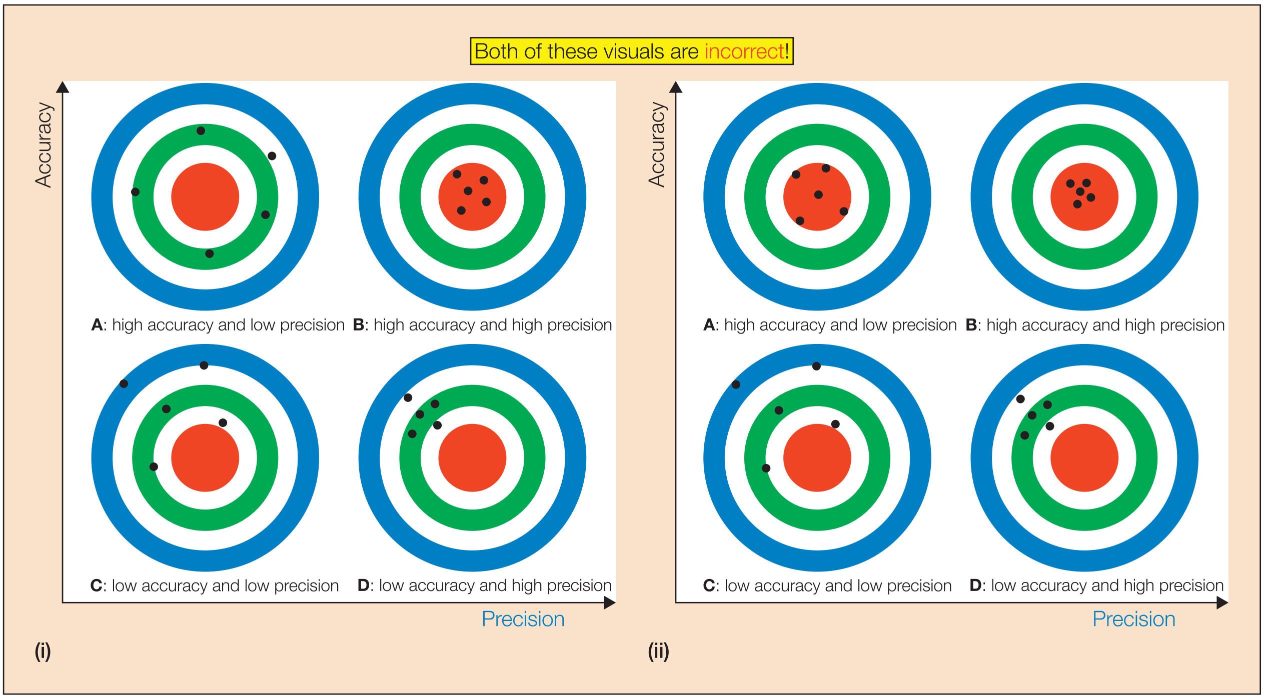

Fig. 1. Commonly-used incorrect representations of Accuracy and Precision.

1094-6969/21/$25.00©2021IEEE

location, a different language, or a different time (say a few months from now) would give different results. But it is reasonable to conclude that currently, the majority of search results, whether two thirds or a slightly different percentage, are teaching incorrect information to readers, and this is very concerning. So, we wrote this short article as a one-stop shop to help educate interested readers about the correct definitions and therefore correct visual representation of accuracy and precision. We tried to keep the technical language accessible to readers whose expertise is outside of Instrumentation and Measurement (I&M). As such, I&M experts will notice some deliberate oversimplifications of concepts.

So why do we say that Fig. 1 is incorrect? We have two reasons for our claim. The first relies on what is commonly meant by "accuracy." In short, parts A of Fig. 1(i) and Fig. 1(ii) assume that because the average position of the individual measured values is close to the bullseye, then accuracy is high. This is incorrect because accuracy should be considered for a single measured value: even by visual inspection, is it not obvious that each black dot of Fig. 1(i) part A is rather far from the target's bullseye, quite similar to the "low accuracy"—labeled part D, and is therefore not accurate? Would you call "accurate" the marksman in Fig. 1(i) part A who always misses the bullseye so badly, compared to part B? Apparently, the answer to the latter question is "yes" for the authors who use this visualization: not really plausible! Fig. 1(ii) tries to make it "look" better by having the black dots in part A appear closer to the bullseye, but this causes another problem: the precision in part A is now quite similar to the "high precision"—labeled part D, so how can part A be "low precision?" In fact, as we will explain later, "high accuracy and low precision" is an oxymoron: a measurement instrument cannot be accurate and imprecise at the same time! The second reason is based on a couple of relevant technical documents, VIM [1] and ISO 5725 [2], that clearly present the relations between accuracy and precision [1], [2]. But before browsing these technical documents for answers on how to avoid the mistakes of Fig. 1, a fundamental clarification is in order: accuracy and precision may refer to features of either measuring instruments or measurement results and consequently measured values. This distinction is particularly important with respect to accuracy because it leads to this question: is the target's bullseye actually known when accuracy is to be evaluated? The short answer is: "yes" in the case of measuring instruments, because the reference value could be obtained from a measurement standard available from the instrument's manufacturer, and "no" in the case of measurement results, because the reference value should be the measurand's true value, which is unknowable according to the *International Vocabulary of Metrology* (VIM) [1, 2.11 Note 1]. Let us keep this basic distinction in mind while reading what follows.

