Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Pharmaceutical and Biomedical Analysis 37 (2005) 313–318

JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS

www.elsevier.com/locate/jpba

BIS repetita non placent—a statistical analysis of the number of sample preparations and number of injections required for chromatographic analyses

Anne-Francoise Aubry[a,*], Robert Noble[b,c], Mark S. Alasandro[a,1], Christopher M. Riley[a,2]

[a] Pharmaceutical R&D Department, Bristol-Myers Squibb Company, Wilmington, DE, USA
[b] Biometrics Department, Bristol-Myers Squibb Company, Wilmington, DE, USA
[c] Department of Mathematics and Statistics, Miami University, Oxford, OH 45056, USA

Received 11 June 2003; received in revised form 27 October 2004; accepted 29 October 2004
Available online 8 December 2004

Abstract

This article attempts to answer the question of how many replicate sample preparations and replicate chromatographic injections must be done to provide accurate results in chromatographic analyses of pharmaceuticals. Using a random selection of chromatographic runs obtained with 1–3 replicate preparations and duplicate injections, the variance associated with preparation-to-preparation and injection-to-injection variability were estimated by a mixed-model statistical analysis. The analysis also predicted the probability that two injections of the same sample preparation are not in agreement with each other. Results indicated that, with modern chromatographic equipment, duplicate injections do not improve the precision. The number of replicate preparations needed to provide accurate results for various types of analysis depends on the type of sample and the desired tightness of the specification limits.

Keywords: Replicate analysis; Replicate injection; Chromatography; Statistical analysis; Pharmaceuticals

1. Introduction

While manufacturers of chromatographic equipment claim superior injection-to-injection reproducibility, not all users (in particular in the pharmaceutical industry), are convinced that they can safely go to single injection without compromising the quality of the results of their chromatographic analyses. Results of an informal survey conducted with nine companies engaged in pharmaceutical analysis indicated that the number of sample preparations/number of injections varies widely within the industry. In this survey, sample preparation/injection combinations included 1/1, 1/2, 1/3, 2/1 (one company each), 3/1 (two companies) and 2/2 (three companies). In each company and often at each site within a company, decisions on the number of replicate analysis are usually made arbitrarily for the purpose of standardization, on the basis of the personal experience and risk acceptance of a handful of scientists. Two factors enter in this decision: a higher number of replicates theoretically improves precision; on the other hand, lowering the number of replicates allows for faster turnaround time, higher productivity and cost saving.

More than any other analytical instrumentation, chromatographic equipment cannot be assumed to be stable over time. A major contributor to this is the chromatographic column, which is susceptible to aging, chemical and physical changes. Other factors are the many mechanical modules, such as pumps or autoinjectors. The potential for day-to-day variability is addressed by introducing daily “system suitability” tests designed for controlling the performance of the

[*] Corresponding author. Tel.: +1 732 227 7624; fax: +1 732 227 3798.
E-mail address: anne.aubry@bms.com (A.-F. Aubry).

[1] Present address: Merck & Co., Inc., P.O. Box 4, West Point, PA 19486-0004, USA.

[2] Present address: Development Sciences, ALZA Corporation, 1900 Charleston Road, Mountain View, CA 94043, USA.

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.10.039

chromatographic system during quantitative analyses [1]. To date, studies of the analytical variability of chromatography have focused on evaluating and controlling the interday precision of quantitative analyses [2,3]. The objective of this work was to study more specifically the repeatability of chromatographic results (within the same chromatographic run) and use statistics to justify the number of replicate sample analyses both in number of samples and number of injections. A data set comprising actual chromatographic data collected within a period of 1 year was assembled and analyzed.

2. Experimental

2.1. Data collection

Randomly selected chromatographic runs used to produce reportable results within the year 2000 were collected. The randomization was done by selecting 12 dates of analysis in a randomization table, three of which were attributed to each group leader. Each group leader was asked to collect data produced on these days. If no analysis was performed on a given day, data produced on the closest date 15 days before or after the target date were collected. A data set of 1036 pairs of injections was assembled. It included results of various types of analysis (assay or potency, preservative assay, counter-ion assay, organic impurities, residual solvents) and various types of samples (drug substance, tablets, capsules, oral solutions and oral suspensions). Only data that were actually reportable (i.e. only chromatographic runs that satisfied system suitability requirements) were included.

