Lessons in biostatistics

Resampling methods in Microsoft Excel® for estimating reference intervals

Elvar Theodorsson

Department of Clinical Chemistry and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

Corresponding author: elvar.theodorsson@liu.se

Abstract

Computer- intensive resampling/bootstrap methods are feasible when calculating reference intervals from non-Gaussian or small reference samples. Microsoft Excel® in version 2010 or later includes natural functions, which lend themselves well to this purpose including recommended interpolation procedures for estimating 2.5 and 97.5 percentiles.

The purpose of this paper is to introduce the reader to resampling estimation techniques in general and in using Microsoft Excel® 2010 for the purpose of estimating reference intervals in particular.

Parametric methods are preferable to resampling methods when the distributions of observations in the reference samples is Gaussian or can transformed to that distribution even when the number of reference samples is less than 120. Resampling methods are appropriate when the distribution of data from the reference samples is non-Gaussian and in case the number of reference individuals and corresponding samples are in the order of 40. At least 500-1000 random samples with replacement should be taken from the results of measurement of the reference samples.

Key words: reference interval; resampling method; Microsoft Excel; bootstrap method; biostatistics

Description of the task

Reference intervals (1,2) are amongst the essential tools for interpreting laboratory results. Excellent theoretical and practical studies have been published since the 1960ies for establishing reference intervals as a concept in contrast to "normal intervals" (3), establishing principles for selecting reference persons (4) and performing the appropriate statistical data analysis (5-8). Reference intervals continue to be an active area of research (2,9) and at the core of practical work in clinical laboratories (6,10).

The reference interval includes 95% the results/ values measured in a representative sample of reference subjects and is bounded by lower and upper reference limits marking the 2.5 and 97.5 percentiles respectively (Figure 1). The reference interval including the upper and lower reference limits are naturally determined in a ranked set of data values if the number of observations is large, e.g. > 1000. When the number of values is small, e.g. in the order of 40-120 and the data are non-Gaussian, resampling methods are useful to estimate the reference interval. The reference population is commonly a population of apparently healthy individuals, but may also consist of any other well-defined population of interest for diagnostic/comparative purposes.

Different methods for calculating reference intervals

Estimating reference intervals means dealing with uncertainties and probabilities. All probabilistic methods are based on assumptions about a theo-
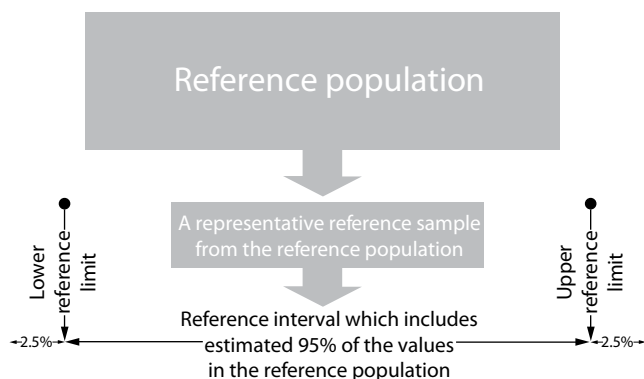
Figure 1. General principles when calculating reference intervals.

retical distribution which fundamentally determines the conclusion that can be drawn from the data. The most commonly used probabilistic methods are the parametric methods that assume that observations in the population are distributed according to the Gaussian/Normal distribution. These are the methods of choice if the data are Gaussian or can be transformed to that distribution since the data themselves with the added knowledge of the distribution of the data in the sample and population enables the user of the data to draw firmer conclusions than if only the data are known.

When the data are non-Gaussian and cannot be transformed to the Gaussian distribution the analyst is left with non-parametric or resampling methods for determining reference intervals. Data from 120 reference intervals are needed for reliably determining non-parametric reference intervals (7,11). Resampling/bootstrap methods (12) take numerous repeated sub- samples with replacement of the available data in order to estimate the distribution of the data in the population, including the reference intervals.

Resampling methods with replacement are commonly called bootstrap methods by many believed to refer to the fairytale of Baron Munchausen pulling himself and his horse out of a swamp by his hair. Bootstraps are in fact the tab, loop or handle at the top of boots allowing the use of fingers pull the boots on. The concept of bootstrap is therefore a metaphor for clever, self-induced salvaging efforts. There is a wealth of current intro-

ductory literature on resampling methods in general (13) and resampling methods for Microsoft Excel® in particular (14).

