



The Impact of Significant Digits in Bioequivalence Studies

Yuqing Liu*¹ and Shein Chung Chow¹¹Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA***Corresponding author:** Yuqing Liu, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA.**Received Date:** March 29, 2021**Published Date:** May 28, 2021

Abstract

In biomedical research, various statistical methods are used to analyze data collected from studies to provide statistical inferences of the unknown parameters of interest. The analysis results up to a few decimal places are usually obtained for scientific validity. The use of different number of decimal places, however, may alter the conclusion of the analysis results. A typical example is the assessment of bioequivalence for generic approval in bioequivalence studies. In practice, there is no uniformly accepted cut-off standard for the numbers of decimal places in biomedical research. In this paper, following the concept of signal-noise in quality control and assurance, we propose a simple procedure in determining the number of decimal places for providing a valid statistical inference and assurance in biomedical research. Based on a non-central t distribution, the proposed procedure is able to control the largest rounding error when considering rounding up the data to certain decimal places. In other words, the proposed method could control the worst possible error adjusted for standard deviation by selecting an appropriate digit that meets the criteria that the non-central parameter is less than a pre-specified critical value. Statistical justification of the proposed procedure was provided and illustrated through a simulated dataset.

Keywords: Analytical Research; Decimal Places; Non-Central t Distribution; Transformation

Introduction

In drug research and development, both non-clinical and clinical studies are often conducted for assessment of treatment effect of the drug product under investigation. In practice, analysis results up to a few decimal places for scientific validity are usually obtained. However, there is no uniformly accepted standard regarding how many decimal places should be used for scientific validity. As an example, the latest version of the AMA Manual of Style only requires interception to reflect the original accuracy of measurement and the original precision of calculation (Committee, AMA, 2020), and the Publication Manual of APA generally recommends keeping one to three rounding decimals in different situations (American Psychological Association, 2020). Therefore, rounding numbers play an essential role in biomedical research, where analysis results must be very accurate for maintaining scientific validity.

For another example, in the assessment of bioequivalence (BE) for generic approval, it is required that the constructed 90% confidence interval of the geometric mean ratio falls entirely within the bioequivalence limit of (80%,125%) with no decimal place, (80.0%, 125.0%) with one decimal place, or (80.00%, 125.00%) with 2 decimal places. Thus, if the constructed confidence interval is marginally significant, the selection of decimal places could alter the conclusion of bioequivalence assessment. For example, suppose that the constructed 90% confidence interval is (79.956, 125.012). If one decimal place is selected, the 90% confidence interval is given by (80.0%, 125.0%) after rounding up to the first decimal place, which falls within the bioequivalence limit of (80.0%, 125.0%) entirely. Hence, we conclude that the test product is bioequivalent to the reference product. If two decimal places are selected, the 90%

confidence interval becomes (79.96%, 125.01%) after rounding up to the second decimal place, which does not fall within the bioequivalence limit of (80.00%, 125.00%). Consequently, we fail to demonstrate bioequivalence between the test product and the reference product. Thus, the use of different decimal places may lead to different conclusions.

Following the concept of signal-noise introduced by Chow (2000), the appropriate number of significant digits for analytical results that satisfy two conditions would be selected. First, there is no significant difference between the truncated data with the raw data. Second, any data with fewer digits would not meet the first condition. In other words, after obtaining analytical results from statistical experiments, the optimal decimal place we should preserve is the smallest digits that could retain the same statistical analytical results as before. In this article, statistical properties and justification for Chow's proposal for selection of appropriate decimal places are studied. The method proposed in this article aims to help determining the number of decimal places in biomedical research such as bioequivalence assessment for generic drugs and bio similarity assessment for biosimilar drug products. At the same time, it provides a user-friendly and simple way to obtain statistical inference while achieving the desired accuracy and precision. The proposed procedure is also very helpful in basic research such as drug discovery when making a critical go/no-go decision. The next section defines significant digits (decimal places) and bioequivalence testing. The proposed procedure for the determination of significant digits is briefly outlined in Section 3, which is illustrated through a numerical study in Section 4. In the Section 5, statistical properties and justification of the proposed procedure are studied. An application of bioequivalence study is given in Section 6. Concluding remarks are provided in the last section of this article.

Simulated Examples on Number of Decimal Places

As indicated in the FDA guidance on bioequivalence study, the

Proposed procedure of Observations Truncation

Table 1: C_{\max} data from a parallel bioequivalence trial.