2.2. Statistical analysis

The statistical analysis was performed using SAS version 8.0 (SAS Institute Inc., Cary, NC, USA). The analysis was two-fold. In the first test, variance component estimation, the portion of the overall result variability that could be attributed to preparation-to-preparation and injection-to-injection differences was evaluated. For this test, the data was first evaluated as a whole and then grouped in a 4 × 2 matrix as follows:

	Drug substance	Solid dosage forms	Liquid dosage forms (GMP)	Liquid dosage forms (GLP)
High concentration tests	×	×	×	×
Low concentration tests	×	×	×	×

where

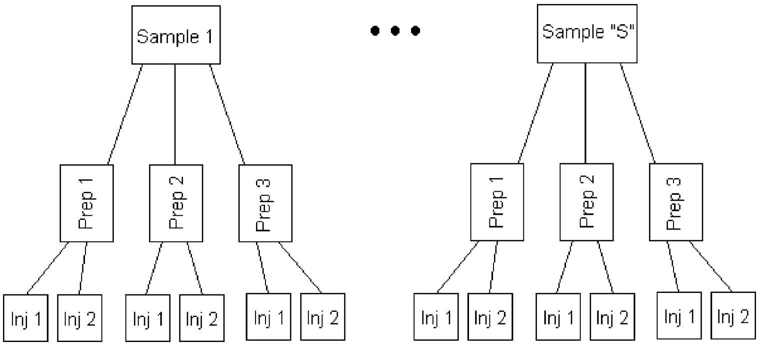
- *High concentration tests* included assay, potency, content uniformity, preservative assay, and counter-ion assay.
- *Low concentration tests* included residual solvents, organic impurities (i.e. related substances and degradation products) and inorganic impurities.
- *Drug substance* included drug substance and powder-in-bottle formulations.
- *Solid dosage forms* included tablet, capsules and blends.
- *Liquid dosage forms* (GMP) included solutions and suspensions for human use.
- *Liquid dosage forms* (GLP) were suspensions for toxicology studies.

In the second test, the probability of having a large difference between two injections was evaluated. The difference between injections was considered acceptable if it did not exceed the limits set a priori and defined in a Standard Operating Procedure, i.e. for high concentration tests, 1% and for low concentration tests, 0.03% for impurity level less than 0.10%, and 0.05% for impurity levels equal to or higher than 0.10%. This test was preformed using Bayesian Statistics. For this test, the significance of parameters, such as product type, instrument, type of test, product, instrument precision, etc. was evaluated.

3. Theoretical background

3.1. Variance component estimation

The response variable (result) is being modeled as a function of three random effects: sample, preparation and injection. The nested hierarchy of the effects is illustrated below:



The linear model is written as

$$Y_{ijk} = \mu + \beta_i + \gamma_{j(i)} + \delta_{k(ij)} + \varepsilon_{ijk}$$

where Y_{ijk} is the result for the ith sample, jth preparation and kth injection; μ the overall mean result; β_i the random effect of the ith sample; $\gamma_{j(i)}$ the random effect of the jth preparation within the ith sample; $\delta_{k(ij)}$ random effect of the kth injection within the jth preparation within the ith sample; and ε_{ijk} is the random error term.

Since the effects are random, the associated distributions are assumed to be the following:

$$\beta_i \overset{iid}{\sim} N(0, \sigma^2_\beta)$$
$$\gamma_{j(i)} \overset{iid}{\sim} N(0, \sigma^2_\gamma)$$
$$\delta_{k(ij)} \overset{iid}{\sim} N(0, \sigma^2_\delta)$$
$$\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$

Based on this model structure, the variance of an observation is given by

$$\text{var}(Y_{ijk}) = \sigma^2_\beta + \sigma^2_\gamma + \sigma^2_\delta + \sigma^2$$

Hence, the variability in the data is partitioned into four components: sample, preparation, injection, and random error. When we take the average of n injections within each preparation by sample combination, then the variance of the average is

$$\text{var}(\bar{Y}_{ij\cdot}) = \sigma^2_\beta + \sigma^2_\gamma + \frac{\sigma^2_\delta + \sigma^2}{n}$$

so as the number of injections increases, the variability in the average due to injection and random error

$$\frac{\sigma^2_\delta + \sigma^2}{n}$$

becomes smaller.

Test results within each classification grouping (product type, analysis type, and stage of development) are considered to be independent of all other test results and assumed to have the same probability of signaling. The customary model for these types of data is the binomial model. We let x represent the number of signals in n tests where each test has a probability p to signal.