Resampling with replacement means that a large number of samples with replacement are then taken from the original sample and the statistic of interest is calculated from this pseudo-population as estimate of the corresponding parameter of the population (Figure 2).
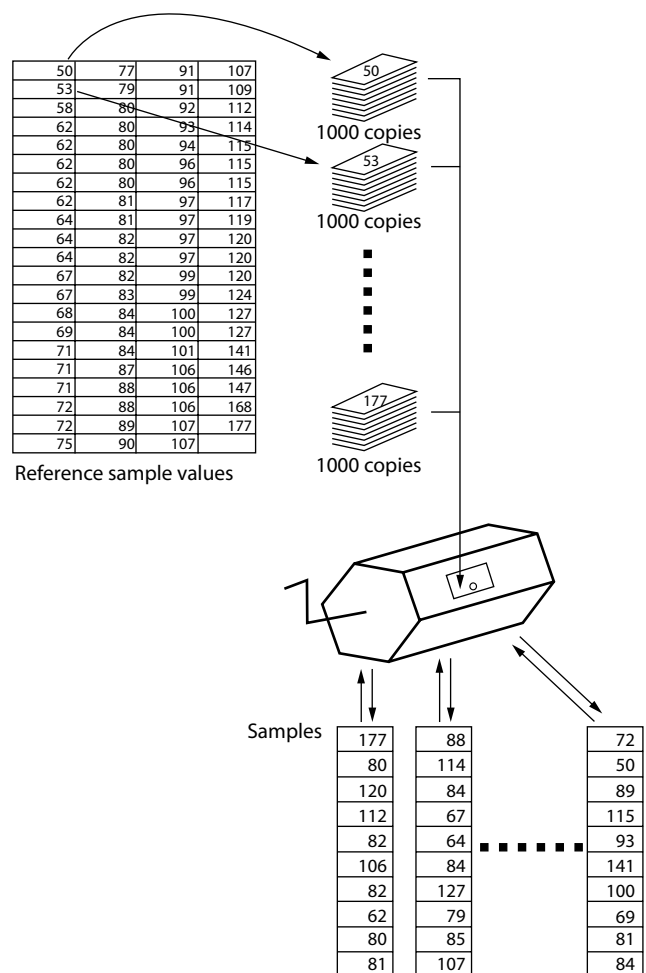
Figure 2. Simplified illustration of the principles of resampling techniques as employed by resampling methods with replacement.
This figure illustrates the use of 83 raw data observations obtained from Geffré *et al*. (24) resampled 1000 times in this case. One thousand copies of each observation are created and thoroughly mixed giving each of them identical possibility of being selected. A multitude of random samples are then taken from the mixture, calculating the 0.025 and 0.975 percentiles for each sample in order to create an estimate of these percentiles in the intended population.

Percentiles

A proper start when calculating reference intervals is to estimate the observation(s) corresponding to 2.5% and 97.5% in a ranked list of reference value data. Both International Federation of Clinical Chemistry (IFCC) (5-7,15) and Clinical and Laboratory Standards Institute (CLSI) (16) recommend the following formulas: lower limit has the rank number 0.025 x (n+1) and the upper limit the rank number 0.975 x (n+1). This method is practicable mainly when the number of reference samples is 120 or more.

While there is agreement on how to calculate the median, agreement on how to calculate quantiles including percentiles is lacking due to the fact that interpolation is frequently needed, especially when the number of observations is small. Unfortunately there is no general agreement on the best way of performing interpolations. In 1996 Hyndman and Fan published an influential paper on calculating quantiles (17). It evaluated the methods used by popular statistics packages with the intention to find a consensus and basis for standardization. Of the nine formulas commonly used, four satisfied five of the six properties desirable for a sample quantile. Hyndman and Fan felt that the "Linear interpolation of the approximate medians for order statistics" (used by the Excel =PERCENTILE.EXC function) was best due to the approximately median-unbiased estimates of the quantiles, regardless of the distribution (17).

Unfortunately only a limited movement towards standardization has taken place since 1996 (18). Most statistical programs stick to their traditions when calculating quantiles probably to maintain backwards computability with older versions.

