ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
23 (2000) 581–589

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Design of experiment and data analysis by JMP® (SAS institute) in analytical method validation

Christine Ye [a,*], June Liu [b], Feiyan Ren [a], Ngozi Okafo [a]

[a] *Analytical Sciences, ALZA Corporation, Mountain View, CA 94303, USA*
[b] *Development Engineering, ALZA Corporation, Mountain View, CA 94303, USA*

Accepted 27 March 2000

Abstract

Validation of an analytical method through a series of experiments demonstrates that the method is suitable for its intended purpose. Due to multi-parameters to be examined and a large number of experiments involved in validation, it is important to design the experiments scientifically so that appropriate validation parameters can be examined simultaneously to provide a sound, overall knowledge of the capabilities of the analytical method. A statistical method through design of experiment (DOE) was applied to the validation of a HPLC analytical method for the quantitation of a small molecule in drug product in terms of intermediate precision and robustness study. The data were analyzed in JMP® (SAS institute) software using analyses of variance method. Confidence intervals for outcomes and control limits for individual parameters were determined. It was demonstrated that the experimental design and statistical analysis used in this study provided an efficient and systematic approach to evaluating intermediate precision and robustness for a HPLC analytical method for small molecule quantitation.

Keywords: Validation; Design of experiment; Precision; Robustness; HPLC

1. Introduction

Validation of an analytical method demonstrates the methods suitability for its intended purpose. Per ICH guideline, characteristics to be considered during the validation include specificity, range, linearity, accuracy, precision, limit of detection and quantitation, and robustness [1].

Intermediate precision of an analytical method expresses the closeness of agreement (degree of scatter) between a series of measurements. The measurements are obtained from multiple samplings from the same homogeneous population under prescribed conditions within a single laboratory. Typical variations to be studied include days, analysts, equipment, etc.

Robustness of an analytical method is its capacity to remain unaffected by small variations in method parameters. The robustness, which is determined by deliberately changing method parameters, provides an indication of its reliability during normal use. Typical variations studied

* Corresponding author. Fax: +1-650-5642258.

0731-7085/00/$ - see front matter
PII: S0731-7085(00)00335-6

for liquid chromatography are flow rate, temperature, different columns, influence of variations in mobile phase composition, pH, etc.

Conventionally, intermediate precision and robustness are studied by varying one parameter while keeping the rest at target levels and then examining the effect of this single change on method performance. This approach has drawbacks such as inability to detect any interactions among parameters and determine control limits when interactions are present. Second, it is unable to measure experimental error jointly, and therefore unable to calculate confidence intervals and control limits accurately and efficiently. Also, it is usually time consuming to perform the test due to multiple parameters to be examined. To enable efficient data collection and scientific data interpretation, a design of experiments (DOE) [2] in combination with statistical evaluation of data obtained from chromatographic analyses of a chemical compound extracted from oral tablet was employed. Data were analyzed in JMP® (SAS institute) using analysis of variance method (ANOVA) by least-square fit [3]. Confidence intervals for outcomes and control limits for individual parameters were determined based on the analysis results. Advantages of using JMP® in method validation will be discussed.

2. Experimental

2.1. Instrumentation and chromatographic conditions for assay of compound A

Chromatographic analyses for compound A were carried out using a HP 1100 system (HP Corp.), with an autosampler and a UV detector. Chromatograms were recorded and processed using ChromPerfect data analysis software (Justice Inc.).

A HP Zorbax Eclipse column (XDB-C18, 5 μm, 4.6 × 150 mm) (HP Corp.) and a Waters Symmetry-C8 column (5 μm, 3.9 × 150 mm) (Waters Corp.) were used with column temperature controlled at 40°C. The composition of the mobile phase was 25 mM phosphoric acid aqueous solution with pH the adjusted to 3.0 with triethylamine and acetonitrile (85:15). Flow rate was 1.5 ml/min and detector wavelength was at 278 nm. Compound A eluted at about 4 min under these conditions.

