Reference ranges in hemostasis testing: necessary but imperfect

Emmanuel J. Favaloro[1], Giuseppe Lippi[2]

[1]Department of Haematology, Sydney Centres for Thrombosis and Haemostasis, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Westmead Hospital, Westmead, NSW, Australia; [2]Section of Clinical Biochemistry, University of Verona, Verona, Italy

Correspondence to: Dr. Emmanuel J. Favaloro. Department of Haematology, Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, Westmead, NSW, Australia. Email: emmanuel.favaloro@health.nsw.gov.au.

Most tests performed in laboratories rely on comparison of a test result, typically a patient sample, against a range of values defined for normal healthy individuals, the concept being that any result that lies outside this range thereby defines a (potentially) 'abnormal' test result. Indeed, it is difficult to otherwise conceptualize an alternative approach in laboratory diagnostics so far. This process comprises an established standard also within hemostasis testing, and can therefore be defined as 'necessary' to establish a deviation from the norm. However, the limitations imposed by this practice are often under-recognized.

First, the reference range, sometimes called a reference interval, is often based on use of a so-called "reference population", which presumes that the reference subjects utilized in the process are all "ostensibly healthy". This will inevitably lead to collection of inaccurate data, since there is no way to accurately identify the true health of these subjects. This is something that we briefly touched on in a recent article (1). Health status is commonly inferred, for example after collection of medical history and/or physical examination, which thereby fails to show 'abnormalities', but which will never be able to fully exclude 'silent pathologies', sometimes present and potentially contributing to bias in the calculated range. Sometimes, subjects are laboratory colleagues and, here, even 'known pathologies' may not be divulged for reasons of privacy. It is obviously infeasible to perform extensive health evaluation of otherwise 'ostensibly healthy' subjects, simply to define a reference range.

Although outliers are often excluded in such processes (i.e., cutting the tails of the value distribution and conventionally including only the 95% confidence interval of the entire population, as for current indications of some international organizations such as the International Federation of Clinical Chemistry or the Clinical and Laboratory Standards Institute) (2), this does not ensure reliability. Indeed, this actually introduces another problem, since in such a process, the normal range will essentially only capture ~95% of the normal population, and by definition some 5% of test results performed on patients will be 'false positives', either falling above or below the normal reference range.

