

# Reference intervals and percentiles – implications for the healthy patient

February 2018



**Suzanne Ekelund**

Principal specialist, Clinical Biochemist, MSc  
Immunoassay Business Unit  
Radiometer Medical ApS  
Denmark

## Summary

The reference interval for a biochemical analyte is usually an interval of values bounded by the reference limit values at certain designated percentiles [1].

As always when choosing a cut-off, the value will determine the trade-off between clinical sensitivity and specificity. When a low cut-off is chosen, the sensitivity increases and the specificity decreases. When a high cut-off is chosen, the specificity increases and the sensitivity decreases.

The most commonly used definition of the reference interval is the interval of values containing the central 95% of a healthy population, i.e. the reference limits are the values at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles, respectively. This could theoretically lead to the assumption that any

patient test result within the interval from the 2.5<sup>th</sup> to the 97.5<sup>th</sup> percentile is considered “normal” and any patient result outside this interval is considered “not normal”. Intervals from zero to the 95<sup>th</sup> or the 99<sup>th</sup> percentiles are also often seen.

This means that 1-5% (dependent on the designated percentiles) of healthy persons may have test results outside the reference interval and thus they could theoretically be considered “not normal”.

In consequence of the way reference intervals are defined, it is important to be aware that:

- Not all results outside the cut-offs mean that the patient is sick.
- If the patient is not sick, it does not mean that a result outside the cut-offs is wrong.

## Abbreviations and definitions

<b>CKMB</b>	Creatine kinase MB isoform
<b>FN</b>	False negative – a negative test result in a person with disease.
<b>FP</b>	False positive – a positive test result in a person without disease.
<b>LoD</b>	Limit of detection.
<b>Percentile</b>	Each of the 100 equal groups into which a population can be divided according to the distribution of values of a particular variable.
<b>Sensitivity</b>	Fraction of persons with disease who get a positive test result with the assay in question.
<b>Specificity</b>	Fraction of persons without disease who get a negative test result with the assay in question.
<b>TN</b>	True negative – a negative test result in a person without disease.
<b>TP</b>	True positive – a positive test result in a person with disease.

## Background

When you perform a diagnostic biochemical test in a patient, you will compare the patient result to a reference interval or to a medical decision limit.

To determine a reference interval you need to test a large number of healthy people, the reference population. Health is a relative condition lacking a universal definition. Before collecting samples for making a reference interval, inclusion and exclusion criteria for participation have to be established. Furthermore, samples must be collected in a way that minimizes the risk of preanalytical errors.

The reference interval for a biochemical analyte is usually the central interval of values bounded by the reference limit values at certain designated percentiles [1, 2]. That is, the reference interval refers to that interval set of values observed in the reference sample group or predicted for the reference population, defined by a specific percentage.

### Distribution of test results from healthy populations

The test results from a healthy population may form a normal distribution, as in Fig. 1. For the