Table 1
Design of experiment for intermediate precision study

Run #	Pattern	Analyst	Day	Instrument
1	−+−	a	2	1
2	+++	b	2	2
3	−−+	a	1	2
4	+−−	b	1	1

2.2. System suitability test

The chromatographic system was equilibrated until a stable baseline was achieved. Six replicate injections of compound A standard solution were performed. The system was suitable for analysis if the chromatograms obtained met the proposed system suitability requirements: capacity factor ($K' \geq 1.5$ and tailing factor (T) $0.5 \leq T \leq 2.5$ for each of the five injections, response variation % relative standard deviation (%RSD) ≤ 3% and retention time (RT) variation %RSD ≤ 5% for five injections.

2.3. Intermediate precision study

Parameters examined in this intermediate study were analyst, day, and instrument. Prior knowledge suggested that interaction of these factors were negligible, a main effect fractional factorial design was generated with these three factors using JMP® software (Table 1). Nine tablets were prepared and tested for drug content for a total of four analyses according to the experimental design. Had it been necessary to study interactions in these experiments, a larger or full fractional factorial design would have been employed.

2.4. Robustness study

In this robustness study, seven factors were selected: % acetonitrile (ACN) in the mobile phase, pH of the mobile phase, detector wave-

Table 2
Design of experiment for robustness study

Run #	Pattern	%ACN	pH	Wavelength (nm)	Temperature (°C)	Flow rate (ml/min)	Buffer concentration (mM)	Column
1	−−−−+++	13	2.8	276	35	1.7	28	A[a]
2	−−+−−−+	13	2.8	280	45	1.3	22	A
3	−+−+−+−	13	3.2	276	45	1.3	28	B[a]
4	−++−+−−	13	3.2	280	35	1.7	22	B
5	+−−+−−−	17	2.8	276	45	1.7	22	B
6	+−+−−+−	17	2.8	280	35	1.3	28	B
7	++−−−−+	17	3.2	276	35	1.3	22	A
8	+++++++	17	3.2	280	45	1.7	28	A

[a] Column A is a symmetry column and column B is a Zorbax column.

length, column temperature, flow rate, buffer concentration and column types (from two different manufacturers with similar packing material). Prior knowledge suggested that these seven factors would not be highly interactive. Therefore, using JMP® software, a fractional factorial design was generated (Table 2). The effects of variations in chromatographic parameters were evaluated using the system suitability test results generated.

3. Results and discussion

DOE has been widely used for the design of multi-factor experiments. One of its advantages is that it provides efficient data collection and helps reduce the workload effectively. For example, the factors being studied in this validation would have yielded a total of six and 14 experiments for intermediate precision and robustness, respectively, without DOE. By applying DOE, the numbers of experiments were reduced to four and eight.

Other advantages of DOE include effective problem structuring, comprehensive data analysis that leads to more precise estimation of experimental errors, and therefore provides more power in detecting any statistical differences. These advantages will be demonstrated in the following discussions.

3.1. Intermediate precision

Intermediate precision was studied by testing nine tablets for drug content varying analyst, days, and instrument according to the design. Percent drug recoveries were obtained (Table 3) and analyzed by JMP®. All the results were first analyzed by least-square modeling to compare the mean difference due to any of these three factors. The model used was $y = \text{analyst} + \text{date} + \text{instrument} + \text{error}$. The result indicated that none of the factors was significant with $\text{Prob} \geq |t|$ as 0.65, 0.66, and 0.77 for analyst, day, and instrument, respectively (Table 4). Equal variance tests were

Table 3
Results for intermediate precision study

Sample #	Run 1	Run 2	Run 3	Run 4
1	97.12	98.34	97.83	98.75
2	98.04	97.69	97.84	98.18
3	93.39	94.77	94.92	95.08
4	94.86	96.27	96.37	96.65
5	96.21	97.76	97.46	97.56
6	95.76	96.12	95.96	95.47
7	91.38	92.01	91.72	91.26
8	99.13	98.98	98.90	98.33
9	98.68	99.03	98.82	98.62
Average	96.65	96.77	96.06	96.66
%RSD	2.34	2.36	2.65	2.55
Upper 95% CI	98.39	98.53	98.02	98.45
Lower 95% CI	94.90	95.02	94.11	94.34

Table 4
Parameter estimation for intermediate precision study

Mean of response	96.54			
Observations	36			
	Parameter estimate			
Term	Estimate	Standard error	t Ratio	Prob$>\|t\|$*
Analyst (a–b)	−0.18	0.40	−0.45	0.65
Day (1–2)	−0.18	0.40	−0.44	0.66
Instrument (1–2)	0.12	0.40	0.29	0.77

* Prob$>|t|$ stands for the probability of observing $|t|$ value which is greater than t ratio. It is obtained from the t distribution table. [Prob$>|t|$] of 0.05 is usually the cut off point to define statistical significance.