Accuracy and precision have an interesting connection with *measurement error*, which is defined by VIM as *measured quantity value minus a reference quantity value* [1, 2.16]. Whenever a reference value is known; i.e., in the case accuracy is considered of a measurement instrument and not of the measurement results, as explained above, the measurement error implied by a measured value can be expressed as consisting of two components: the systematic error and the random error. According to the VIM, *systematic error* is the *component of measurement error that in replicate measurements remains constant or varies in a predictable manner* [1, 2.17].

Examples of causes of systematic errors include lack of calibration, the operator reading the meter not at eye level but at an angle, and the long-term instability of the measuring instrument. In Fig. 1(i) and Fig. 1(ii) parts C and D, we can see that due to systematic errors, the average position of the black dots is off the red bullseye, always towards the top left. This is also known as *bias*, which is indeed an estimate of the systematic error [1]. This offset is knowable only if the position of the bullseye is known: systematic errors are knowable only if a reference value is given.

In a complementary way, the VIM defines *random error* as the *component of measurement error that in replicate measurements varies in an unpredictable manner* [1, 2.19]. Examples of causes of random errors include the effect on instrument indications of noise interference and fluctuations in environmental conditions. In Fig. 1(i) parts A and C, we can see that due to random errors, the black dots are further from each other compared to those in parts B and D. In each part, random errors can be described by a probability distribution function with zero expectation (since by definition offsets are accounted for by systematic errors) and a variance that increases if the magnitude of random errors increases. It is important to understand that random errors are not related to a reference value: even if the bullseye is hidden, the information about the relative spread of the black dots would remain visible.

The distinction between systematic and random errors shows that a better understanding requires a third feature, in addition to accuracy and precision, which the VIM and the ISO 5725 series of technical standards call *trueness*. The basic idea is simple: similar to measurement error which consists of systematic and random errors, accuracy also consists of trueness and precision of the instrument. And like measurement errors which affect the accuracy of the instrument, systematic errors affect its trueness and random errors affect its precision. With this premise the definitions given by ISO 5725 become even more meaningful, which defines *trueness* as *the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value* [2, 3.7].

Note the usage of "test results," not measurement results or measured values; this emphasizes that trueness may be attributed to a measuring instrument, and the same applies also to the definitions that follow below. This definition tells us that because the average position of black dots in parts C and D of Fig. 1(i) and Fig. 1(ii) is not close to the target's bullseye, trueness is low. Vice versa, trueness is high for parts A and B because the average in each part is indeed very close to the target's bullseye. This means that it would have been correct if the vertical axis in Fig. 1 had been labeled "trueness" instead

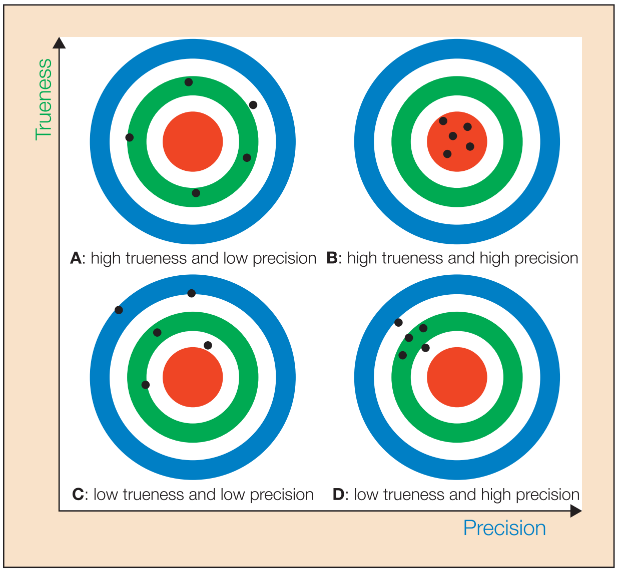

Fig. 2. One way to correct the incorrect Fig. 1. However, trueness is not a widely-used feature, so this visualization might not be very helpful in practice.

of "accuracy!" Trueness is then the feature of an instrument that indicates its ability to avoid systematic errors: the less the systematic errors (i.e., the closer the average of the measured values to the reference value) the greater the trueness.