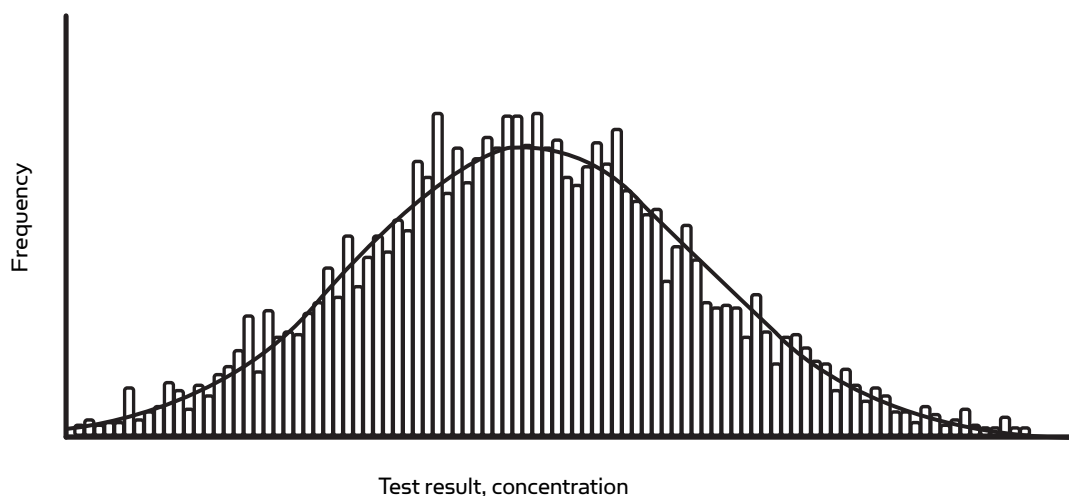


FIG. 1: Example – distribution of test results in a healthy population.

sake of simplicity, the examples in the following will focus on this type of distribution.

The results may be shown in a histogram or its normal curve overlay.

As mentioned above, the reference interval refers to an interval set of values observed in the reference sample group or predicted for the reference population, defined by a specific percentage [1].

The most commonly used definition of the reference interval is the interval of values containing the central 95% of a healthy population, i.e. the reference limits are the values at the 2.5<sup>th</sup>

and 97.5<sup>th</sup> percentiles, respectively. It means that any patient result within the interval from the 2.5<sup>th</sup> to the 97.5<sup>th</sup> percentile is per definition considered “normal” and any patient result outside this interval is per definition considered “not normal”. In cases where the distribution is nonparametric (not normal distribution), the percentiles can still be used as limits as in Fig. 3.

If the lower limit is close to or cannot be distinguished from zero, the reference interval chosen might be from zero to the 95<sup>th</sup> percentile as in Fig. 4. In this case any patient result equal to or below the 95<sup>th</sup> percentile is theoretically assumed to be “normal” and any patient result above this value to be “not normal”.

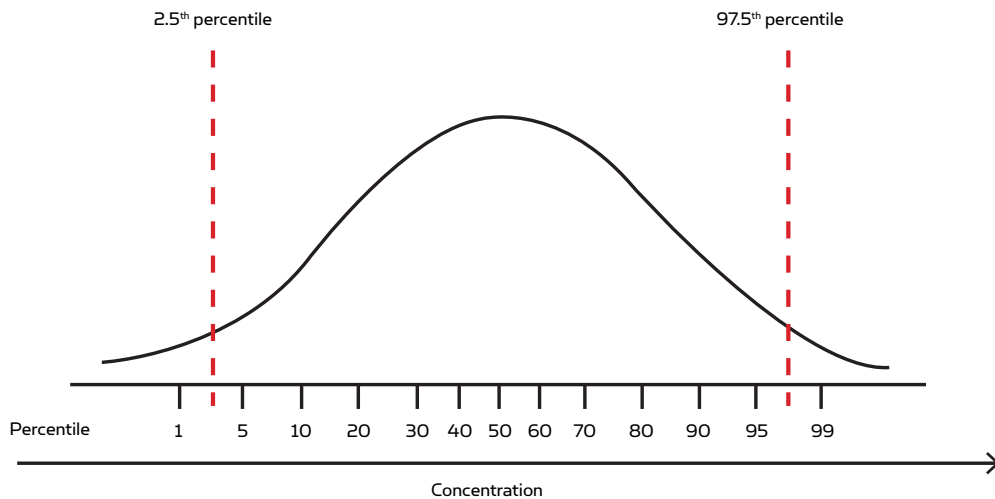


FIG. 2: Example - distribution of test results in a healthy population with reference limits 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles, respectively.