Table 5
Comparison of within run variation in intermediate precision study

	Run ID	Observation	Standard deviation	
	1	9	2.27	
	2	9	2.28	
	3	9	2.55	
	4	9	2.43	
Test	F Ratio	DF #	DF denominator	Prob$>F$*
O'Brien (0.5)	0.0427	3	32	0.99
Brown–Forsythe	0.0869	3	32	0.97
Levene	0.0628	3	32	0.98

* Prob$>F$ stands for the probability of observing F value which is greater than F ratio. It is obtained from the F distribution table. [Prob$>F$] of 0.05 is usually the cut off point to define statistical significance.

performed and demonstrated that there was no significant within run variation difference among the four runs, with Prob $\geq F$ as 0.99, 0.97, and 0.98 for the O'Brien, Brown–Forsythe, Levene tests, respectively (Table 5). Therefore, there were no analyst to analyst, day to day and instrument to instrument differences in the study, and data from those four runs were pooled together for confidence interval (CI) calculation (Table 6). Upper 95% CI of the mean was calculated as 97.32% and the lower 95% CI of the mean was 95.76%. The confidence interval was determined to be $\pm$ 0.78% if the number of determinations was 36. In pharmaceutical development, drug content uniformity is usually performed with $n = 10$. With that number of experimental determinations, the confidence interval was concluded to be $\pm$ 1.47% (Table 6). This indicated that with 95% probability, the difference between test value and true value will be no greater than $\pm$ 1.47%.

Table 6
Estimation of confidence interval of intermediate precision

Mean	96.54
Standard deviation	2.30
Standard error mean	0.38
Upper 95% mean	97.31
Lower 95% mean	95.76
T ($n = 36$, 95%)	2.03
N	36

Half width of CI $= \pm t_{(n=36,\ 95\%)} \times S/N^{1/2} = \pm 2.03 \times 2.29/N^{1/2}$
Half width of CI = 0.78%, $n = 36$
Half width of CI = 1.47%, $n = 10$

Table 7
System suitability results for robustness study

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
%RSD RT	0.22	0.31	0.06	0.09	0.78	0.20	0.96	0.24
%RSD peak area	0.24	0.06	0.11	0.09	0.73	0.06	1.43	0.06
K'_1	2.18	1.6	1.32	2.21	0.63	0.76	1.24	0.72
K'_2	2.17	1.59	1.38	2.21	0.6	0.76	1.22	0.71
K'_3	2.17	1.59	1.39	2.21	0.6	0.76	1.22	0.71
K'_4	2.17	1.59	1.38	2.22	0.6	0.76	1.21	0.71
K'_5	2.17	1.58	1.34	2.22	0.6	0.76	1.2	0.71
K'_6	2.16	1.58	1.43	2.22	0.6	0.76	1.2	0.7
T_1	1.17	1.11	0.98	1.00	0.99	0.94	1.23	1.03
T_2	1.11	1.11	1.00	1.01	1.00	0.97	1.24	1.23
T_3	1.07	1.23	1.00	1.02	0.96	0.98	1.25	1.21
T_4	1.06	1.23	1.01	1.02	0.94	0.97	1.14	1.17
T_5	1.17	1.16	1.02	1.05	0.96	0.98	1.04	1.11
T_6	1.09	1.09	1.02	0.94	0.98	1.01	1.19	1.07

Table 8
Parameter estimation on tailing factor (T)

Summary of fit

R square	0.75
Root mean square error	0.052
Mean of response	1.07
Observations	48

Parameter estimates

Term	Estimate	Standard error	t ratio	Prob$>\|t\|$
ACN % (13–17)	0.002	0.0075	0.22	0.83
PH of MP (2.8–3.2)	−0.02	0.0075	−1.94	0.059
Wavelength (276–280)	−0.0004	0.0075	−0.06	0.96
Column temperature (35–45)	0.001	0.0075	0.11	0.91
Flow rate (1.3–1.7)	0.01	0.0075	1.50	0.14
Buffer concentration (22–28)	0.08	0.0075	1.44	0.16
Column (A–B)[a]	0.08	0.0075	10.44	<0.0001

[a] Column A, symmetry A; column B, Zorbax B.