CLSI has published an authoritative and practical standard "Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline EP28-A3c" (11) and the whole subject is authoritatively summarized by Harris and Boyd (7).

Several excellent software packages are available for the statistical analysis of reference intervals (5,19-23) but they are not as commonly used as they should e.g. due to cost constraints. Recently

Geffré et al. published "Reference Value Advisor", a set of comprehensive freeware macros for Microsoft Excel® to calculate reference intervals (24). It – together with the original paper - can be downloaded at http://www.biostat.envt.fr/spip/spip.php?article63 which implements the procedures recommended e.g. by EP28-A3c. Increased availability of low-cost computing machinery and resampling methods has made distribution independent computer intensive methods accessible in routine work (25).

The fundamental purpose of statistics in general and statistics in reference interval estimation is to extrapolate from a limited number of observations/data to the whole population of observations. All statistical methods are based on assumptions which fundamentally decide the conclusion that can be made from the data. Parametric statistical methods which are the most widely used, assume that the population values are distributed according to a mathematically well-defined distribution - the Gaussian/Normal distribution. Parametric methods were initially described in the first half of the twentieth century when methods for performing extensive calculations were primarily manual and semi-automated at best. Using distribution functions to describe natural phenomena including research data, substantially simplified the calculations required for statistical analysis. Without these known statistical distribution functions it would have been impossible to use advanced statistical methods in practice.

The wide availability of low-cost computing power in the 1980ies substantially changed the options for working with the data as they turn out irrespective of theoretical distributions. A seminal work on resampling methods by Jones in 1956 (26) was followed by the very influential works by Efron et al. (12,25) which brought these methods in the mainstream of data analysis.

The purposes of the present paper are to provide detailed instructions on the use of distribution-independent computer intensive resampling/bootstrap methods using standard functions of Microsoft Excel® 2010 or later for estimating reference intervals and to put resampling methods into a theoretical and practical perspective.

Description of the resampling method

Microsoft Excel® 2013 for Windows was used on a personal computer running Windows 7 operating system on an Intel® Core® i7-3520M CPU 2.90GHz processor and 16 GB RAM. The 83 reference results were obtained from a data set of reference values of creatinine published by Geffré *et al.* (24).

A simple method based on ranked results was initially used (27,28). The results were sorted using the sorting function in Microsoft Excel® and the rank numbers of the reference interval computed using the following formulas: lower limit has the rank number 0.025 x (n+1) and the upper limit the rank number 0.975 x (n+1) where n is the total number of reference samples.

The first randomly drawn reference value

The first randomly drawn observation in the first cell C1 was created by the following function =INDEX($A$1:$A$10000,RANDBETWEEN(1,COUNT($A$1:$A$10000))) (Figure 2).

The INDEX function in Microsoft Excel® is used to locate a single data value amongst data in a table by specifying the row and column numbers corresponding to the data point to be retrieved. The format is = INDEX(range_to_search_in, row, column).

The $ sign before the row and column numbers is used to mark the fixed data column and row to be used (Figure 2).

The COUNT function in Microsoft Excel® totals the number of cells in a selected range of cells that contain a specific type of data, e.g. cells containing data in a range of cells that can contain empty cells. In this case it counts the number of observations entered between row 1 and 10 000 - COUNT($A$1:$A$10000). It is used here in order to eliminate the need for the user to keep track of the number of reference values in the reference sample.

RANDBETWEEN(1,COUNT($A$1:$A$10000))

picks a random number from the reference sample values entered in column A, in this case 83 reference values. You can speed up the calculation by giving the number of data points in your particular data set, e.g. in this case by RANDBETWEEN(1,83) or the complete formula =INDEX($A$1:$A$83,RANDBETWEEN(1,83)). The increased speed is caused by the fact that Excel then does not need to repeatedly determine the number of reference values you have entered.

Subsequent randomly drawn reference values

Drag the small quadrant in the lower right corner of cell C1 to column AG83 thereby creating 30 samples, each of 83 random observations from the original observations. Extending the columns to the right you can increase the number of samples e.g. to 1000 random samples.

