It is likely that these combined efforts will become decisive for formulation of future receptor models, which might better match all the experimental evidence than the contemporary schemes.

Acknowledgments. Comments on this topic and kind help by Drs Denis J. Crankshaw and David F. Sargent are gratefully acknowledged. Supported by Swiss National Science Foundation.

1 Laduron, P. M., Biochem. Pharmac. *33* (1984) 833.
2 Nickerson, M., Nature *178* (1956) 697.
3 Pliška, V., Heiniger, J., Müller-Lhotsky, A., Pliska, P., and Ekberg, B., J. biol. Chem. *261* (1986) 16984.
4 Soloff, M. S., and Swartz, T. L., J. biol. Chem. *249* (1974) 1376.
5 Soloff, M. S., in: Hormone-Receptor Interaction – Molecular Aspects, p. 129. Ed. G. S. Levey. Marcel Dekker, Inc., New York 1976.
6 Crankshaw, D. J., Branda, L. A., Matlib, M. A., and Daniel, E. E., Eur. J. Biochem. *86* (1978) 481.
7 Crankshaw, D. J., Romaniuk, E., and Branda, L. A., Gynec. Obstet. Invest. *14* (1982) 202.
8 Fuchs, A. R., Fuchs, F., Husslein, P., Soloff, M. S., and Fernström, M. J., Science *215* (1982) 1396.
9 Riemer, R. K., Goldfien, A. C., Goldfien, A., and Roberts, J. M., Endocrinology *119* (1986) 699.
10 Pliška, V., Marbach, P., Vašák, J., and Rudinger, J., Experientia *33* (1977) 367.
11 Pliška, V., and Marbach, P., Eur. J. Pharmac. *49* (1978) 213.
12 Lindeberg, G., Vilhardt, H., Larsson, L.-E., Melin, P., and Pliška, V., J. Receptor Res. *1* (1980) 389.
13 Pliška, V., J. Receptor Res. *4* (1984) 371.
14 Leff, P., and Harper, D., J. theor. Biol. *140* (1989) 381.
15 Pliška, V., J. Receptor Res. *8* (1988) 245.
16 Moore, G. J., and Kwok, Y. C., Life Sci. *41* (1987) 505.
17 Birdsall, N. J. M., Burgen, A. S. V., and Hulme, E. C., Molec. Pharmac. *14* (1978) 723.
18 Furchgott, R. F., and Bursztyn, P., Annls N.Y. Acad. Sci. *144* (1967) 882.
19 Pliška, V., Trends pharmac. Sci. *8* (1987) 50.
20 Pliška, V., Arzneim.-Forsch. *16* (1966) 886.
21 Ariëns, E. J., and van Rossum, J. M., Archs int. Pharmacodyn. *110* (1957) 257.
22 van Rossum, J. M., Archs int. Pharmacodyn. *143* (1963) 299.
23 Schild, H. O., Br. J. Pharmac. *2* (1947) 189.
24 Pliška, V., and Heiniger, J., Int. J. Pept. Prot. Res. *31* (1988) 520.
25 Free, S. M., and Wilson, J. W., J. med. Chem. *7* (1964) 395.
26 Hruby, V. J., in: Perspectives in Peptide Chemistry, p. 207. Eds A. Eberle, R. Geiger and T. Wieland. S. Karger, Basel 1981.
27 Wood, S. P., Tickle, I. J., Treharne, A. M., Pitts, J. E., Mascarenhas, Y., Li, J. Y., Husain, J., Cooper, S., Blundell, T. L., Hruby, V. J., Buku, A., Fischman, A. J., and Wyssbrod, H. R., Science *232* (1986) 633.
28 Hruby, V. J., Trends pharmac. Sci. *8* (1987) 336.
29 Fauchère, J.-L., Charton, M., Kier, L. B., Verloop, A., and Pliška, V., Int. J. Pept. Prot. Res. *32* (1988) 269.
30 Pliška, V., and Charton, M., in: Peptides – Chemistry, Structure and Biology, p. 290; Proc. of the Eleventh American Peptide Symposium, La Jolla 1989. Eds J. E. Rivier and G. R. Marshall. ESCOM, Leiden 1990.
31 Birdsall, N. J. M., Burgen, A. S. V., and Hulme, E. C., in: Cholinergic Mechanisms and Psychopharmacy, p. 25. Ed. D. J. Jenden. Plenum Press, New York 1978.
32 Furchgott, R. F., Fedn Proc. *37* (1978) 115.
33 Sargent, D. F., and Schwyzer, R., Proc. natl Acad. Sci. USA *83* (1986) 5774.
34 Sargent, D. F., Bean, J. W., Kosterlitz, H. W., and Schwyzer, R., Biochemistry *27* (1988) 4974.