In a normally distributed normal range, around 2.5% will be false low and around 2.5% will be false high, whereas proportions will differ for non-normally distributed data. Irrespective, this can lead to potential for false diagnosis of disease where such disease does not exist. As an example, a commonly applied normal reference range for von Willebrand factor (VWF) is 50–200 U/dL, with a median value close to 100 U/dL (thus defining a non-normally distributed normal range). Deficiency of VWF can define von Willebrand disease (VWD), which is a bleeding disorder (3). However, the reference range of 50–200 U/dL is a partially 'artificial construct' and has limited utility in terms of diagnosing VWD. Most simply, a patient with a 'normal level' of 51 U/dL cannot necessarily be diagnosed as not having VWD, and another patient with an 'abnormal level' of 49 U/dL cannot necessarily be diagnosed as having VWD. Indeed, given the imprecision inherent in laboratory tests, including those for VWF, and the inter-individual variability in VWF levels, the same patient can easily give the 'different' results (49 and 51 U/dL) in the same assay run using the same sample. Moreover, as suggested above, around 1/20 test results for any individual assay based on 95% confidence intervals will reflect a false 'positive'. Thus, diagnosis of disease requires much more than simple clinical
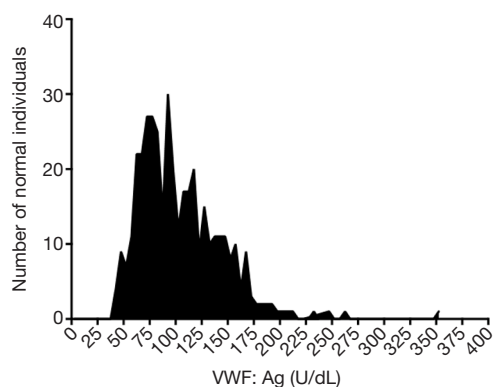
Figure 1 Distribution of VWF antigen results in a sampling of nearly 400 normal individuals (5). The distribution is non-normal, although log transformation yields a near normal distribution to then enable calculation of a 95% confidence interval for normal range estimation. For this sample set, the resulting normal range would be defined as 47.8–200.0, although this has been 'rounded' to 50–200 in the Westmead laboratory to enable a more simplified range for clinical use, especially considering additional minor variation for other VWF assays. Irrespective, there are results below 48 U/dL (n=9; ~2.3%) and below 50 U/dL (n=13; ~3.3%); thus, around 2–3% of normal individuals tested in this assay will yield results below the normal reference range (false positive). VWF, von Willebrand factor.
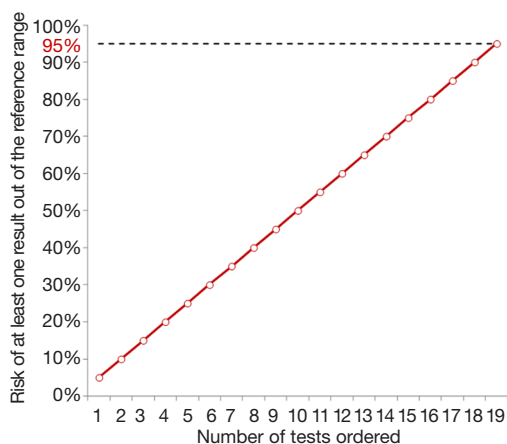
Figure 2 Probability of obtaining at least one false positive test results according to the number of tests ordered.

evaluation of test numbers. For VWD, this would mean test results from a panel on VWF tests, perhaps well below the normal reference range normal cut off (e.g., below 30 U/dL), repeated at least once for confirmation, together with clinical and family history of mucocutaneous

bleeding (3,4).

Moreover, in the real world of laboratory testing, an analysis of normal individuals is never as 'perfect' as actually enabling the generation of such a rounded normal reference range of '50–200 U/dL', given as the above example. In the Westmead laboratory, for example, we have established normal reference ranges for many VWF tests. One historical example [for VWF antigen tested by enzyme-linked immunosorbent assay (ELISA)] is shown in *Figure 1*. Here, analysis of almost 400 normal individual samples yielded the data shown, which after log transformation to approximate a normal distribution, yielded a calculated normal reference range of 47.8–200 U/dL (5), close to but not exactly the same as 50–200 U/dL. Similar data for VWF collagen binding yielded a calculated normal reference range of 52.0–231.1 U/dL. When faced with such 'minor' variance from a 'commonly' employed ('harmonious') range approximating 50–200 U/dL, laboratories will often opt for a 'rounding' to effectively avoid clinicians having to deal with different and 'difficult to recall' ranges for different tests that measure several analytes associated to a single disease (VWD in this example). This rounding assists clinicians by providing a simplified schema, but it is imperfect because it reduces the accuracy of disease diagnosis.

Naturally, the larger the number of normal individuals used in a normal range assessment process, the more accurate the normal range, and conversely, the smaller the number, the less accurate the normal range. Too small a number produces very inaccurate normal ranges. Given the difficulty for most laboratories to test large numbers of normal individuals, either because of unavailability or excessive cost, there is a balance created between these. It is then of little surprise that some variation is observed for normal ranges identified by different laboratories for the same analytes. In such cases, there is benefit to harmonization of test practice, and establishment of 'generic' normal ranges, such as 150×10^6–450×10^6/L for platelet count, even though a true normal range evaluation will never yield that 'perfect range' in every laboratory.

Just as 1/20 test results may reflect a false 'positive', this then means that the more tests performed on any given individual, the more likely the risk of a false positive in any individual patient. Essentially, if a patient has 20 tests performed, then by chance one of those tests will likely be a false positive (above or below the reference interval) for that patient. In total, this risk actually increases linearly with the number of test requested, as shown in *Figure 2*, thus exceeding 95% probability of one false positive test result

when the number of tests ordered is 19 or more.