Calculating the 0.025 and 0.975 percentiles

Calculate the percentiles, e.g. in row 85 and row 86, respectively using the formulas =PERCENTILE.EXC(I1:I83;0.025) and =PERCENTILE.EXC(I1:I83;0.975), respectively (Figure 2). The PERCENTILE.EXC function, available only in 2010 – or later versions of MS Office is used in order to implement a recommended method for percentile value interpolation (17). The Microsoft Excel® simple = PERCENTILE function will not work for the present purpose (see below). The error "#NUM!" in the PERCENTILE.EXC function occurs if the supplied value of k is is < 1/(n+1) or > n/(n+1) (where n is the number of reference sample values) or - the supplied array is empty. This means that random samples of at least 39 observations are needed to calculate the 2.5 and 97.5 percentiles.

Calculate the medians of the random sample estimates of the population percentiles

Lastly calculate the median of the samples estimates of the population 0.025, and 0.975 percentiles using the median function e.g. = MEDIAN(C85:AG85) and =MEDIAN(C86:AG86) (Figure 2).

Detailed instructions to create the calculation sheet are presented in table 1.

30 samples of 83 observations each resulted in the reference interval 53.5-165.8, 333 and 1000 samples of 83 observations each and resulted identical reference intervals of 53.9-147. In comparison the reference interval calculated by Geffré *et al.* was 53.4-146.9 (24).
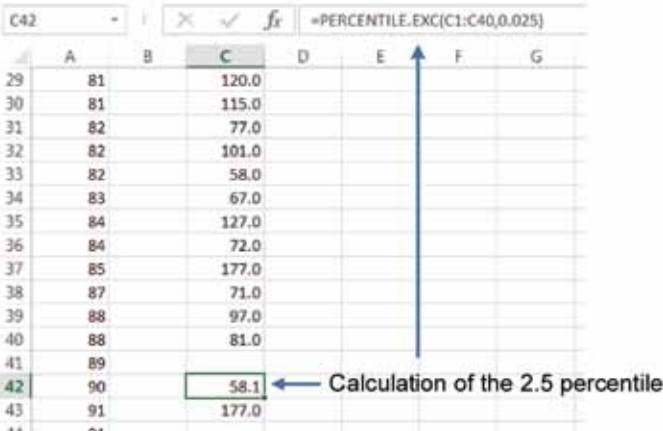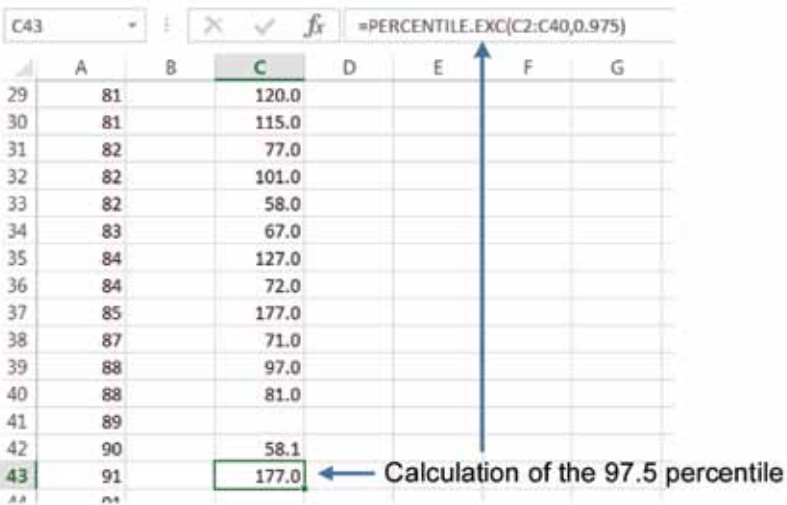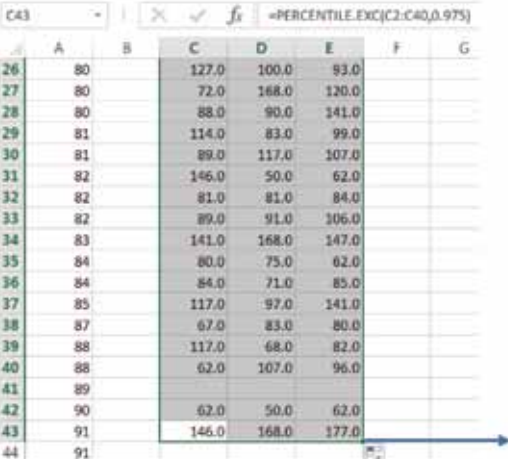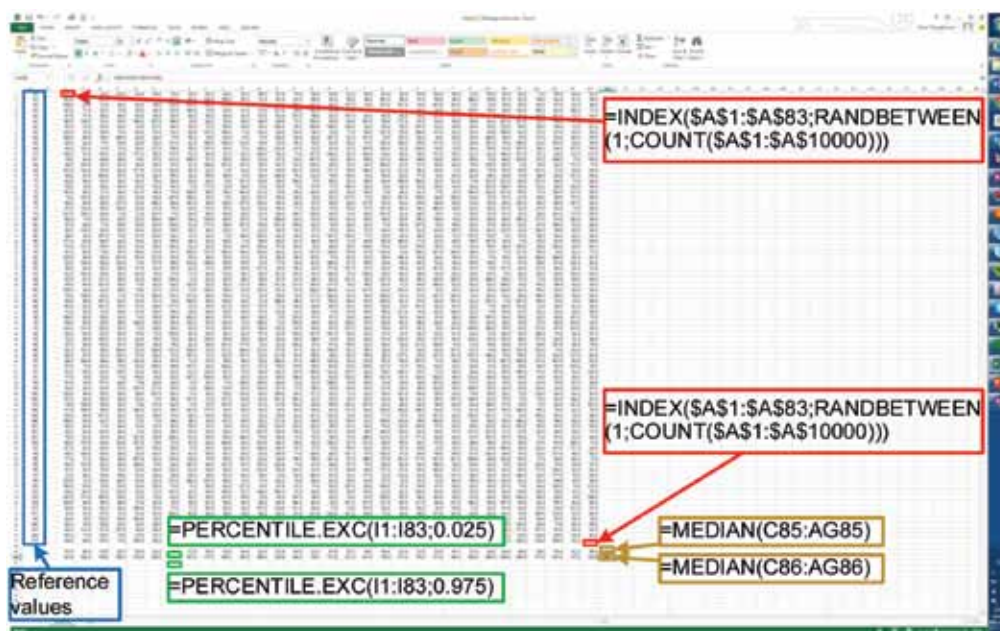
FIGURE 3. An example showing 30 random samples drawn from the 83 reference values shown in Figure 2.
Only 30 random samples are shown here in order to be able to show the principle in a single figure/screen. The 83 reference values are shown in the column on the far left (column A) marked with a blue box.