3.2. Robustness

System suitability tests were performed according to the design in Table 2, across eight chromatographic runs. Results were obtained for area response, retention time, tailing factors and capacity factors of the peak of interest, and %RSD was calculated and examined for robustness (Table 7).

3.2.1. Variation in retention time and area response

Percent RSD for retention time for eight experimental runs was between 0.06 and 0.96% (Table 7), which is well within the proposed acceptance criterion of $\leq 5\%$. Percent RSD for area response was from 0.06 to 1.43%, which also passed the proposed acceptance criterion of $\leq 2\%$.

3.2.2. Tailing factor

Tailing factors (T) for each of the 48 injections (six injections per run × eight runs) were entered in JMP® software and analyzed using the ANOVA method by least-square fit (Table 8).

Fitting results revealed that Prob > $|t|$ was greater than 0.05 for %ACN, wavelength, mobile phase, flow rate, buffer concentration, and column temperature. It demonstrated that no significant differences were observed when changing the above factors within the tested ranges. Statistically significant difference was observed between the two column types, which was demonstrated by Prob > $|t|$ smaller than 0.05. However, parameter estimate (change for the value of T when the factor is changed from center point) for column type was 0.08, which was very small compared with the mean value of 1.07. Prediction profile on 95% confidence interval (Fig. 1) showed that all T values would be within acceptance criterion, $0.5 \leq T \leq 2.5$, if parameters were changed within their testing ranges. Therefore, the conclusion was made that the tailing factor was acceptable when the chromatographic parameters were changed within the tested range.

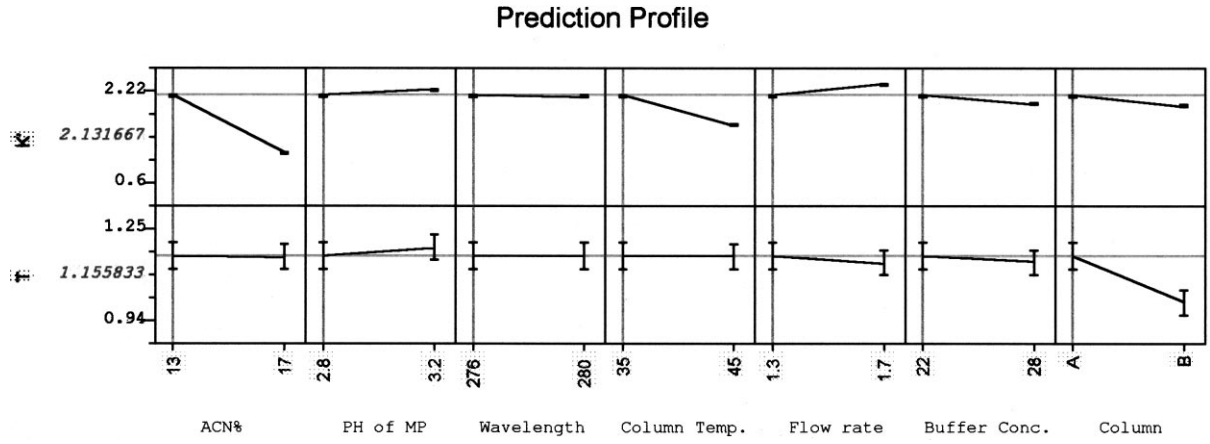

Fig. 1. Prediction profile for T and K'.