X	X_0	X_1	X_2	Y	Y_0	Y_1	Y_2
1.381	1	1.4	1.38	1.007	1	1.0	1.01
1.298	1	1.30	1.3	0.995	1	1.0	0.99
1.177	1	1.2	1.18	0.892	1	0.9	0.89
1.294	1	1.3	1.29	1.129	1	1.1	1.13
1.257	1	1.3	1.26	0.977	1	1.0	0.98
1.237	1	1.2	1.24	1.031	1	1.0	1.03
1.368	1	1.4	1.37	1.011	1	1.0	1.01
1.253	1	1.3	1.25	0.718	1	0.7	0.72
1.530	2	1.5	1.53	1.085	1	1.1	1.09
1.424	1	1.4	1.42	0.930	1	0.9	0.93
1.220	1	1.2	1.22	0.931	1	0.9	0.93

90% confidence interval of the geometric mean ratio between a test product and a reference product is regarded as the indicator to for evaluation of bioequivalence (FDA, 2003; FDA, 2006; Chow and Liu, 2008). Meanwhile, there are two different boundary and calculated methods for the 90% confidence interval of raw data and log-transformed data, respectively. To demonstrate that observations with different digits will lead to different conclusions, (Table 1) gives a simulated C_{\max} dataset with 24 subjects from a parallel-group bioequivalence trial comparing a test product with a reference product. The simulated data were generated from normal distributions $N(1.3, 0.1^2)$ for the test drug X and $N(1.0, 0.1^2)$ for the reference drug Y . In addition, the BE analysis is performed based on the raw data (no log-transformation data) through the confidence interval approach, and bioequivalence would claim if and only if the 90% confidence interval constructed by the mean difference of reference and test drug covered by interval $(-0.2\mu_R, 0.2\mu_R)$, where μ_R is the mean of the reference treatment or drug.

Therefore, the BE limit of this example is $(-0.2, 0.2)$, and the BE analysis should reject bioequivalence because the real difference between the mean of test and reference drugs is $1.3 - 1.0 = 0.3$. As can be seen, (Table 2) summarizes the confidence intervals of C_{\max} data with different decimal places. However, different decimal places of observations lead to different bioequivalence conclusions. Thus, if we do not keep enough decimal digits, we could falsely conclude BE. To indicate that the BE limit with different digits will also cause different results, (Table 3) provides some possible scenarios of the exponential confidence interval of log-transformed data and the BE study results. The bioequivalence for log-transformed data will imply when the exponential confidence interval $[e^L, e^U]$ is covered by $[80\%, 125\%]$, where $[L, U]$ is 90% confidence interval of transformed treatment effect. (Table 3) illustrates the impact of significant digits of the BE limit, which could also mislead the BE analysis.

1.313	1	1.3	1.31	0.966	1	1.0	0.97
1.412	1	1.4	1.41	0.944	1	0.9	0.94
1.239	1	1.2	1.24	1.087	1	1.1	1.09
1.3	1	1.3	1.30	1.064	1	1.1	1.06
1.292	1	1.3	1.29	1.082	1	1.1	1.08
1.223	1	1.2	1.22	1.063	1	1.1	1.06
1.268	1	1.3	1.27	0.920	1	0.9	0.92
1.482	1	1.5	1.48	0.954	1	1.0	0.95
1.149	1	1.1	1.15	1.212	1	1.2	1.21
1.387	1	1.4	1.39	0.932	1	0.9	0.93
1.162	1	1.2	1.16	1.025	1	1.0	1.03
1.476	1	1.5	1.48	0.819	1	0.8	0.82
1.467	1	1.5	1.47	0.959	1	1.0	0.96

X : the original data from test drug; X_i : the original data with i decimal digits.

Y : the data from the reference drug; Y_i : the data with i decimal digits.

Table 2: Summary results of the bioequivalence study.

Decimal Places of Observations	90% Confidence Interval	BE limit	Pass/Fail
0	(-0.030, 0.113)	(-0.2, 0.2)	Pass
1	(0.279, 0.388)	(-0.2, 0.2)	Fail
2	(0.277, 0.379)	(-0.2, 0.2)	Fail

Table 3: The impact of significant digits of BE limit.