TABLE 1. Detailed instructions to create the calculation sheet.

Step	Instruction
Input the data from the reference persons	In column A. As many data values as needed up to 10 000 data points, each in its own cell |	A
1	50
2	53
3	58
4	62
5	62
6	62
7	62
8	62
9	64
10	64	
Selecting the random samples from the data from the reference persons	• In column C, row 1 enter =INDEX($A$1:$A$10000;RANDBETWEEN(1;COUNT($A$1:$A$10000))) • Select the solid black quadratic box in the lower right corner of column c, row 1 and pull it downwards the number of rows corresponding to the number of observations you want in each sample from the reference data This will copy the cells downwards, as many as you wish, making e.g. random picks of 40 from the samples from the reference persons. It is crucial for this application that you make at least 39 random picks.

TABLE 1. Detailed instructions to create the calculation sheet.

Step	Instruction
Calculate the 2.5 percentile	 Calculation of the 2.5 percentile
Calculate the 97.5 percentile	 Calculation of the 97.5 percentile
Create several random samples	Copy the column C you have created to the right to make as many random samples from the reference sample as you wish: If you wish to create 500 random samples, draw/copy 500 columns to the right and if you wish to create 1000 random samples, 1000 columns to the right

Step	Instruction
Calculate the median of the 2.5 percentiles	`=MEDIAN(C42:BC42)` AT, AU, AV, AW, AX, AY, AZ, BA, BB, BC, **BD** 115.0 115.0 168.0 82.0 80.0 106.0 92.0 115.0 75.0 101.0 80.0 88.0 69.0 168.0 94.0 58.0 69.0 72.0 120.0 88.0 85.0 64.0 107.0 71.0 115.0 147.0 94.0 97.0 62.0 64.0 64.0 91.0 106.0 82.0 109.0 82.0 120.0 62.0 93.0 147.0 62.0 68.0 71.0 101.0 80.0 97.0 177.0 84.0 109.0 62.0 79.0 141.0 80.0 99.0 72.0 109.0 64.0 71.0 88.0 64.0 97.0 68.0 72.0 112.0 71.0 77.0 96.0 99.0 71.0 85.0 87.0 58.0 127.0 71.0 67.0 62.0 64.0 89.0 120.0 71.0 114.0 81.0 71.0 120.0 91.0 75.0 50.0 83.0 100.0 106.0 97.0 107.0 62.0 82.0 77.0 146.0 64.0 168.0 80.0 115.0 109.0 75.0 79.0 115.0 177.0 107.0 115.0 119.0 81.0 97.0 97.0 107.0 82.0 119.0 106.0 115.0 71.0 93.0 97.0 85.0 97.0 114.0 67.0 91.0 81.0 92.0 88.0 83.0 107.0 91.0 89.0 90.0 97.0 120.0 67.0 89.0 72.0 82.0 80.0 141.0 62.0 72.0 106.0 106.0 147.0 107.0 80.0 114.0 91.0 141.0 62.0 53.1 50.3 62.1 53.2 58.1 50.3 53.2 62.0 53.2 **53.2** 147.0 168.0 168.0 168.0 177.0 147.0 177.0 168.0 177.0 168.0 **177.0** If the mean is preferred instead of the median, the function AVERAGE should be used instead of MEDIAN
Calculate the median of the 97.5 percentiles	`=MEDIAN(C43:BC43)` AT, AU, AV, AW, AX, AY, AZ, BA, BB, BC, **BD** 115.0 115.0 168.0 82.0 80.0 106.0 92.0 115.0 75.0 101.0 80.0 88.0 69.0 168.0 94.0 58.0 69.0 72.0 120.0 88.0 85.0 64.0 107.0 71.0 115.0 147.0 94.0 97.0 62.0 64.0 64.0 91.0 106.0 82.0 109.0 82.0 120.0 62.0 93.0 147.0 62.0 68.0 71.0 101.0 80.0 97.0 177.0 84.0 109.0 62.0 79.0 141.0 80.0 99.0 72.0 109.0 64.0 71.0 88.0 64.0 97.0 68.0 72.0 112.0 71.0 77.0 96.0 99.0 71.0 85.0 87.0 58.0 127.0 71.0 67.0 62.0 64.0 89.0 120.0 71.0 114.0 81.0 71.0 120.0 91.0 75.0 50.0 83.0 100.0 106.0 97.0 107.0 62.0 82.0 77.0 146.0 64.0 168.0 80.0 115.0 109.0 75.0 79.0 115.0 177.0 107.0 115.0 119.0 81.0 97.0 97.0 107.0 82.0 119.0 106.0 115.0 71.0 93.0 97.0 85.0 97.0 114.0 67.0 91.0 81.0 92.0 88.0 83.0 107.0 91.0 89.0 90.0 97.0 120.0 67.0 89.0 72.0 82.0 80.0 141.0 62.0 72.0 106.0 106.0 147.0 107.0 80.0 114.0 91.0 141.0 62.0 53.1 50.3 62.1 53.2 58.1 50.3 53.2 62.0 53.2 **53.2** 147.0 168.0 168.0 168.0 177.0 147.0 177.0 168.0 177.0 168.0 **177.0** If the mean is preferred instead of the median, the function AVERAGE should be used instead of MEDIAN

The computer time needed for re-calculating 1000 samples of 83 observations each from a sample of 83 observations was less than 1 second.

Discussion

The resampling model built by standard functions of Microsoft Excel® 2013 and illustrated here resulted in the practically identical reference interval calculated by Microsoft Visual Basic for Applications and described by Geffré *et al*. (24).