Commonly, the normal distribution is used to approximate the binomial in order to obtain confidence interval estimates. Since the probability of signaling is small and the number of tests is large, the normal approximation would be considered to be adequate if the number of signals was greater than five for each analysis type [4]. In the present data set, a large portion of the data would not be well approximated using the classical methods.

An alternative would be to model the data via a Bayesian model structure. We maintain the binomial structure described previously, but now consider the probability of signaling to be a random variable also. We assume Jeffrey's non-informative conjugate prior distribution, Beta(0.5, 0.5), on p [5].

It can be shown that the posterior density of the probability of signaling given the data follows a Beta$(0.5 + x, 0.5 + n - x)$. The point estimate corresponds to the mean of the posterior density,

$$\hat{p} = \frac{0.5 + x}{1 + n}$$

The interval estimate, also known as a credible set, is obtained by taking the highest posterior density (HPD) [6].

For this test, the significance of parameters, such as product type, instrument, type of test, product, instrument precision, etc. was evaluated.

4. Results and discussion

The data set used in this study included a total of 1036 pairs of injections. The materials tested included drug substance samples, tablets, capsules, solutions, suspensions and powder blends. The analytical tests performed on drug substance comprised purity assays, residual solvents determinations, counter-ion determinations and organic impurity assays. The analytical tests for drug products included potency assays, content uniformity, preservative assays and organic impurity assays (i.e. related substances). All results were expressed in percent, either percent purity (for drug substance), percent of main component (for organic impurities), weight percent (counter ion) or percent of target (main component analysis in drug product). Because several types of tests were included, individual results ranged from 0.00% (for trace analysis) to 100% or more (for main component analysis results).

4.1. Variance estimates

The variance component estimates for the complete set are presented in Table 1. This table summarizes the contribution of each random effect, i.e. sample, preparation, and injection to the observed variability. Variability that could not be attributed to any of the three random effects was captured in the residual variance. As expected, the largest contribution was the effect of sample, reflecting simply that the sample set included a wide range of individual results. The effect of preparation within each sample was also significant and is discussed in detail below.

Table 1
Variance estimates for the complete data set

Covariance parameter	Estimate
Sample	2187
Preparation (within sample)[a]	1.123
Injection (within sample× preparation)[b]	0.0000
Residual	0.1752

[a] Variance associated with preparations for each sample.
[b] Variance associated with injections for each preparation by sample combination.

Table 2
Variance estimates for the parameter “preparation (within sample)” for data subsets by test and products

Type of test	Drug substance	Solid dosage forms	All liquid dosage forms
High concentration	0.1439	0.8405	2.8803
Low concentration	0.0000	0.0000	0.0000

Note: High concentration tests: assay, content uniformity, preservative assay and counter ion; low concentration tests: residual solvents and related substances.

On the other hand, the variance associated with the effect of injection was zero, indicating that injection-to-injection repeatability did not contribute to the overall observed variability. This observation suggests that there is no need to make multiple injections of a solution to improve method precision. The same conclusion was reached when the data set was broken up into subsets by type of material and type of test. In all cases, the estimated variance was zero, indicating conclusively the uselessness of replicate injections.

Contrary to what was reported for injections, data in Table 1 indicate that the contribution of preparation-to-preparation variability to the overall observed variability was significant. To investigate this effect further, the data set was divided into six subsets according to the type of material and the type of test and the variance estimated for each subset. Variance estimates are presented in Table 2. The variance estimates for low concentration tests were zero or close to zero for all types of materials indicating that for impurity determination, one preparation would give the same precision as multiple preparations. This may be counter-intuitive since lower concentrations are typically associated with higher standard deviations. This would be true if results were expressed in area counts or micrograms per milliliter. However, in this case, because the results are expressed in percent of the main component, the deviation becomes negligible.

For high concentration tests, the variance estimate was lowest for drug substance, reflecting the greater homogeneity of powders over formulated drug products. The variance estimate was highest for liquid products. If liquid products are further subdivided into GLP liquids (formulations used in animal toxicity studies) and GMP liquids (formulations used in clinical studies), the high variance is associated with analyses of the toxicology formulations as shown in Table 3. The higher variability for GLP samples was attributed to the fact that the sampling of the suspensions, done at the site of formulation, was not as accurate as is typical for an analytical laboratory.

Table 3
Variance estimates for the parameter “preparation (within sample)” for liquid dosage forms

Type of test	GLP liquid dosage forms	GMP liquid dosage forms
High concentration	3.3304	0.4922
Low concentration	0.0003	NA

Note: High concentration tests: assay, preservative assay; low concentration tests: related substances; NA: insufficient data available.