Possible scenario of 90% confidence interval	# of decimal places	Rounding result	BE limit	pass/fail
(79.693,124.504)	0	(80, 125)	(80, 125)	Pass
	1	(79.7, 124.5)	(80.0, 125.0)	Fail
(79.956,125.012)	0	(80, 125)	(80, 125)	Pass
	1	(80.0, 125.0)	(80.0, 125.0)	Pass
	2	(79.96, 125.01)	(80.00, 125.00)	Fail

Suppose X is a continuous random variable with standard deviation σ , and X^* is its truncated value after rounding up to the d^{th} decimal place with the same variance. We said X^* is not statistically different from X if we fail to reject the following null hypothesis at the α level of significance

$$H_0 : \mu_X = \mu_{X^*} \text{ vs } H_a : \mu_X \neq \mu_{X^*} \quad (1)$$

where μ_X and μ_{X^*} are the population means of X and X^* , respectively. Meanwhile, when X and X^* are not statistically distinguishable, the d^{th} decimal places are considered significant digits we would like to keep. For all d , $|X - X^*|$, the maximum possible error due to the truncation would be less than 10^{-d} . As an example, for floating number $a.bcef$, if $d = 2$, the smallest and largest values for a given number with two decimal places are $a.bc$ and $a.b(c+1)$, respectively. Hence, the possible error is $0.00ef$ or $(0.01 - 0.00ef)$, which means that the maximum possible error must less than $0.01 = 10^{-2} = 10^{-d}$. Intuitively, to restrict the worst-case rounding error small

enough so that the distortion of the distribution due to rounding error would be negligible. An idea of signal-noise in quality control and assurance is applied to compare this error with X 's standard deviation σ . The significant digits can then be chosen by taking the minimum d digits such that

$$10^{-d} / \sigma \leq C$$

where C is a constant, which is to be chosen such that the truncated observation in X^* is not statistically different from X at the α level of significance. In practice, a convention choice of C is $C = 10\%$. To provide a better understanding of the proposed procedure, the results for various choices of C given σ is summarized in (Table 4). It can be seen that a smaller value of C would require more decimal places to achieve the desired accuracy and precision, the same for σ .

A Numerical Study

A numerical study is conducted to demonstrate the use of the proposed procedure. Thirty analytical results x_i , $i=1,2,\dots,30$ were

generated from $N(\pi, 0.01^2)$, which are given in (Table 5). For $C = 10\%$, we have

$$10^{-d} / \sigma = 10^{-d} / 0.01 \leq 0.1$$

Thus, the minimum number of d that satisfies the above expression is $d = 3$. Therefore, the number of significant decimal digits is chosen to be 3. Now consider four data sets $X_{ij}, j = 1, 2, 3, 4$, which are truncated at the j^{th} decimal places of X_i respectively. Then a paired t -test is performed in (Table 6) to test if X_{1i}, X_{2i}, X_{3i} are significant-

ly different from the rest data sets, which shows that the rounding error could alter the distribution significantly. These results also prove that the proposed procedure works well because X_{3i} is not significantly different from X_{4i} . However, it should be noted that the conventional choice of C may be conservative in this case, since X_{2i} is also not significantly different from X_{3i} and X_{4i} . Therefore, we need to use different thresholds on different data instead of blindly using 10%.

Table 4: Significant decimal digits for various selection of C given σ .

σ	C				
	1%	5%	10%	15%	20%
0.01	4	4	3	3	3
0.1	3	3	2	2	2
0.5	3	2	2	2	1
1	2	2	1	1	1
2	2	1	1	1	1

Table 5: Simulation data set for two-sample t -test.

i	X_i	X_{1i}	X_{2i}	X_{3i}	X_{4i}
1	3.151756	3.2	3.15	3.152	3.1518
2	3.127301	3.1	3.13	3.127	3.1273
3	3.136724	3.1	3.14	3.137	3.1367
4	3.149354	3.1	3.15	3.149	3.1494
5	3.151683	3.2	3.15	3.152	3.1517
6	3.134783	3.1	3.13	3.135	3.1348
7	3.157186	3.2	3.16	3.157	3.1572
8	3.134598	3.1	3.13	3.135	3.1346
9	3.13699	3.1	3.14	3.137	3.1370
10	3.134906	3.1	3.13	3.135	3.1349
11	3.130847	3.1	3.13	3.131	3.1308
12	3.148668	3.1	3.15	3.149	3.1487
13	3.154195	3.2	3.15	3.154	3.1542
14	3.154655	3.2	3.15	3.155	3.1547
15	3.133849	3.1	3.13	3.134	3.1338
16	3.144525	3.1	3.14	3.145	3.1445
17	3.131453	3.1	3.13	3.131	3.1315
18	3.147936	3.1	3.15	3.148	3.1479
19	3.141931	3.1	3.14	3.142	3.1419
20	3.132252	3.1	3.13	3.132	3.1323
21	3.162851	3.2	3.16	3.163	3.1629
22	3.149818	3.1	3.15	3.150	3.1498
23	3.145238	3.1	3.15	3.145	3.1452
24	3.142632	3.1	3.14	3.143	3.1426

Table 6: Pairwise comparisons.