0014-4754/91/030216-06$1.50 + 0.20/0

Research Articles

The use of significance limits in graphical data representations

J. Elsner and R. Knutti

Institute of Toxicology, Swiss Federal Institute of Technology and University of Zurich, CH-8603 Schwerzenbach (Switzerland)
Received 31 January 1990; accepted 20 July 1990

Summary. Significance limits are proposed as an alternative to the use of standard deviation, standard error, or confidence or tolerance limits when experimental data are presented in a graphical form. This measurement of uncertainty allows graphical t-tests to be used both for the estimation of data variance and for an approximate statistical comparison between two or more data sets.
Key words. Uncertainty limits; error bars; graphical t-test; significance limits.

Graphical representations are used in all areas of experimental science as an efficient tool for comparing two or more data sets. Over the past years, the new approach of exploratory data analysis (EDA)[1] has elevated data visualization to the status of an accepted statistical technique. Also, because of the increasing availability of computers, graphical representations have become more feasible. Since the values of the means or medians do not by themselves give enough information for the comparison of data sets, it is essential that they be accompanied by an appropriate measurement of uncertainty. EDA proposes the characterization of data sets graphically by medians, hinges, quartiles, and depth. This allows the representation of many features of a set, such as amount of spread, symmetry, and the presence of outliers. However, this representation, and also all classical uncertainty ranges (standard error, confidence limits, standard deviations, and tolerance limits), do not permit statistical differences between data sets to be inferred from the graphical representations alone.

Figure 1 shows the individual values, the means, and a variety of representations of uncertainties in increasing