But what about *precision*? It is *the closeness of agreement between independent test results obtained under stipulated conditions* [2, 3.12]. Thus, precision depends only on the distribution of random errors and does not relate to a reference value. We can see in Fig. 1(i) that the black dots are close to each other in parts B and D, while they are further apart in parts A and C. In that sense, we can say that the figure is correctly visualizing the precision of the instrument. Precision is then the feature of an instrument that indicates its ability to avoid random errors: the less the random errors (i.e., the closer the measured values to each other) the greater the precision. This gives us the first corrected version of the incorrect Fig. 1, shown in Fig. 2.

Problem solved? Not really, because compared to accuracy, trueness is rarely used in natural sciences and engineering; hence, we still have to include accuracy in the visual. To do so, we first need to know what *accuracy* is. According to ISO 5725 it is *the closeness of agreement between a test result and the accepted reference value* [2, 3.6]. Hence, accuracy is defined for a single measurement, unlike trueness and precision which require a sample of values from replicated measurements. Equally important is to note that accuracy like trueness, and unlike precision, depends on a reference value. And in fact, the VIM—which considers accuracy to be related to measurement results instead of measuring instruments (it defines accuracy as "closeness of agreement between a measured quantity value and a true quantity value of a measurand" [1, 2.13])—is explicit in noting that accuracy is not evaluated quantitatively, and that *a measurement is said to be more accurate when it offers a smaller measurement error* [1, 2.13, Note 1] and of course the same holds for trueness.

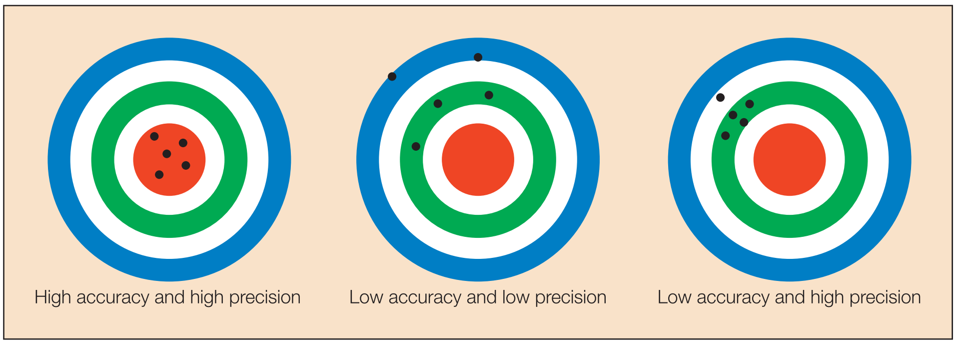

Fig. 3. Some authors prefer this visualization.

Although accuracy is defined for a single measurement, we are not really interested in the accuracy of a specific value, whose offset from the reference value depends in a non-predictable way by random errors. In other words, we do not want to accept cases in which a high accuracy is obtained by chance. What the definition of ISO 5725 actually means instead is that an instrument is accurate if *each* result it produces is (or at least *most* of the results it produces are) accurate.

The consequences are now clear:

1. Since accuracy depends on both systematic error and random error that affect any single test or measurement, instrument manufacturers can evaluate quantitatively the accuracy of their instrument, despite the fact that there are no standardized procedures to evaluate accuracy as a function of trueness and precision.
2. For an instrument to have a high accuracy, it needs to have both a high trueness and a high precision. Therefore, the concept of "high accuracy and low precision" as shown in part A of both Fig. 1(i) and Fig. 1(ii) is nonsense, because precision contributes to accuracy. For this reason, a representation like the one in Fig. 3 is correct but still partial, since it remains implicit about the role that the trueness and precision have to determine accuracy. Even better is therefore the representation in Fig. 4, which shows how trueness and precision independently contribute to accuracy.

Back to our Google search results, the 26 results with the correct visuals consisted of 23 and 3 illustrations that were similar to Fig. 3 and Fig. 4, respectively. Not surprisingly, none used any visual similar to Fig. 2 which shows trueness and precision but not accuracy.