Comparison	t-statistic	p-value
X_i vs X_{1i}	2.368	0.027 < 0.05
X_i vs X_{2i}	1.761	0.092
X_i vs X_{3i}	-1.438	0.164

X_{1i} vs X_{4i}	-0.440	0.664
X_{1i} vs X_{2i}	-2.180	0.040<0.05
X_{1i} vs X_{3i}	-2.379	0.026<0.05
X_{1i} vs X_{4i}	-2.369	0.027<0.05
X_{2i} vs X_{3i}	-1.818	0.082
X_{2i} vs X_{4i}	-1.763	0.091
X_{3i} vs X_{4i}	1.334	0.195

Statistical Justification

Without loss of generality, suppose X follows a normal distribution with mean μ and variance σ^2 , i.e., $X \sim N(\mu, \sigma^2)$. By properly truncation, X^* can still approximately normally distributed with mean μ_{x^*} and variance σ^2 , where μ_{x^*} may be different from μ_x due to rounding error. Set $Y = X - X^*$ considering that X and X^* are dependent, then under sample with n subjects of the population, the following t -statistic can be used to test the null hypothesis $H_0: \mu_x = \mu_{x^*}$ given in

$$t = \sqrt{n} \left(\bar{X} - \bar{X}^* \right) / \sqrt{s_x^2 + s_{x^*}^2 - 2\text{Cov}_{\text{obs}}(X, X^*)} = \bar{Y} / \sqrt{s_y^2/n}$$

where s_x , s_{x^*} and s_y are sample standard deviation of X , X^* and Y respectively. Meanwhile, $\text{Cov}_{\text{obs}}(X, X^*) = \left[\sum_{i=1}^n (X_i - \bar{X})(X_i^* - \bar{X}^*) \right] / (n-1)$ is the sample covariance of X and X^* . Under the null hypothesis, the paired t statistic follows a t distribution with $n - 1$ degrees of freedom. Generally, we reject null hypothesis if $|t| > t_{\alpha/2, n-1}$ where $t_{\alpha/2, n-1}$

is the $(1 - \alpha/2)^{\text{th}}$ quantile for a t distribution with $n - 1$ degrees of freedom. Thus, under the alternative hypothesis $H_a: \mu_x - \mu_{x^*} \neq 0$, the t statistic could be written as

$$t = \frac{\bar{Y}}{\sqrt{s_y^2/n}} \sim t_{n-1}(\delta)$$

where $t_{n-1}(\delta)$ denotes a non-central t distribution with a non-centrality parameter

$$\delta = \sqrt{n} \mu_y / \sigma_y = \sqrt{n} (\mu_x - \mu_{x^*}) / \sqrt{[\sigma_x^2 + \sigma_{x^*}^2 - 2\text{Cov}(X, X^*)]} \quad (2)$$

Obviously, $\delta = 0$ under H_0 . On the other hand, since X^* rounds to d^{th} decimal places, the maximum possible error due to truncation would be less than 10^{-d} and $|\mu_x - \mu_{x^*}| \leq 10^{-d}$. Thus, a small value of $\frac{\sqrt{n} 10^{-d}}{\sigma_y}$ would guarantee that X^* is not significantly different from X . Meanwhile, as $|\delta|$ decreases, the non-central t distribution under the alternative hypothesis will get closer to the central t distribution, which under the null hypothesis. Thus, as shown in the (Figure 1), if $|\delta|$ is small enough, the probability of correctly rejecting the null hypothesis, which is the power of the t -test, will be very low. Therefore, we want to find a boundary of δ , i.e.