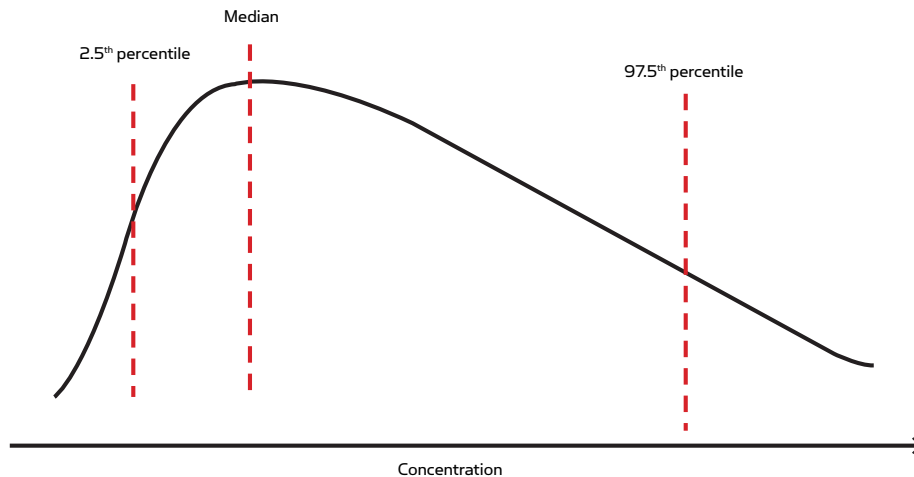


FIG. 3: Example - nonparametric distribution of test results in a healthy population with reference limits 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles, respectively (adapted from [7]).

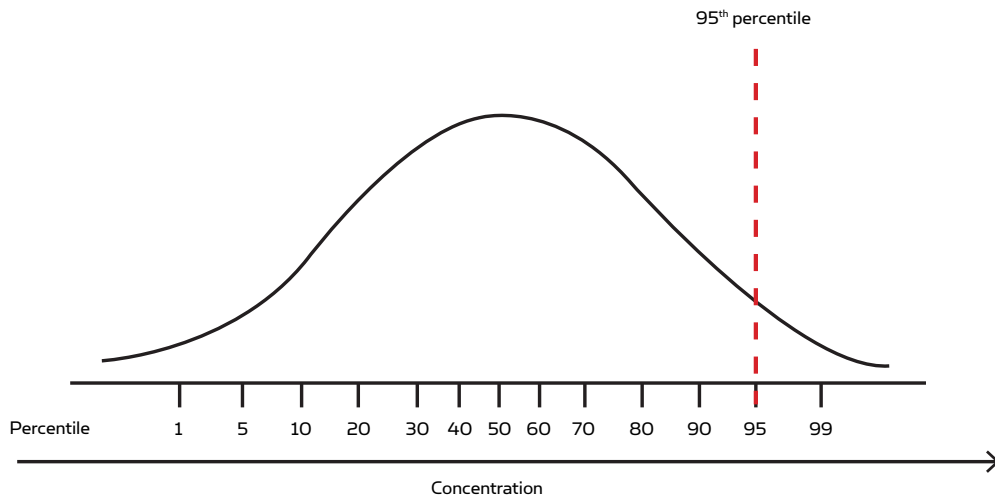


FIG. 4: Example – distribution of test results in a healthy population with no lower limit and upper limit 95<sup>th</sup> percentile.

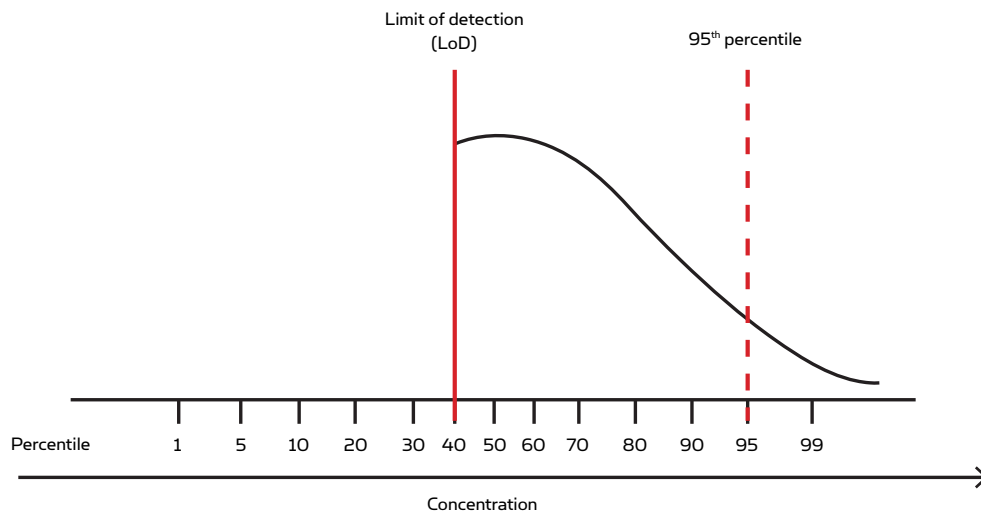


FIG. 5: Example – distribution of test results in a healthy population with no lower limit but many results below the limit of detection, and upper limit 95<sup>th</sup> percentile.