Resampling methods are free from the assumption that the observations are distributed according to a certain theoretical distribution, but importantly assume that the underlying population distribution is practically the same as that in a particular sample. This means that a sufficient number of observations is needed in the sample to make sure that it represents the population. Tu and Zhang have found that applying resampling methods to a random sample of 50 from true Gaussian distribution only resulted in coverage of the true confidence interval estimate in 88% of the cases, even though the improved bias-corrected-and-accelerated bootstrap (29) is used. Good (13) recommends in the order of 100 observations in the sample which is close to the 120 individuals/observations recommended by IFCC (8,30) and the CLSI (16). Geffré *et al*. (24) found that resampling methods are not appropriate when the number of reference samples less than 40. Coscun *et al*. (31) calculated reference intervals for glucose, creatinine, blood urea nitrogen (BUN), and triglycerides from random samples from 20 to 120 individuals using

resampling methods. Their results confirm that 40 reference samples represent a prudent minimum. At least 500 but preferably 1000 or more (31) resamples are generally recommended which is supported by the results of the present study.

Bjerner *et al.* (32) have shown that parametric methods may be preferable to resampling methods even when the number of reference samples is less than 120 when the distributions of observations in the reference samples is Gaussian or can transformed to that distribution.

Microsoft Excel®

Microsoft Excel® is currently the most widely used and powerful spreadsheet program. Its widespread availability and versatility constitutes its major advantages enabling the users to perform powerful descriptive, inferential graphical and statistical analysis using even methods not yet programmed in statistical packages. The program has, especially in its earlier versions been criticized for its mathematical (33), statistical (34) and graphical properties. Several of the limitations have been overcome in the most recent versions of Excel (35), since 2010.

Is Microsoft Excel® appropriate for resampling estimations?

Amongst the advantages of using Microsoft Excel® for calculations in general and for resampling in particular is that the software is widely available and used for administrative purposes in the laboratory and facilitates easy data handling and visualization. Using the natural functions of Microsoft Excel® 2010 or later the resampling procedures are made highly visual, providing the user with hands on visual contact with all the random samples taken and the calculations made. This is, however, also the disadvantage of the use of the natural functions gallery of Microsoft Excel® for this purpose since there is a risk of errors when performing calculation on large tables of data – e.g. the 1000 columns needed for the present purpose.

Conclusion

Resampling methods are recommended by IFCC and are widely available in software for handling reference values (5,19-24) but not as yet as yet a part of the CLSI guideline (16). Resampling methods are appropriate when the distribution of data from the reference samples is non-Gaussian and in case the number of reference samples are in order of at least 40. At least 500-1000 random samples with replacement should be taken from the reference samples. Resampling methods are conveniently implemented in versions of Microsoft Excel® 2010 or later as shown here.

Potential conflict of interest

None declared.

References

1. Gräsbeck R, Saris NE. Establishment and use of normal values. Scand J Clin Invest 1969;26:S62-3.

2. Siest G, Henny J, Grasbeck R, Wilding P, Petitclerc C, Queralto JM, et al. The theory of reference values: an unfinished symphony. Clin Chem Lab Med 2013;51:47-64.http://dx.doi.org/10.1515/cclm-2012-0682.

3. Murphy EA. The normal, and the perils of the sylleptic argument. Perspect Biol Med 1972;15:566-82.http://dx.doi.org/10.1353/pbm.1972.0003.

4. Gräsbeck RA, T. Solberg, H E. Reference Values in Laboratory Medicine. Chichester: John Wiley & Sons; 1981.

5. Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal program. Clin Chem Lab Med 2004;42:710-4.http://dx.doi.org/10.1515/CCLM.2004.121.

6. Horn PS, Pesce AJ, American Association for Clinical Chemistry. Reference intervals: a user's guide. Washington, DC: American Association for Clinical Chemistry; 2005. p. p.

7. Harris EK, Boyd JC. Statistical bases of reference values in laboratory medicine. New York: M. Dekker; 1995. xiv, 361 p. p.

8. Solberg HE. The theory of reference values Part 5. Statistical treatment of collected reference values. Determination of reference limits. J Clin Chem Clin Biochem 1983;21:749-60.

9. Henny J, Petitclerc C, Fuentes-Arderiu X, Hyltoft Petersen P, Queraltó JM, Schiele F, et al. Need for revisiting the concept of reference values. Clin Chem Lab Med 2000;38:589-95. http://dx.doi.org/10.1515/CCLM.2000.085.

10. Horowitz GL. Establishment and Use of Reference Values. In: Burtis CA, Ashwood ER, Bruns DE, editors. Tietz textbook of Clinical Chemistry and Molecular Diagnostics. Fifth Edition ed. St. Louis: Elsevier; 2012. p. 95-118.http://dx.doi.org/10.1016/b978-1-4160-6164-9.00005-6.