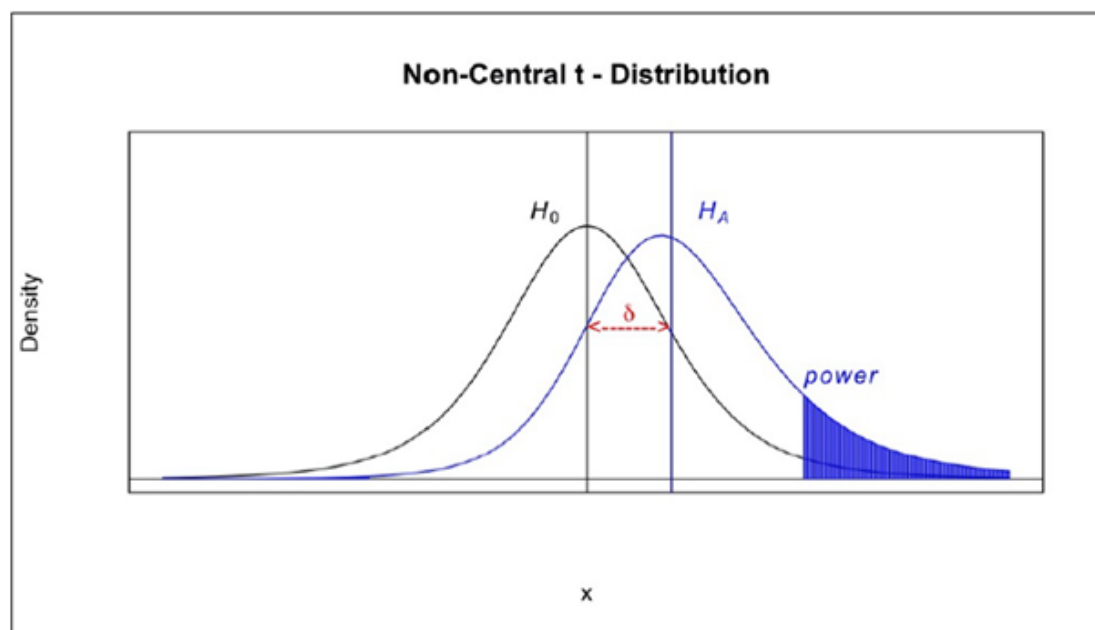


Figure 1: Density and power of non-central t-distribution.

$$\text{Power} = P(\text{reject } H_0 | H_0 \text{ is false}) = P\left(|t| > t_{\frac{\alpha}{2}, n-1} \mid \delta \neq 0\right) < 10\%$$

Based on the ratio of μ/σ , confidence level α , and degree of freedom df , power could be calculated by

$$\text{power} = 1 - F_{df, \delta}\left(t_{df, 1-\frac{\alpha}{2}}\right) + F_{df, \delta}\left(t_{df, \frac{\alpha}{2}}\right) < 10\% \quad (3)$$

where $F_{df, \delta}(t)$ is the CDF of non-central t distribution with non-central parameter δ .

Accordingly, power would keep increasing as δ enlarges, and a proper limitation of δ could guarantee the power small enough under the alternative hypothesis. Thus, we could not reject H_0 with a relatively high probability. (Table 7) demonstrate the relationship between non-central parameter, sample size, and power, which shows that the power would smaller than 10% when $\mu/\sigma = 0.1$ and sample within 30 subjects. Therefore, for the constant C , generally given $C = 10\%$, we said X^* is not significantly different from X when

Table 7: Relation between n , δ and power, by μ/σ .

n	$\mu/\sigma = 0.2$		$\mu/\sigma = 0.1$		$\mu/\sigma = 0.05$	
	δ	power	δ	power	δ	power
3	0.346	0.056	0.173	0.051	0.087	0.050
4	0.400	0.060	0.200	0.052	0.100	0.051
5	0.447	0.064	0.224	0.054	0.112	0.051
6	0.490	0.069	0.245	0.055	0.122	0.051
7	0.529	0.074	0.265	0.056	0.132	0.051
8	0.566	0.078	0.283	0.057	0.141	0.052
9	0.600	0.083	0.300	0.058	0.150	0.052
10	0.632	0.088	0.316	0.059	0.158	0.052
11	0.663	0.092	0.332	0.060	0.166	0.053
12	0.693	0.097	0.346	0.062	0.173	0.053
13	0.721	0.102	0.361	0.063	0.180	0.053
14	0.748	0.107	0.374	0.064	0.187	0.053
15	0.775	0.112	0.387	0.065	0.194	0.054
16	0.800	0.116	0.400	0.066	0.200	0.054
17	0.825	0.121	0.412	0.067	0.206	0.054
18	0.849	0.126	0.424	0.069	0.212	0.055
19	0.872	0.131	0.436	0.070	0.218	0.055
20	0.894	0.136	0.447	0.071	0.224	0.055
21	0.917	0.141	0.458	0.072	0.229	0.055
22	0.938	0.146	0.469	0.073	0.235	0.056
23	0.959	0.151	0.480	0.074	0.240	0.056
24	0.980	0.156	0.490	0.076	0.245	0.056
25	1.000	0.161	0.500	0.077	0.250	0.057
26	1.020	0.165	0.510	0.078	0.255	0.057
27	1.039	0.170	0.520	0.079	0.260	0.057
28	1.058	0.175	0.529	0.080	0.265	0.057
29	1.077	0.180	0.539	0.082	0.269	0.058
30	1.095	0.185	0.548	0.083	0.274	0.058

Table 8: The pass range and rounding error of significant digits.