Table 4
Expected improvement in the test precision (expressed as variance) with increasing number of preparations (high concentration tests)

	Drug substance	Solid dosage forms	GMP liquid dosage forms	GLP liquid dosage forms
Variability	0.1439	0.8405	0.4922	3.3304
One preparation	0.7	1.8	1.4	3.6
Two preparations	0.5	1.3	1.0	2.5
Three preparations	0.4	1.0	0.8	2.1

Based on these observations, the next step was to attempt to determine the optimal number of preparations for purity/potency assays. Table 4 summarizes the variance associated with one, two or three preparations for drug substances, solid dosage forms and liquid dosage forms. From the variance estimates, the precision (expressed as variance) for one, two and three preparation was calculated using the following formula:

$$V = V_{\text{residual}} + \frac{V_{\text{prep}}}{n} \quad (1)$$

where V is the variance estimate for the test/product type, V_{residual} the estimate of the background variability, V_{prep} the variance estimate for the preparation within sample and n is the number of replicate sample preparations.

As the number of preparations n increases, the term V_{prep}/n approaches zero and V approaches V_{residual}. In other words, there is a lower limit to the assay precision that is equal to V_{residual}. The larger the V_{prep} compared to V_{residual}, the more improvement in the precision can be gained from multiple preparations. However, deciding on the exact number of replicates must take into account other considerations, such as acceptance limits, phase of development, and the purpose of the lot or batch (e.g. submission versus clinical, release versus stability).

In summary, the results of this statistical test indicated that injection-to-injection variability does not contribute to the precision of the quantitative determination for any product and any test. Similarly, for low concentration tests (determination of impurities), the preparation-to-preparation variability does not contribute to the precision indicating that there is no gain in precision by doing more than one preparation. However, for high concentration tests, preparation-to-preparation variability contributed significantly to the precision. The variability was lower for drug substance than for drug products. This can be explained by sample homogeneity, which is higher for powders than for solid or liquid formulations. In addition, sample preparation procedures for drug products are typically more complex than for drug substances and may contribute to the observed variability.

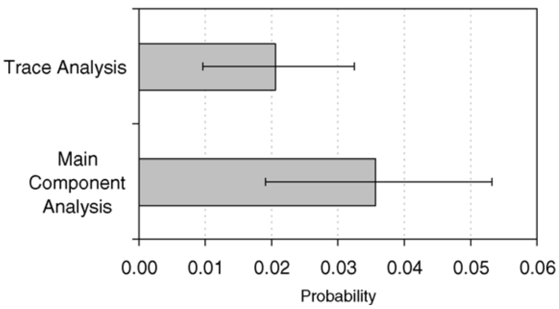


Fig. 1. Probability of a signal by type of test.

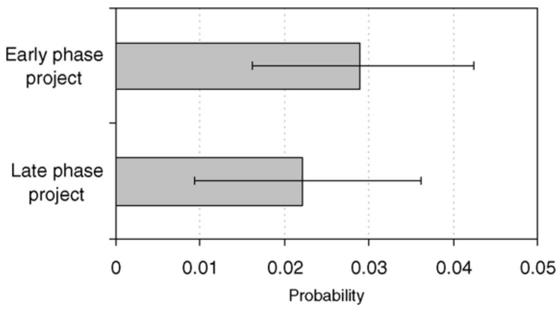


Fig. 3. Probability of a signal by stage of development of a drug candidate.

4.2. Estimation of the probability of inconsistency between duplicate injections

The probability that two injections of the same sample preparation are not consistent was estimated using Bayesian statistics. Acceptable differences between replicate injections had been established prior to starting the study and are provided in the experimental section. Failures to meet the criteria (i.e. unacceptable difference between injections of the same sample solution) were identified as signals in the statistical test. The probability of a signal in a test was calculated for various subsets of the data set and is presented in bar graphs. Fig. 1 presents the probability of a signal per type of test. There was no significant difference in the occurrence of injection-to-injection discrepancy between low concentration and high concentration tests. The mean probability was 3% or less but the upper confidence limit was about 5%. Looking at specific chromatographic runs that included failures, it can be noted that most occurrences of inconsistency between injections in high concentration tests were related to autosampler failures while in low concentration tests, they were mostly due to integration inconsistency. In the former, the instrument was the direct cause of the inconsistency, while in the latter, the data processing system or the analyst was at fault. Fig. 2 presents the probability of a signal graphed by type of product tested. Results for drug substance and solid dosage forms (including blends) were consistent with the previous result of an upper confidence limit of 5%. However, results for liquid dosage forms indicated a significantly higher probability of failure. When liquid products were separated between suspensions and solutions, only solutions showed a significant increase in failure rate. Since there were only two chromatographic runs included in the set for solutions and that one of the chromatographic run had several failures, the high rate of failure is not considered representative of this type of drug product. Finally, Fig. 3 presents results by phase of development. There was no significant difference between early and late projects even though early projects tend to use less rugged methods.