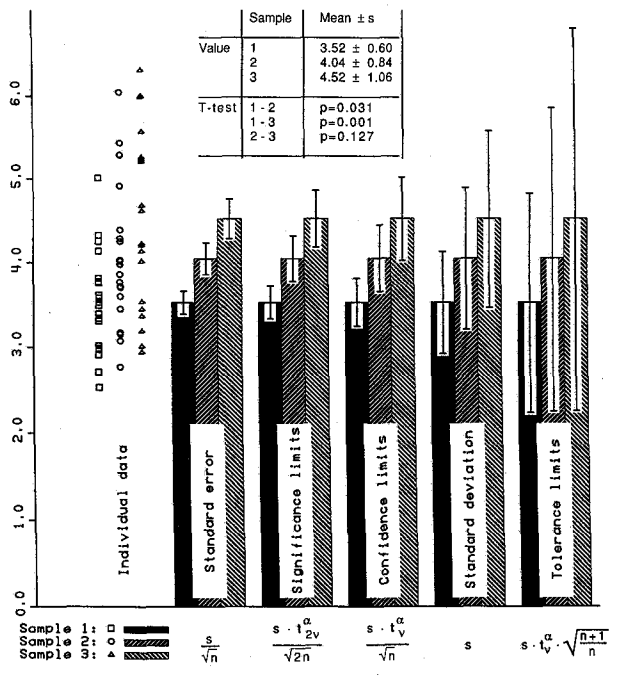

Figure 1. The significance limits of three normally distributed random samples of size 20 (described in the table) are compared with the individual data, the standard errors, the confidence limits, the standard deviations and the tolerance limits as represented by ranges above and below the sample means. These uncertainty values are expressed in the formulae below the bars [7] as functions of the standard deviation s, the sample size n and the t-test ratio **t**. The significance criterion is $\alpha = 0.05$, and the degree of freedom $\nu = n-1$. Whether the ranges defined by the significance limits overlap or not reflects the result of t-tests at the selected significance criterion, as shown in the table, a property which cannot be attributed to the other uncertainty measures.

order for three sets of 20 data points generated by a normal distribution random value generator [2]. The sets have the means and standard deviations listed in the figure. Also listed are t-test results, considering non-homogeneous standard deviations [3]. The individual data are shown on the left of figure 1. The first bar-cluster shows the smallest and, maybe for this reason, most popular uncertainty representation, the standard error, indicating the standard deviation of the mean. The intuitively obvious action when studying such a graph is to estimate the differences between the means from the overlap of the ranges defined by the standard errors. The fact that none of these ranges overlap would suggest that all the data sets may be considered as statistically different. This is not in agreement with the t-test, which shows a lack of significance at the $\alpha = 0.05$ level for the comparison between sets 2 and 3.

Confidence limits, represented in the third bar-cluster, are proposed by many statisticians as being the least biased uncertainty measurement. A confidence limit defines a range within which the mean of the data sample lies with a confidence of 95%, or alternatively, the mean of the sample is statistically different from any fixed value outside this range. This example suggests, again considering the overlapping characteristics of the ranges, that only sets 1 and 3 are statistically different. This conclusion is again not in agreement with the t-test, which indicates that sets 1 and 2 are also different at the $\alpha = 0.05$ level. Checking the location of the mean of one sample with respect to the confidence limit of another sample, a procedure which corresponds more closely to the definition of the confidence limit, may reveal inconsistencies as demonstrated by samples 2 and 3: the mean of sample 3 is outside the range defined by the confidence limit of sample 2, but the mean of sample 2 is within the range defined by the confidence limit of sample 3.

The standard deviations and tolerance limits (limits which contain 95% of the values of an inferred normal distribution), represented in the two last bar-clusters, suggest that none of the comparisons reach statistical difference; this is also in contradiction to the t-test results. These two uncertainty measures are well suited for a characterization of the individual data samples, but are no help for estimating the differences between the sample means.

For the purpose of being able to infer statistical differences between two or more data sets in graphical data representations, a new measurement of uncertainty is defined, which we call the significance limit τ. The significance limits define a range above and below the mean which excludes all significance ranges of other data sets with the same size and variance, which are significantly different from the original data set at the $\alpha = 0.05$ level. The test ratio for a t-test of two data sets of equal sample size and variance has the formula:

$$\mathbf{t} = \frac{d}{s \cdot \sqrt{\frac{2}{n}}}$$

where d is the difference between the means, s the standard deviation and n the sample size common to the two samples. The significance limit τ of a single data set, defined as half the difference d of two samples with a value of **t** corresponding to the significance level α, and the degree of freedom of $2\nu = 2n-2$, is thus calculated as:

$$\tau = \frac{s \cdot \mathbf{t}^{\alpha}_{2\nu}}{\sqrt{2n}}$$

As can be seen from the second bar-cluster in figure 1, which shows significance limits at the $\alpha = 0.05$ level, the overlapping characteristics of the significance limits of the three random samples correspond precisely to the t-test results at the same level.