# of digits	BE limit (%)	Pass range of 90% CI (%)	Maximum Rounding Error (%)
0	(80,125)	[79.5,125.5)	log80-log79.5
1	(80.0,125.0)	[79.95,125.05)	log80-log79.95
2	(80.00,125.00)	[79.995,125.005)	log80-log79.995
3	(80.000,125.000)	[79.9995,125.0005)	log80-log79.9995

$$\frac{|\mu_y|}{\sigma_y} \leq C \Leftrightarrow |\delta| \leq C', C' = C\sqrt{n}$$

At the same time, d is the minimum decimal that could let the above inequation always holds. We know that $\forall d, |\mu_x - \mu_{x^*}| \leq 10^{-d}$, and then $|\delta| = \frac{\sqrt{n}|\mu_x - \mu_{x^*}|}{\sigma_y} \leq \frac{\sqrt{n}10^{-d}}{\sigma_y}$. Thus, d should be the only number that can satisfy

$$\frac{\sqrt{n}10^{-(d+1)}}{\sigma_y} < \frac{\sqrt{n}10^{-d}}{\sigma_y} \leq C < \frac{\sqrt{n}10^{-(d-1)}}{\sigma_y}$$

After simplification, d should be

$$-\log C\sigma_y \leq d < 1 - \log C\sigma_y \quad (4)$$

The above argument could be applied similarly to a more general situation where a transformation is performed. Let $f(X)$ be the function of transformation of X , while $f(X)$ assumes normality. In this case, the hypothesis of interest becomes to

$$H_0 : f(\mu_x) = f(\mu_{x^*}) \text{ vs } H_a : f(\mu_x) \neq f(\mu_{x^*}) \quad (5)$$

For function which satisfied $\bar{f}(X) \approx f(\bar{X})$ and $\bar{f}(X^*) \approx f(\bar{X}^*)$, the null hypothesis can be tested by the hypothesis of

$$H_0 : \bar{f}(X) = \bar{f}(X^*) \text{ vs } H_a : \bar{f}(X) \neq \bar{f}(X^*)$$

By Taylor's expansion,

$$f(\bar{X}) - f(\bar{X}^*) \approx f'(\mu_x)(\bar{X} - \bar{X}^*)$$

Thus, set $Y' = f(X) - f(X^*)$, the above hypotheses can be tested by the following test statistic

$$t_f = \frac{\sqrt{n}Y'}{\sqrt{s_{Y'}^2}} \sim t_{n-1}(\delta)$$

Therefore, under the null hypothesis, t_f approximately follows a t distribution with $n - 1$ degrees of freedom and under alternative hypothesis, t_f can be written as

$$t_f \approx \frac{\sqrt{n}Y}{\sqrt{s_Y^2}} \sim t_{n-1}(\delta)$$

where δ is still the same noncentrality parameter as defined in (2). Meanwhile, if we choose the significant decimal digits property, we could guarantee δ small enough, and the probability that X and X^* are statistically different small as well. In conclusion, the proposed procedure works as well for data after transformation. Generally, AUC (area under the plasma concentration-time curve) and C_{\max} (time to achieve maximum concentration) in bioavailability and bioequivalence studies skew to the right. As the result, log-transformation is recommended. Thus, to illustrate the use of the proposed procedure for transformed data, consider a log-transformation, i.e., $f(X) = \log_e(X)$

$$H_0 : \log_e(\mu_x) = \log_e(\mu_{x^*}) \text{ vs } H_a : \log_e(\mu_x) \neq \log_e(\mu_{x^*})$$

Then $f'(\mu_x) = \log_e'(\mu_x) = \frac{1}{\mu_x}$, and the test statistic is given by

$$t_{\log_e} \approx \frac{\mu_x \sqrt{n} [\log_e(\bar{X}) - \log_e(\bar{X}^*)]}{\sqrt{s_x^2 + s_{x^*}^2 - 2\text{Cov}_{\text{obs}}(X, X^*)}} \sim t_{n-1}(\delta)$$

In this case, the proposed procedure is useful for the determination of the number of significant decimal digits to maintain a certain degree of accuracy and precision for the assessment of bioequivalence.