The probability of a signal in the test on the absolute differences between injections was not significantly related to the instrument precision (i.e. R.S.D. of six or more injections of the same solution done as part of system suitability) (p-value = 0.7942). This is an important conclusion as it indicates that these failures, when they are instrument related are random occurrences not directly related to the overall performance of the instrument during the analysis. Tightening the system suitability criterion would not decrease the frequency of these injection failures. The probability of a signal in the test on the absolute differences between injections was not significantly related to the instrument (p-value > 0.9995), product (p-value = 0.4732), or type of analysis (p-value = 0.1632).

In summary, the probability of two injections of the same sample exceeding the acceptable difference was less than 5% (based on upper confidence interval) overall. There was no significant increase in the probability of injection-to-injection imprecision in early-stage projects versus late-stage projects

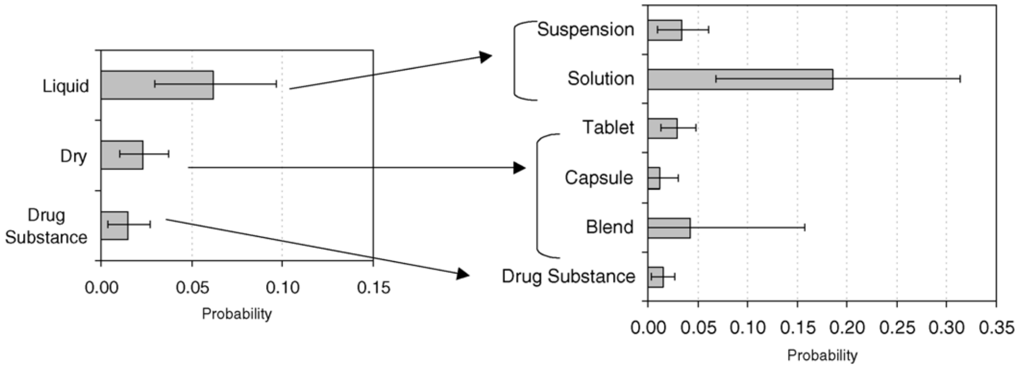


Fig. 2. Probability of a signal by type of material tested.

and the probability of failure was not related to the overall performance of the chromatographic system as measured by the R.S.D. between standard injections.

The data can be used to estimate the rate of injection failure. Assuming that in each failed pair of injections, only one injection was incorrect, if single injections instead of duplicate injections were made, it is anticipated that failures (whether instrument related or analyst-related) will occur half as frequently, causing up to 2–3% of the results to be inaccurate. This relatively high rate may be acceptable to some companies depending on their risk acceptance level. Others may consider alternatives to duplicate injections to give them added confidence in their results. One example would be to compare impurity profiles from duplicate preparations of drug substance to ascertain that there were no external contamination, integration error or instrument error.

5. Conclusion

The results of this study indicate that the number of injections does not affect the precision of the quantitative determinations. Results also indicate that a single preparation is sufficient to achieve good precision for the determination of impurities. For purity/potency assay, the number of preparations must be decided based on the desired precision, taking into account acceptance criteria and purpose of the testing. However, the study revealed that when duplicate injections are used, there is up to 5% chance (upper confidence limit) of observing poor injection-to-injection repeatability.

References

[1] W.B. Furman, et al., Pharm. Technol. 22 (1998) 58–64.

[2] M. Schiavi, E. Rocca, P. Ventura, Chemometrics Intell. Lab. Syst. 2 (1987) 303–312.

[3] W. Horwitz, R. Albert, J. Assoc. Off. Anal. Chem. 67 (1984) 81–90.

[4] R.L. Scheaffer, W. Mendenhall, L. Ott, Elementary Survey Sampling, WWS-Kent Publishing, 1990.

[5] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman & Hall, 1997.

[6] G. Casella, R.L. Berger, Statistical Inference, Duxbury Press, 1990.