Figure 2 shows the ratios between the significance limit and the other uncertainty limits as functions of the sample size. The significance limits belong to the same class as the standard error and the confidence limits, since the ratio between the uncertainty measurements tends to a constant as the sample size increases (the significance limit tends to zero with increasing sample size). As can be

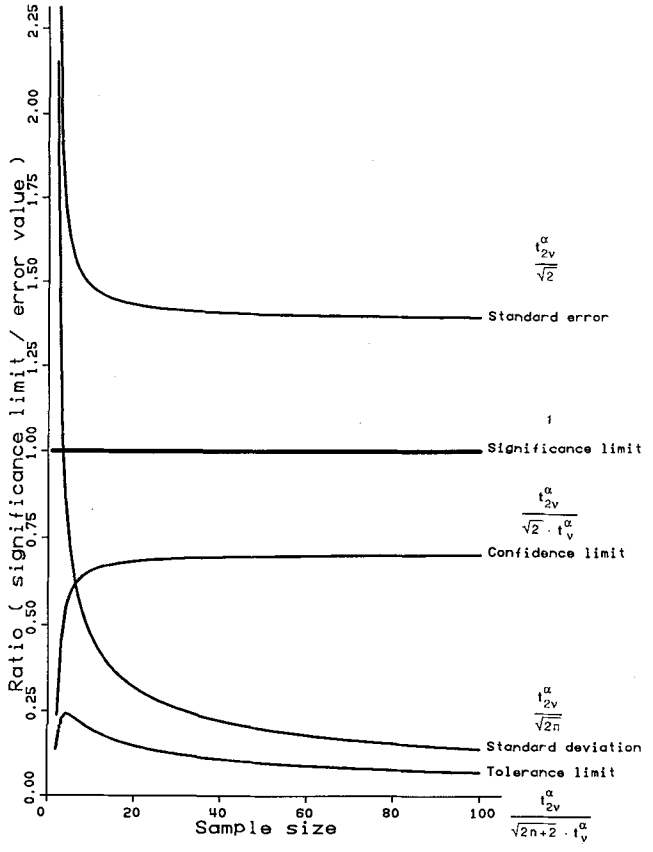

Figure 2. The ratios between the significance limits and other uncertainty limits expressed as functions of the sample size. A value larger than one signifies that the labelled uncertainty is smaller than the significance limit. The formulae express the ratios as displayed.

seen from figure 2, the behavior of the ratio at low sample sizes is quite interesting. Taking a sample size of 4, the standard error is only little more than half of the significance limit, while the confidence limit is almost twice the significance limit. Thus, the significance limit may also be regarded as a true compromise between the standard error and the confidence limit.

It is important to insist on the fact that the use of significance limits in graphical representations does not replace a correct statistical analysis. Especially multiple graphical comparisons may only be interpreted correctly after having done, for example, a one- or a multi-way ANOVA (if the ANOVA resulted in a significant effect the graphical significances may be used, e.g., for the determination of a 'no statistical effect level' in a dose-response relation). Other authors propose graphical Multiple Comparison Procedures which take into account the number of possible comparisons[4]. To our knowledge these techniques, which are not context-independent, have never been used in common graphical representations. The definition of significant limits as presented here is less sophisticated but much easier to understand and simpler to use if the rules of good statistical practice are respected.

Additional caution in the interpretation of the graphical readings is mandatory in cases with unequal variances or unequal sample sizes. As already mentioned, significance limits are defined with the assumption that the variances and sizes of all samples are equal[5]. The error under other conditions may be estimated by calculating the true significance level of two sets whose significance ranges touch. In this case, the graphical t-test would imply that $p = 0.05$, whereas a correct t-test would show that $p \geq 0.05$. In figure 3, true t-test probabilities[3] are plotted as functions of the ratio between the standard deviations of two equally-sized samples. For very small sample sizes, the error in the probability p increases rapidly as the standard deviation ratio moves away from one. For larger sample sizes, the error stays at a tolerable level if the ratio is between 0.5 and 1. This inference error is of the same order of magnitude as the one which occurs when a simple t-test (which does not take into consideration variance inhomogeneity) is done for two samples with unequal variances. Notwithstanding that one of the basic requirements for using significance limits is not fulfilled in the example of figure 1 (equality of the standard deviations), the error bar magnitudes partially compensate for this fact and reflect qualitatively the t-test results.