Application on BE study

In a bioequivalence study, suppose X_T is a pivotal parameter of test drug with standard deviation σ and X_R is for reference drug with the same variance. If d^{th} decimal places of BE limit is considered the significant digit we would like to keep, the confidence interval also needs to be rounded up with d^{th} digits.

For raw data of samples X_T and X_R , by two sample t-test of hypothesis testing $H_0 : X_T, X_R$ are bioequivalent vs $H_a : X_T, X_R$ are not bioequivalent. The confidence interval under α confidence level of the difference of their population mean $\mu_{X_T} - \mu_{X_R}$ is $[L, U]$ as follows

$$L = (\bar{X}_T - \bar{X}_R) - t_{n_T+n_R-2, \alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$U = (\bar{X}_T - \bar{X}_R) + t_{n_T+n_R-2, \alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$s_p^2 = \frac{[(n_T - 1)s_{X_T}^2 + (n_R - 1)s_{X_R}^2]}{(n_T + n_R - 2)}$$

where \bar{X}_T, \bar{X}_R are samples mean, n_T, n_R are samples size and $s_{X_T}^2, s_{X_R}^2$ are samples variance. We said T and R are bioequivalence if interval $[L, U]$ is totally covered by the BE limit $[-0.2\mu_r, 0.2\mu_r]$,

which means $\left[\frac{L}{\mu_r}, \frac{U}{\mu_r} \right]$ needs inside $[-20\%, 20\%]$. If BE limit with d significant digits, the pass range would be

$$\left[-20 - \frac{10^{-d}}{2}, 20 + \frac{10^{-d}}{2} \right] \times 100\%$$

In general, let $f(X)$ be the function of transformation of X , while $f(X)$ assumes normality. For transformed data $f(X_T)$ and $f(X_R)$, the hypothesis becomes to

$$H_0 : f(\mu_{X_T}) = f(\mu_{X_R}) \text{ vs } H_a : f(\mu_{X_T}) \neq f(\mu_{X_R})$$

The 90% confidence interval of $f(\mu_{X_T}) - f(\mu_{X_R})$ is $[L, U]$, where

$$L = \left(\bar{f}(X_T) - \bar{f}(X_R) \right) - t_{n_T+n_R-2, \frac{\alpha}{2}} \sqrt{s_{f_p}^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$U = \left(\bar{f}(X_T) - \bar{f}(X_R) \right) + t_{n_T+n_R-2, \frac{\alpha}{2}} \sqrt{s_{f_p}^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$s_{f_p}^2 = \frac{\left[(n_T - 1) s_{f(X_T)}^2 + (n_R - 1) s_{f(X_R)}^2 \right]}{(n_T + n_R - 2)}$$

By conclusion in Section 5, when $|\delta| \leq C\sqrt{n}$, $[L^*, U^*]$ are not significant different with $[L, U]$, where L^* and U^* are the truncated value of L and U , respectively.

Therefore, for log-transformed data $\log X_T$ and $\log X_R$ in BE study, under the hypothesis

$$H_0 : \log_e X_T, \log_e X_R \text{ are bioequivalent}$$

$$H_a : \log_e X_T, \log_e X_R \text{ are not bioequivalent}$$

The $(1 - \alpha) \times 100\%$ confidence interval of $(\mu_{\log_e X_T} - \mu_{\log_e X_R})$ is $[L_{\log_e}, U_{\log_e}]$ as follow

$$L_{\log_e} = \left(\overline{\log_e X_T} - \overline{\log_e X_R} \right) - t_{n_T+n_R-2, \frac{\alpha}{2}} \sqrt{s_{\log_e p}^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$U_{\log_e} = \left(\overline{\log_e X_T} - \overline{\log_e X_R} \right) + t_{n_T+n_R-2, \frac{\alpha}{2}} \sqrt{s_{\log_e p}^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right)}$$

$$\text{where } s_{\log_e p}^2 = \frac{\left[(n_T - 1) s_{\log_e X_T}^2 + (n_R - 1) s_{\log_e X_R}^2 \right]}{(n_T + n_R - 2)}, \quad \mu_{\log_e X_T} \text{ and}$$