Table 9
Parameter estimation on capacity factor (K')

Summary of fit	
R Square	0.999
Root mean square error	0.016
Mean of response	1.33
Observations	48

Parameter estimates

Term	Estimate	Standard error	t ratio	Prob > $\|t\|$
ACN% (13–17)	0.51	0.0023	219.29	<0.0001
PH of MP (2.8–3.20)	−0.05	0.0023	−21.08	<0.0001
Wavelength (276–280)	0.01	0.0023	4.87	<0.0001
Column temperature (35–45)	0.26	0.0023	112.62	<0.0001
Flow rate (1.3–1.7)	−0.10	0.0023	−41.26	<0.0001
Buffer concentration (22–28)	0.08	0.0023	32.97	<0.0001
Column (A–B)[a]	0.09	0.0023	39.46	<0.0001

[a] Column A, symmetry A; column B, Zorbax B.

3.2.3. Capacity factor

Capacity factors (K') for each of the 48 injections were also analyzed using the ANOVA method by least -squares fit (Table 9). The fitting result revealed that each of the seven factors showed significant differences between the two levels tested. However, parameter estimates for pH of mobile phase, wavelength, flow rate, buffer concentration, and column type were -0.05, 0.01, -0.1, 0.08, and 0.09, respectively, which

Contour Profiler

Horiz	Vert	Factor	Current Setting
X	_	ACN%	13.5
_	_	PH of MP	2.8
_	X	Column Temp.	37
_	_	Flow rate	1.3
_	_	Buffer Conc.	28
_	_	Column	1

Response	Lower 95% Individual K'
Current K'	1.51
Low limit	1.5
High limit	None

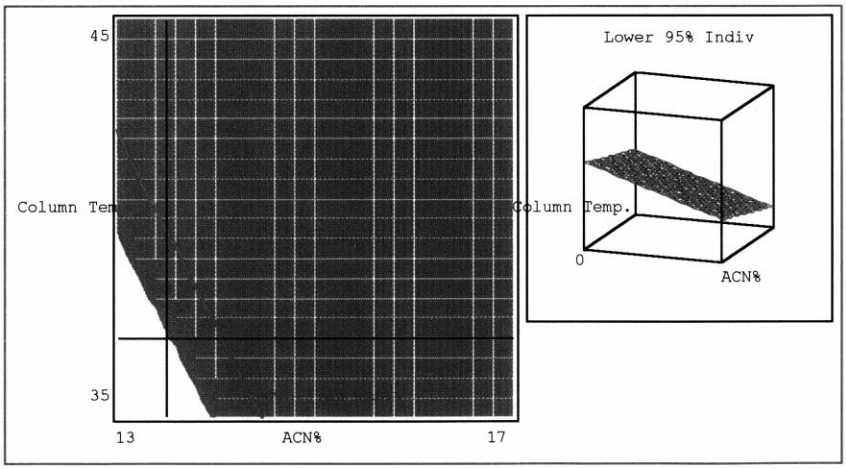

Fig. 2. Determination of control limits for %ACN and column temperature ($K' \geq 1.5$).

Contour Profiler

Horizontal	Vertical	Factor	Current Setting
X	–	ACN%	14.9
–	–	PH of MP	2.8
–	X	Column Temp.	40
–	–	Flow rate	1.3
–	–	Buffer Conc.	28
–	–	Column	1

Response	Lower 95% Individual K'
Current K'	1.00
Low limit	1.0
High limit	None

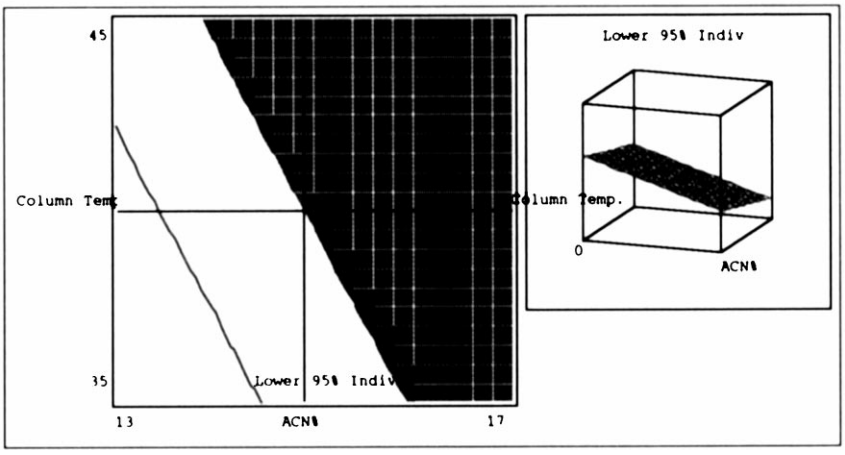

Fig. 3. Determination of control limits for %ACN and column temperature ($K' \geq 1.0$).