Many methods are not able to give measurable results in all healthy subjects and a fraction of the healthy subjects will get the result  $<[LoD]$ . Then the distribution can look like Fig. 5. And as above, any patient result equal to or below the 95<sup>th</sup> percentile is assumed to be “normal” (including those with the result  $<[LoD]$ ), and any patient result above this value “not normal”.

A few analytes use as upper limit the 99<sup>th</sup> percentile. This is the case for e.g. the cardiac markers troponin I, troponin T and CKMB. It is

even recommended in guidelines to use the 99<sup>th</sup> percentile for these analytes [3]. In this case, any patient result equal to or below the 99<sup>th</sup> percentile is assumed to be “normal” and any patient result above this value “not normal”.

### Why are extreme results eliminated from reference intervals?

Now, why is the upper limit of a reference interval defined as a percentile below 100%?

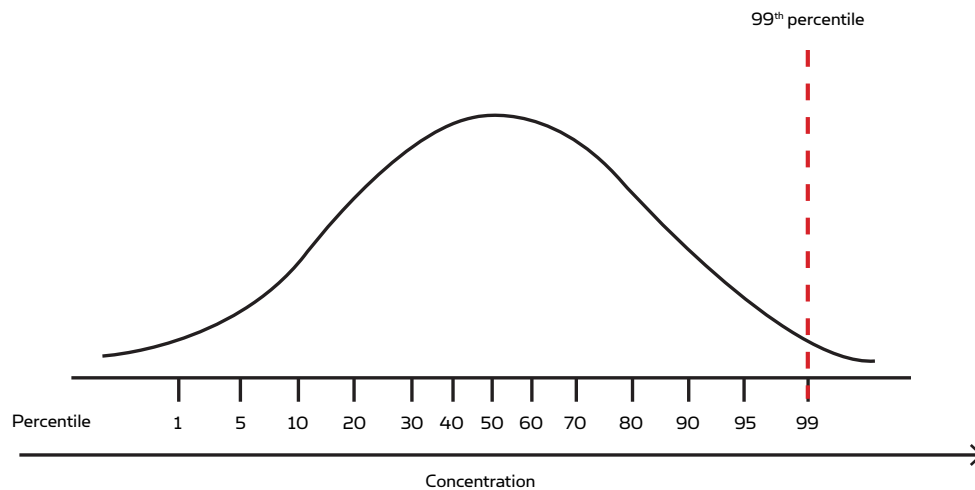


FIG. 6: Example – distribution of test results in a healthy population with no lower limit and upper limit 99<sup>th</sup> percentile.

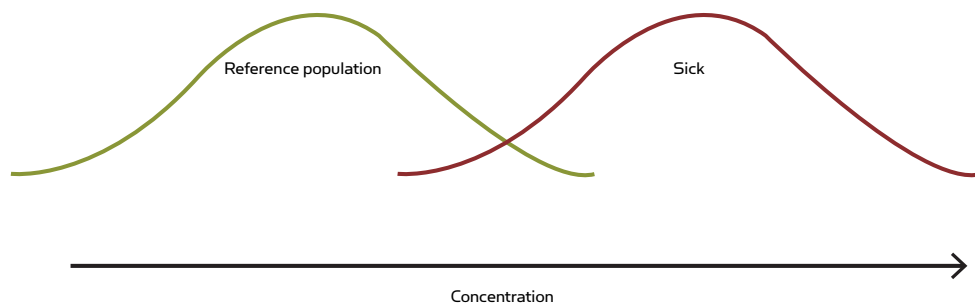


FIG. 7: Example – distributions of test results in a healthy population and in a sick population.