11. CLSI. Defining, Establishing, and Verifying reference Intervals in the Clinical Laboratory; Approved Guideline – Third Edition. Document C28-A3: CLSI; 2010.

12. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993. 436 p.http://dx.doi.org/10.1007/978-1-4899-4541-9.

13. Good PI. Resampling methods: a practical guide to data analysis. 3rd ed. ed. Boston: Birkhäuser; 2006.

14. Christie D. Resampling with Excel. Teaching Statistics 2004;26:9-14.http://dx.doi.org/10.1111/j.1467-9639.2004.00136.x.

15. Solberg HE. The statistical theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta 1984;137:97F-114F.http://dx.doi.org/10.1016/0009-8981(84)90319-X.

16. CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline EP28-A3c. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

17. Hyndman RJ, Fan Y. Sample Quantiles in Statistical Packages. The American Statistician 1996;50:361-5.

18. Wikipedia. Quantile. Available at: http://en.wikipedia.org/wiki/Quantile.Accessed May 17 2015.

19. Solberg HE. RefVal: a program implementing the recommendations of the International Federation of Clinical Chemistry on the statistical treatment of reference values. Comput Methods Programs Biomed 1995;48:247-56.http://dx.doi.org/10.1016/0169-2607(95)01697-X.

20. Analyse-it Software L. Analyse-IT. 2015.

21. MedCalc. Acacialaan 22, 8400 Ostend, Belgium: MedCalc Software bvba; 2015.

22. Linnet C. CBstat: A Program for Statistical Analysis in Clinical Biochemistry - CD. 2007.

23. Kenny D, Solberg HE. RefVal for windows--migrating to the Windows environment with the help of Visual Basic. Clin Chim Acta1993;222:19-21.http://dx.doi.org/10.1016/0009-8981(93)90088-L.

24. Geffre A, Concordet D, Braun JP, Trumel C. Reference Value Advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel. Veterinary clinical pathology / American Society for Veterinary Clinical Pathology. 2011;40:107-12.http://dx.doi.org/10.1111/j.1939-165X.2011.00287.x.

25. Efron B, Tibshirani R. Statistical-Data Analysis in the Computer-Age. Science 1991;253:390-5.http://dx.doi.org/10.1126/science.253.5018.390.

26. Jones HL. Investigating the properties of a sample mean by employing random subsample means. JASA 1956;51:54-83.http://dx.doi.org/10.1080/01621459.1956.10501311.

27. Solberg HE. The theory of reference values. Part 5. Statistical treatment of collected reference values determination of reference limits. J Clin Chem Clin Biochem 1983;21:749-60.

28. Linnet K. Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. Clin Chem 2000;46:867-9.

29. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Stat Sci 1986;1:54-75.http://dx.doi.org/10.1214/ss/1177013815.

30. Solberg HE. Approved recommendation on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta 1987;170:S13-S32.http://dx.doi.org/10.1016/0009-8981(87)90151-3.

31. Coskun A, Ceyhan E, Inal TC, Serteser M, Unsal I. The comparison of parametric and nonparametric bootstrap methods for reference interval computation in small sample size groups. Accredit Qual Assur 2013;18:51-60.http://dx.doi.org/10.1007/s00769-012-0948-5.

32. Bjerner J, Theodorsson E, Hovig E, Kallner A. Non-parametric estimation of reference intervals in small non-Gaussian sample sets. Accredit Qual Assur 2009;14:185-92.http://dx.doi.org/10.1007/s00769-009-0490-2.

33. Almiron MG, Lopes B, Oliveira ALC, Medeiros AC, Frery AC. On the Numerical Accuracy of Spreadsheets. J Stat Softw 2010;34:1-29.

34. McCullough BD, Wilson B. On the accuracy of statistical procedures in Microsoft Excel 97. Comput Stat Data An 1999;31:27-37.http://dx.doi.org/10.1016/S0167-9473(99)00004-3.

35. Melard G. On the accuracy of statistical procedures in Microsoft Excel 2010. Computation Stat 2014;29:1095-128. http://dx.doi.org/10.1007/s00180-014-0482-5.