For some test systems, for example antiphospholipid antibodies for identification of antiphospholipid syndrome (6), other limits are used for calculating reference ranges, such as 99th percentile. The advantage of such a process is reduced risk of false positive. Nevertheless, the downside may be reduced sensitivity and higher risk of false negative (i.e., patient is positive, but missed by testing which gives a result in the normal range).

The take-home message for this article is that not all abnormal test results reflect disease or abnormality, and a normal test result does not always exclude a disease. Sometimes, this just represents 'noise' around the normal-abnormal cut-off values. Sometimes, such outcomes are due to analytical error, although these are minimized in modern laboratories due to stringent quality control procedures. Sometimes abnormal test results reflect pre-analytical issues, rather than patient status, because of a compromised test sample quality, including that this has potentially been collected and processed appropriately (7,8). Another take home message here for clinicians is to avoid 'fishing expeditions', and ordering 'all' tests 'just in case'—which on occasion may occur to avoid future medico-legal issues related to a 'failure to test' for (and thus identify) a certain disease condition (9). One other recommendation is to repeat any test that does not seem to match the clinical condition, using a fresh sample collected on another occasion, and perhaps even using a different test methodology. For example, heterophile antibodies can yield false normal results in severely VWF deficient patients using some test methodologies (such as latex agglutination), whereas a different test process (e.g., ELISA) may yield the correct result (10).

Acknowledgments

Footnote

Provenance and Peer Review: This article was a standard submission to the journal. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jlpm.2017.04.05). Emmanuel J Favaloro serves as an unpaid editorial board member of *Journal of Laboratory and Precision Medicine* from April 2017 to March 2019. Giuseppe Lippi serves as the unpaid Editor-in-Chief of *Journal of Laboratory and Precision Medicine* from November 2016 to October 2021. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Favaloro EJ, Lippi G. Translational aspects of developmental hemostasis: infants and children are not miniature adults and even adults may be different. Ann Transl Med 2017. [Epub ahead of print].
2. Clinical and Laboratory Standards Institute (CLSI). Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline EP28-A3CS. Third Edition. Wayne: CLSI, 2010.
3. Curnow J, Pasalic L, Favaloro EJ. Treatment of von Willebrand Disease. Semin Thromb Hemost 2016;42:133-46.
4. Favaloro EJ, Pasalic L, Curnow J. Laboratory tests used to help diagnose von Willebrand disease: an update. Pathology 2016;48:303-18.
5. Favaloro EJ, Soltani S, McDonald J, et al. Reassessment of ABO blood group, sex, and age on laboratory parameters used to diagnose von Willebrand disorder: potential influence on the diagnosis vs the potential association with risk of thrombosis. Am J Clin Pathol 2005;124:910-7.
6. Favaloro EJ, Wong RC. Antiphospholipid antibody testing for the antiphospholipid syndrome: a comprehensive practical review including a synopsis of challenges and recent guidelines. Pathology 2014;46:481-95.
7. Lippi G, Salvagno GL, Montagnana M, et al. Quality

standards for sample collection in coagulation testing. Semin Thromb Hemost 2012;38:565-75.

8. Adcock Funk DM, Lippi G, Favaloro EJ. Quality standards for sample processing, transportation, and storage in hemostasis testing. Semin Thromb Hemost 2012;38:576-85.

9. Lippi G, Favaloro EJ, Franchini M. Dangers in the practice of defensive medicine in hemostasis testing

for investigation of bleeding or thrombosis: part I--routine coagulation testing. Semin Thromb Hemost 2014;40:812-24.

10. Favaloro EJ, Mohammed S. Towards improved diagnosis of von Willebrand disease: comparative evaluations of several automated von Willebrand factor antigen and activity assays. Thromb Res 2014;134:1292-300.