A reference interval should not rely on samples with extreme results. Outliers, if they exist, will occur in that fraction of the samples [4]. In addition, it appears that for certain analytes an important contributor to results in the high end may be subclinical disease [5]. Upper limits for these analytes can be lowered by further screening of the reference individuals; e.g. for troponin it has been shown that adding estimated glomerular filtration rate, NT-proBNP results and echocardiography for inclusion/exclusion can decrease the determined 99<sup>th</sup> percentile by approximately 50% [6, 7]. However, elimination of extreme results in reference interval studies by the use of more extensive screenings is expensive and therefore it might not be applied for economic reasons.

### Implications of eliminating extreme results from reference intervals

The distribution of test results from a sick population will for the many analytes have some overlap with the distribution of test results from the group of healthy reference individuals. An example is shown in Fig. 7.

The percentile used as cut-off determines the clinical sensitivity (fraction of patients with true-positive (TP) results in the sick group) and thereby the fraction of patients with false-negative (FN) results in the sick group. It also determines the clinical specificity (fraction of patients with true negative (TN) results in the healthy group) and thereby the fraction of patients with false positive

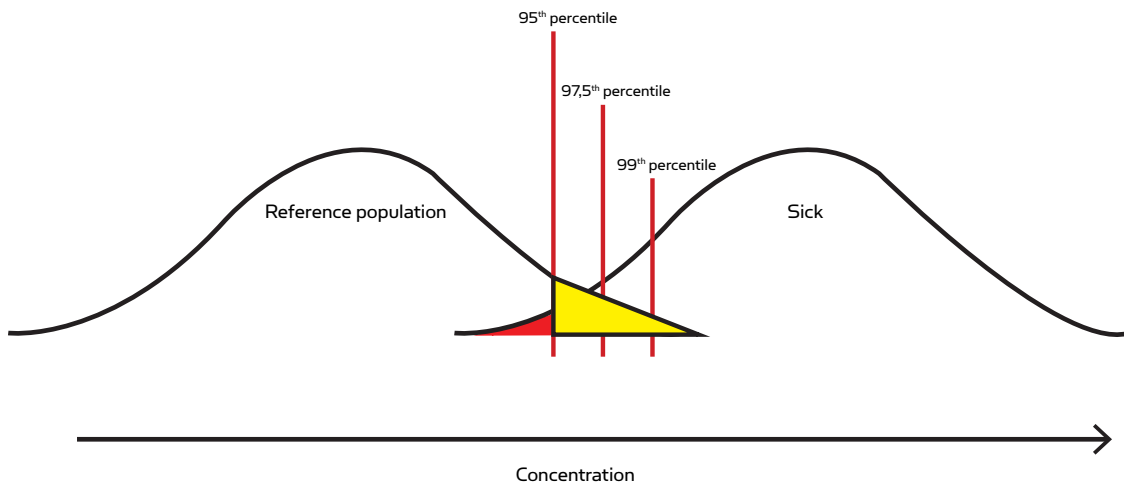


FIG. 8: Example - when the 95<sup>th</sup> percentile is the upper limit of the reference interval, the healthy persons in the yellow part of the distribution will get a false-positive result, and the sick persons in the red part of the distribution will get a false-negative result.