On the same subject, another incorrect visual representation which is less commonly used but still worth mentioning is shown in Fig. 5, which shows that the values obtained from replicated measurements of the same measurand are distributed between V_{min} and V_{max}, with a probability distribution

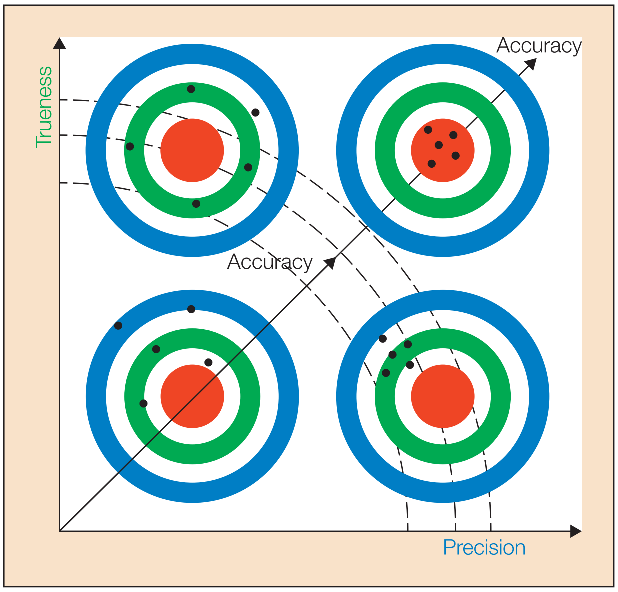

Fig. 4. This correct visualization has also been used by some authors.

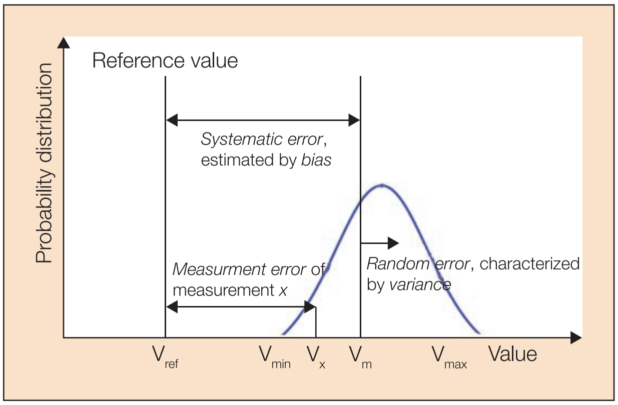

Fig. 6. The correct version of the incorrect Fig. 5.

represented by the blue curve. Let us assume that the figure corresponds to results returned by a measurement instrument, and not measurements and their results. Fig. 5(i) still has three problems:

1. It reports the offset of V_m (the average of the distribution) from the reference value V_{ref} as accuracy, instead of trueness.
2. It conveys the message that the greater the distance between V_m and V_{ref}, the greater the accuracy, which in fact is the exact opposite!
3. Similarly, it conveys the message that the greater the distance between V_{min} and V_{max}, the greater the precision, which in fact is the exact opposite!

Fig. 5(ii) has the same problem (3), as well as problem (2) but for trueness instead of accuracy.

Fig. 6 shows the corrected version of this form of visualization, which refers to components of errors, again under the assumption that a reference value is somehow known. Notice that we have also changed the label of the vertical axis from "probability density" to "probability distribution," because "density" implies the continuity of values, while measurement values are defined on a discrete scale due to finite resolution of measuring instruments.

Concluding, let us reiterate that accuracy should be part of the characterization of a measurement instrument but not part of the characterization of measurement results, since evaluating accuracy requires a reference value to be known, which is not the case for measurement results. In this latter case, measurement uncertainty is the feature that summarizes the distribution of measured values. According to the *Guide to the Expression of Uncertainty in Measurement* (GUM), *measurement uncertainty* is a *parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand* [3, 2.2.3].

In general, measurement uncertainty comprises several components that may manifest themselves with either random