The significance limits as defined above have been used in several publications[6]. In all cases they allowed quite accurate graphical t-tests, rendering the use of additional symbols (such as stars) for the indication of significances superfluous. Nevertheless, correct statistical analyses (t-tests, Wilcoxon tests, ANOVA etc.) were also performed,

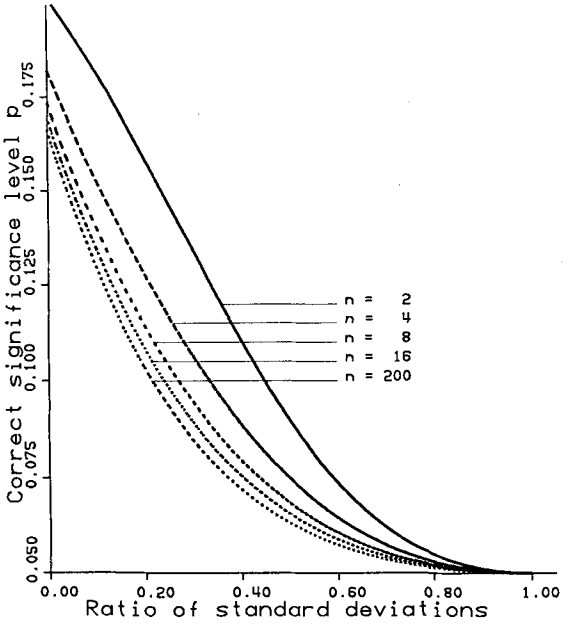

Figure 3. The correct significance levels for different sample sizes n as functions of the ratio between the standard deviations of two samples which are estimated from graphical t-tests to be just significantly different at the $\alpha = 0.05$ level.

and they generally corresponded to the graphical readings. Thus, significance limits are a useful tool for graphical data representations, since they indicate the statistical uncertainty of a value, and also allow approximate statistical comparisons between two or more data sets.

1 Velleman, P. F., and Hoaglin, D. C., Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston, MA, 1981.

2 Three different sets of values with an a priori mean of 0 and standard deviation of 1 were generated. The first set was multiplied with 0.6 (thereby reducing its a priori standard deviation by the same factor) and incremented by 3.78; the second set was just incremented by 4.05 and the third set by 4.25. The a posteriori means and standard deviations are listed in fig. 1. F-tests for comparing the variances resulted in following error probabilities: set 1 vs set 2: $p = 0.078$; set 1 vs set 3: $p = 0.009$; set 2 vs set 3: $p = 0.164$.

3 Welch, B. L., Biometrika *36* (1949) 293. This procedure for unequal variances converges to the one for equal variances when the standard deviations become equal. Therefore, it has been used for all non-graphical t-tests.

4 Andrews, H. P., Snee, R. S., and Sarner, M. H., Am. Statistn *34* (1980) 195; Hochberg, Y., Weiss, G., Hart, S., J. Am. statist. Ass. *77* (1982) 767; Godfrey, K., N. Engl. J. Med. *313* (1985) 1450.

5 For ease of implementation and for clarity (context independence), only information from the considered individual data set should be included in the definition of an uncertainty measure. Therefore, unequal variances and sample-sizes, as well as comparison multiplicity were not taken into account for the definition of the significance limits. It is, however, important that the user of significance limits becomes aware of the imprecisions which may result from these simplifications.