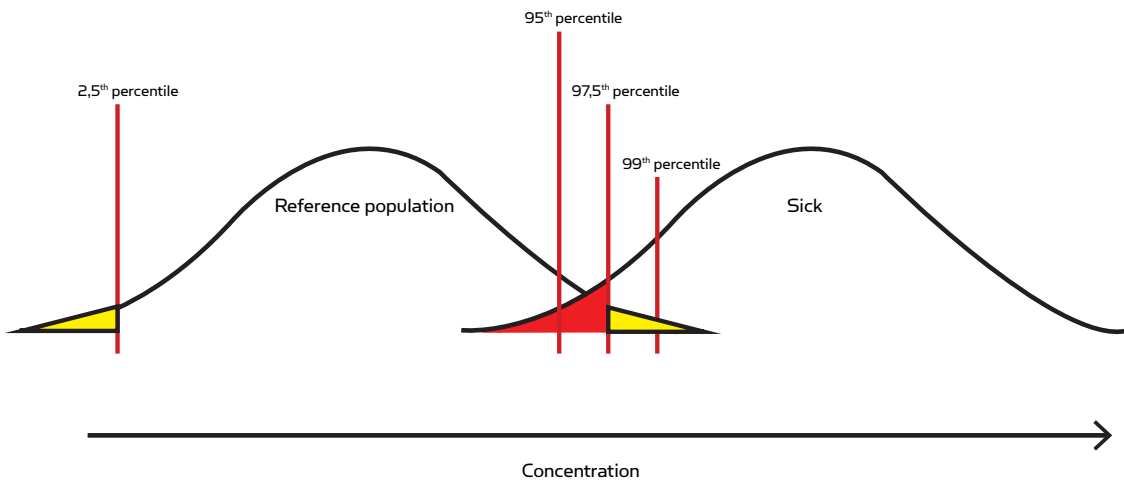


FIG. 9: Example - when the 2.5<sup>th</sup> percentile is the lower limit and the 97.5<sup>th</sup> percentile is the upper limit of the reference interval, the healthy persons in the yellow parts of the distribution will get a false-positive result, and the sick persons in the red part of the distribution will get a false-negative result.

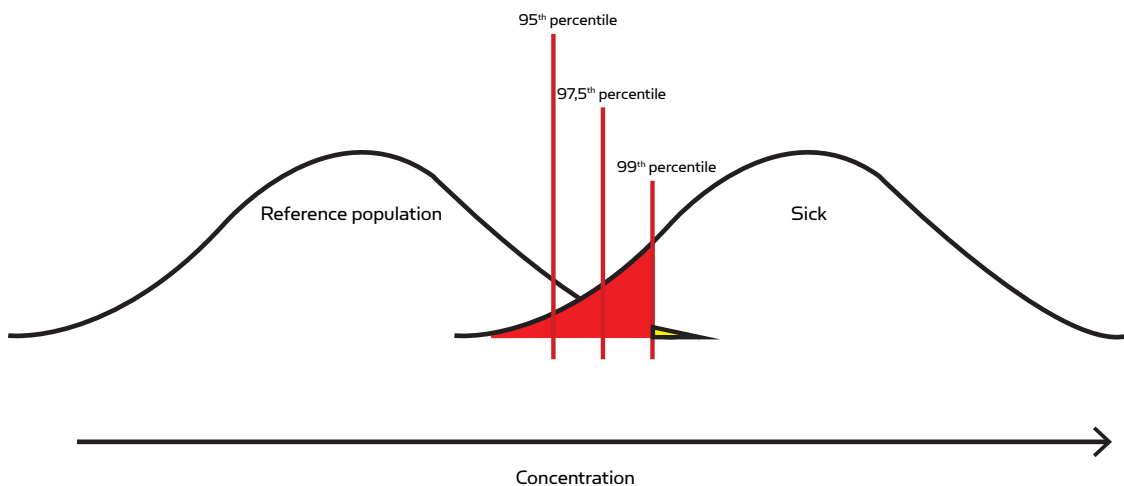


FIG. 10: Example - when the 99<sup>th</sup> percentile is the upper limit of the reference interval, the healthy persons in the yellow part of the distribution will get a false-positive result, and the sick persons in the red part of the distribution will get a false-negative result.

(FP) results in the healthy group. The choice of percentile/cut-off is always a trade-off between sensitivity and specificity.

As can be seen from Fig. 8, Fig. 9 and Fig. 10, the higher the upper percentile that is used for cut-off, the lower the number of healthy persons that will get a false-positive result. However, the higher the upper percentile that is used for cut-off, the higher the number of sick persons that will get a false-negative result.

Usually we want to have as high a clinical sensitivity as possible, meaning that we will prefer a lower cut-off to avoid missing a diagnosis in a sick patient and paying the price of increasing the number of healthy persons who get a false-positive result.