were very small compared with the mean value of 1.33. Therefore, they will not be considered as primary factors in changing the capacity factor. Percent ACN and column temperature were shown to have the most significant impact on K' with parameter estimates of 0.51 and 0.26, respectively. To determine the control limits for %ACN and column temperature, contour profiles were obtained using %ACN, column temperature, flow rate, pH of mobile phase, and buffer concentration as variables (Fig. 2). Wavelength was excluded from this calculation because of its extremely insignificant impact on K'. Flow rate, pH of mobile phase, buffer concentration and

column type were, respectively, set at 1.3 ml/min, 2.8, 28 mM, and 1 (type B: Zorbax B), each of which provide the low end of K', since the criterion for K' was ≥ 1.5 for each injection. Lower 95% individual K' (calculated based on two sides 95% confidence interval) was used in the calculation to ensure that all injections would meet the criteria with high probability. A contour profile for %ACN and column temperature (Fig. 2) was first obtained with the low limit set at 1.5 ($K' \geq 1.5$). The clear area in the profile represents the range where column temperature and %ACN can be changed concomitantly to achieve $K' \geq 1.5$, while the shaded area represents the range where $K' < 1.5$. From this profile, the control limit can be set with a temperature of 35–37°C and %ACN 13–13.5%, which was outlined by the block in the lower left hand corner of the profile.

It was noted at this point that the control limit was impractically set too tight, which led us to explore the feasibility of lowering acceptance criterion for K'. Generally, controlling $K' \geq 1.5$ plays a key role when separation of impurity from peak of interest is involved. In this particular method, there was no separation involved because only a single peak of interest (compound A) was being analyzed. Therefore, the acceptance criterion for K' was lowered to ≥ 1.0 to increase analysis efficiency and allow for practical control limits for %ACN and column temperature. The limit of $K' \geq 1.0$ was applied to the response contour (Fig. 3). The new control limits were determined to be temperature 35–40°C, and %ACN 13–15%.

In conclusion, the acceptance criteria for system suitability were: %RSD for area $\leq 2.0\%$, %RSD for RT $\leq 5.0\%$, $0.5 \leq T \leq 2.5$ and $K' \geq 1.0$. Control limits for all the HPLC parameters tested in this robustness study will be set as: %ACN $14\% \pm 1\%$, column temperature 35–40°C, pH of mobile phase 3.0 ± 0.2, buffer concentration 25 ± 3 mM, wavelength 278 ± 2 nm, flow rate 1.5 ± 0.2 ml/min. When analyses are performed within the above control limits, it will give at least 97.5% assurance that system suitability acceptance criteria will be achieved for each injection.

4. Conclusions

An intermediate precision and robustness study for this analytical method validation were successfully completed through DOE and data were analyzed by JMP® (SAS Institute). Confidence intervals for percent drug content were determined, and control limits for chromatographic parameters were set in a reasonable and reliable range. It was demonstrated that DOE with JMP® software can result in efficient data collection, comprehensive and rapid data analysis, and accurate conclusions. The combined DOE and JMP® software provides an efficient tool for the systematic analysis of multi-factorial design of analytical method validation.

Acknowledgements

The authors are grateful to Bing Luo and Anissa Wong for performing intermediate precision studies. Funding for this study was provided by the ALZA Corporation on behalf of the Crescendo Pharmaceuticals Corporation.

References

[1] International Conference on Harmonization: Guideline on Validation of Analytical Procedures: Methodology; availability, Federal Register, vol. 62, No. 96, 1997.
[2] G.E.P. Box, W.G. Hunter, J.S. Hunter (Eds.), Statistics for Experimenters: an Introduction to Design, Data analysis, and Model Building, Wiley, New York, 1978.
[3] JMP® Statistics and Graphic guide, 3.1, SAS Institute, Inc. (1995).