6 See for example Elsner, J., Looser, R., and Zbinden, G., Neurobehav. Toxic. Terat. *1*, suppl. 1 (1979) 163; Schlatter, J., Elsner, J., and Zbinden, G., Neurobehav. Toxic. Terat. *5* (1983) 413; Looser, R., Elsner, J., and Zbinden, G., Psychopharmacology *84* (1984) 323; Elsner, J., Neurobehav. Toxic. Terat. *8* (1986) 573; Elsner, J., Fellmann, Ch., and Zbinden, G., Neurobehav. Toxic. Terat. *10* (1988) 3; Elsner, J., Hodel, B., Suter, K. E., Oelke, D., Ulbrich, B., Schreiner, G., Cuomo, V., Cagiano, R., Rosengren, L. E., Karlsson, J. E., and Haglid, K., Neurobehav. Toxic. Terat. *10* (1988) 155; Elsner, J., Alder, S., and Zbinden, G., Psychopharmacology *96* (1988) 194; Tannhauser, S. L., Elsner, J., Tannhauser, M., Barros, H. M. T., and Tannhauser, M. A., Brazilian J. med. biol. Res. *22* (1989) 213; Balduini, W., Lombardelli, G., Peruzzi, G., Cattabeni, F., and Elsner, J., Neurobehav. Toxic. Terat. *11* (1989) 339; Elsner, J., Alder, S., and Fellmann, Ch., Neurobehav. Toxic. Terat. *12* (1989) 7.

7 For the formulae of the confidence and tolerance limits see e.g. Wissenschaftliche Tabellen Geigy, Vol. 3, p. 206–207. Ciba-Geigy AG, Basel 1980.

0014-4754/91/030221-04$1.50 + 0.20/0

Aluminium in repair membranes and Al, Ca and P_i in the haemolymph of Al-injected shell-repairing snails (*Helix pomatia* L.)

M. Reineskog

Department of Zoophysiology, Uppsala University, Box 560, S-751 22 Uppsala (Sweden)
Received 29 June 1990; accepted 18 September 1990

Summary. In $AlCl_3$-injected shell-repairing snails, *Helix pomatia* L., the Al-associated decrease of the weights of the shell-repair membranes was unrelated to the Al-concentration in the membranes. In the haemolymph the concentration of Al was related to the dose of injected Al, while the concentration of Ca was increased by the highest Al-dose only. No phosphate was detected in either controls or Al-injected snails. It is concluded that Al inhibits the growth of the $CaCO_3$-crystals by mechanisms other than incorporation in, or adsorption to, the crystals.
Key words. Aluminium; snail; shell-repair; calcium; phosphate.

Aluminium (Al) has been shown in laboratory investigations to affect the process of shell-repair in the snail *Helix pomatia* L. Both the weights and the relative calcium concentration in the shell-repair membranes are reduced in snails injected with $AlCl_3$ [1]. These effects resemble those recorded in the skeletal bones both from patients suffering from the Al-associated syndrome of dialysis osteomalacia and from Al-treated experimental animals [2–4]. Aluminium has furthermore been demonstrated to inhibit the calcification of demineralized shell-repair membranes in vitro, and to reduce the formation of calcium carbonate ($CaCO_3$), the mineral found in the shell of the snail, in a pure physical-chemical system [5]. Similar effects of Al have also been reported concerning the in vitro formation of calcium phosphate [6, 7]. Aluminium possibly interferes with crystal growth by adsorbing to the crystal surfaces. The presence of Al in the lower-weight repair membranes of the Al-injected snails thus indicates a direct effect of Al on the growth of the $CaCO_3$-crystals, while the absence points towards an indirect mechanism.

Aluminium is also known to interfere with phosphate metabolism [3, 8]. Energy-requiring processes like the production of the organic matrix of the repair membrane and the transport of ions across the mantle epithelium (see Watabe [9] for a review of the shell-repair process) may thus be affected. A disturbed phosphorus metabolism in the Al-treated snails might be detected as a reduced concentration of inorganic phosphate (P_i) in the haemolymph.

The induction of acidosis, e.g. by the injection of an acidic solution, increases the Ca-concentration in the haemolymph [10]. Elevated Ca-concentrations in animals injected with $AlCl_3$, which is an acidic solution, thus