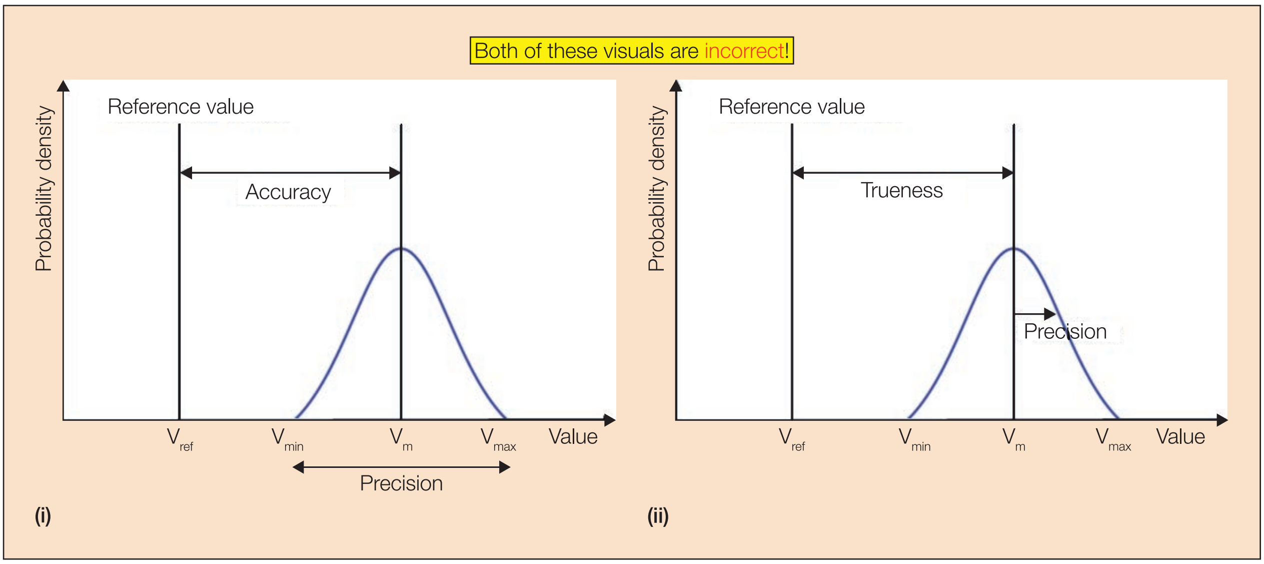

Fig. 5. Another form of incorrect visualization.

or systematic effects [4], [5]. Some components may be evaluated by applying statistical methods on the distribution of the measured values from replicated measurements (called "type A" methods [3, 2.3.2]), while other components require non-statistical methods, based on experience or a priori information (called "type B" methods [3, 2.3.3]). The accuracy of the adopted measuring instrument affects the component of measurement uncertainty called "instrumental uncertainty": the greater the accuracy of the instrument, the less the instrumental uncertainty in the results it produces. Interested readers are encouraged to read [4] for a quick tutorial on uncertainty and [5] for an overview of measurement fundamentals.

References

[1] *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms*, 3rd ed., 2012. JCGM 200:2012, Joint Committee for Guides in Metrology (VIM). [Online]. Available: http://www.bipm.org/en/publications/guides/vim.html.

[2] *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions*, ISO 5725-1, International Organization for Standardization, 1994.

[3] *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*, 2008, Joint Committee for Guides in Metrology, JCGM 100:2008. (GUM, originally published in 1993). [Online]. Available: http://www.bipm.org/en/publications/guides/gum.html.

[4] A. Ferrero and S. Salicone, "Measurement uncertainty," *IEEE Instrum. Meas. Mag*., vol. 9, no. 3, pp. 44-51, Jun. 2006.

[5] L. Mari, P. Carbone, and D. Petri, "Measurement fundamentals: a pragmatic view," *IEEE Trans. Instrum. Meas*., vol. 61, no. 8, pp. 2107-2115, Aug. 2012.

Shervin Shirmohammadi (M '04, SM '04, F '17) is currently a Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada, where he is Director of the Distributed and Collaborative Virtual Environment Research Laboratory and his research focuses on multimedia systems and networks, including measurement techniques and applied AI for networking, video streaming, and health systems. He received his Ph.D. degree in electrical engineering from University of Ottawa, Canada and currently serves as the Editor-in-Chief of *IEEE Transactions on Instrumentation and Measurement*.

Luca Mari is a Full Professor of measurement science at the Cattaneo University–LIUC, Castellanza, Italy, where he teaches courses on measurement science, statistical data analysis, and system theory. He is currently the Chairman of the TC1 (Terminology) and the Secretary of the TC25 (Quantities and units) of the International Electrotechnical Commission (IEC) and an IEC expert in the WG2 (VIM) of the Joint Committee for Guides in Metrology (JCGM). He has been the Chairman of the TC7 (Measurement Science) of the International Measurement Confederation (IMEKO).

Dario Petri (M'92, SM'05, F'09) is a Full Professor in measurement science and electronic instrumentation at the Department of Industrial Engineering of the University of Trento, Italy. He received the M.Sc. and Ph.D. degrees in electronics engineering from the University of Padua, Italy, in 1986 and 1990, respectively. He is an IEEE Fellow member and the recipient of the 2020 IEEE Joseph F. Keithley Award for "contributions to measurement fundamentals and signal processing techniques in instrumentation and measurement." Dr. Petri's research activities are focused on digital signal processing applied to measurement problems, data acquisition systems, and fundamentals of measurement theory.