In a few cases, like for CKMB, troponin I and troponin T, the 99<sup>th</sup> percentile is used as cut-off because it is recommended by guidelines [2]. The use of the 97.5<sup>th</sup> percentile, which is the most common upper limit for clinical biochemistry tests, as cut-off was proposed by the National Academy of Clinical Biochemistry already in 1999 [4]. However, this would likely cause anxiety in the 2.5% of healthy individuals, who would get a result assumed to be “not normal”. Furthermore, it could have insurance implications for these patients. Therefore the 99<sup>th</sup> percentile is used for these analytes.

Results outside a reference interval are commonly assumed to be “not normal” [8, 9]. However, many and among them Guidi (2006) claim [7] that this is a conceptual mistake because:

- All reference individuals were selected based on the same predetermined criteria.
- All the results either far from or around the reference limits are only punctual representation of the biological variation.
- The analytical imprecision will have an influence on the actual results.

These facts are well known and understood in laboratory settings, but they may be underestimated in the clinical practice.

When the cut-off is at the 95<sup>th</sup> percentile, it means that 5% of healthy persons will get a result that is above the reference interval. In other words, a healthy person has a 1 out of 20 chance of having a so-called false-positive test result. If two tests are ordered, the probability that the second test is  $\leq 95^{\text{th}}$  percentile is also 95%, but the probability that this is the case for both tests is  $0.9025 = 0.90 \sim 90\%$ . If 10 tests are ordered, which often occur, the probability that all are  $\leq 95^{\text{th}}$  percentiles is  $0.60 \sim 60\%$  [10].

In consequence of the way reference intervals are defined, it is important to be aware that:

- Not all results outside the cut-offs mean that the patient is sick.
- If the patient is not sick, it does not mean that a result outside the cut-offs is wrong.

## References

1. CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline – Third Edition. CLSI document EP28-A3c. Wayne, PA: Clinical and Laboratory Standards Institute; 2008. Corrected version October 2010
2. Higgins C. An introduction to reference intervals (1) - some theoretical considerations. [www.acutecaretesting.org](https://www.acutecaretesting.org/en/articles/an-introduction-to-reference-intervals-1--some-theoretical-considerations) Jan 2010. <https://www.acutecaretesting.org/en/articles/an-introduction-to-reference-intervals-1--some-theoretical-considerations>
3. Thygesen K, Alpert JS, Jaffe AS, *et al.* Third universal definition of myocardial infarction. *J Am Coll Cardiol* 2012; 60: 1581–598
4. Horn PS, Pesce, AJ. Reference intervals: an update. Review. *Clin Chim Acta* 2003; 334: 5-23
5. Hickman PE, Lindahl B, Potter JM *et al.* Is it time to do away with the 99<sup>th</sup> percentile for cardiac troponin in the diagnosis of acute coronary syndrome and the assessment of cardiac risk? *Clin Chem* 2014; 60(5): 734-736
6. Collinson PO, Heung YM, Gaze D, *et al.* Influence of population selection on the 99<sup>th</sup> percentile reference value for cardiac troponin assays. *Clin Chem* 2012; 58: 219 –225
7. Koerbin G, Abhayaratna WP, Potter JM, *et al.* Effect of population selection on 99<sup>th</sup> percentile values for a high sensitivity cardiac troponin I and T assays. *Clin Biochem* 2013; 46: 1636–643
8. Guidi GC, Lippi G, Solero GP, *et al.* Managing transferability of laboratory data. Review. *Clin Chim Acta* 2006; 374(1-2): 57-62
9. Guidi GC, Salvagno GL. Reference intervals as a tool for total quality management. Special issue: Quality in laboratory diagnostics: from theory to practice. *Biochimica Medica* 2010; 20(2): 165-172
10. Islin H. Interpretation of laboratory results. [www.acutecaretesting.org](https://www.acutecaretesting.org/en/articles/interpretation-of-laboratory-results) Jan 2010. <https://www.acutecaretesting.org/en/articles/interpretation-of-laboratory-results>