$\mu_{\log_e X_R}$ are the population means, $\overline{\log_e X_T}$ and $\overline{\log_e X_R}$ are samples mean, $s_{X_T}^2$ and $s_{X_R}^2$ are samples variance of log-transformed data of test drug and reference drug, respectively.

$$\text{Since } \log_e(\mu_{X_T}) - \log_e(\mu_{X_R}) \in [L_{\log_e}, U_{\log_e}] \Leftrightarrow \frac{\mu_{X_T}}{\mu_{X_R}} \in [L, U] = [\exp(L_{\log_e}), \exp(U_{\log_e})]$$

We said X_T and X_R are bioequivalence if interval $[L, U]$ is totally covered by the BE limit $[80\%, 125\%]$. If BE limit with d significant

$$\text{digits, the pass range would be } \left[80 - \frac{10^{-d}}{2}, 125 + \frac{10^{-d}}{2} \right] \times 100\%$$

Thus, the significant digits of confidence interval could be determined by the methods as above. For untransformed data and log-transformed data, the significant digits should be the smallest

d that could reach $[L^*, U^*]$ are not significant different with $[L, U]$. In addition, for all d , the maximum rounding error due to the truncation would be less than 10^{-d} . Meanwhile, the rounding error of confident interval after log-transformed would less than $\log_e 10^{-d}$. (Table 8) shows the pass range of log-transformed data under BE limit with different significant digits.

Discussion

Statistical properties and justification of the proposed procedure of determining the number of significant digits in biomedical research such as bioequivalence and/or biosimilar studies were studied under the assumption of normality. In practice, other distributions such as Weibull distribution may be more appropriate to describe data collected from biomedical research such as dissolution study of the oral solid dosage form of a drug product. In this case, a similar concept could be applied to provide a valid statistical justification. In the descriptive analysis, the decimal places of descriptive statistics, such as mean, standard deviation, minimum, maximum, range, relative standard deviation (RSD), and the coefficient of variation (CV), and statistical inferences such as confidence intervals and p-values are always a concern, which is related to maintain the desired degree of accuracy and precision. In the interest of consistency, it is recommended that the same number of rounding numbers be used for descriptive statistics and statistical inferences derived from the analysis results. In some cases, the analytical results may be expressed in a scientific form (e.g., 1.32×10^5 or 9.2×10^{-7}) and the proposed procedure can be applied to its significant part (i.e., 1.32 for 1.32×10^5 or 9.2 for 9.2×10^{-7}) or its log (base 10) transformation.

When analytical results involving different datasets, it is suggested that each dataset keeps its own decimal digits determined by its standard deviation to maintain the same degree of accuracy and precision. A typical example is the assessment of dose proportionality in pharmacokinetic (PK) study. The purpose of a dose proportionality study is usually to show that there is a linear relationship between dose and AUC within a given range. In other words, doubled the dose, the AUC value is expected to be doubled. However, a high dose will generally produce a large variability in AUC values. As a result, low doses, median dose, and high dose are expected to have different numbers of decimal places to achieve the same degree of accuracy and precision. In the interest of keeping the same rounding numbers, we may consider the AUC values adjusted for dose and then apply the proposed procedure to determine the number of significant digits.

Acknowledgement

None.

Conflict of Interest

The authors have declared no conflict of interest.

References

1. AMA (2020) AMA Manual of Style: A Guide for Authors and Editors A Guide for Authors and Editors. Oxford University Press, New York, USA.
2. APA (2020) Publication manual of the American Psychological Association 7th (edn.), American Psychological Association, Washington DC, USA.
3. Chow SC (2000) Significant digits in analytical research. Presented at Amgen, Inc, Thousand Oaks, CA.
4. Chow SC and Liu JP (2008) Design and Analysis of Bioavailability and Bioequivalence Studies. 3rd (Edn.), Chapman and Hall/CRC Press, Taylor & Francis, New York.
5. Chow SC, Song F, Bai H (2016) Analytical similarity assessment in biosimilar studies. *AAPS J* 18(3): 670-677.
6. Daniel WW and Cross CL (2018) Biostatistics: a foundation for analysis in the health sciences. Wiley, New York, USA.
7. FDA (2003) Guidance on Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations. Center for Drug Evaluation and Research (CDER), the US Food and Drug Administration, Rockville, USA.
8. FDA (2006) Guidance for Industry: Bioequivalence Guidance. Center for Veterinary Medicine. U.S. Food and Drug Administration, Department of Health and Human Services, Rockville, USA.