

# A

# Managing Variability during Aerodynamic Particle Size Measurement

Liz Cliff<sup>1</sup>, Debbie Kraus<sup>2</sup>, Tamsin Jenkins<sup>1</sup>, Sarah Walker<sup>1</sup>  
 Pharmaceutical Sciences<sup>1</sup>, Pfizer Global R&D, Sandwich, UK.  
 Non-Clinical Statistics<sup>2</sup>, Pfizer Global R&D, Sandwich, UK.

## INTRODUCTION

Aerodynamic particle size (APS) measurement is a key quality attribute of inhaled drug products. A number of factors can have an effect on APS measurement some of which can be understood and controlled (e.g. environmental testing conditions), some factors may be understood but outside our control (e.g. device to device variability) and some factors may not yet be identified. Formulation/Process development will typically take place at a time when sources of variability associated with APS measurement are not fully understood. Consequently, there is great potential for comparison between different formulation batches to be confounded with analytical method biases in the situation where many batches are to be compared as is the case when using a statistical design of experiments (DoE) approach. This poster describes how we have attempted to minimise analytical bias during development of a new product.

## SOURCES OF VARIABILITY

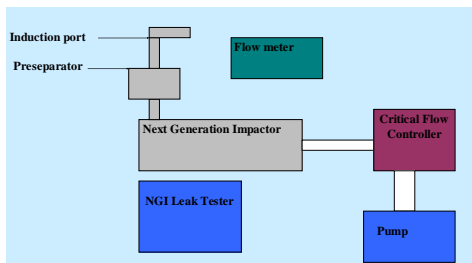
Next Generation Impaction (NGI) testing is a critical performance test for dry powder inhaled products. This apparatus is used in the determination of the particle size distribution. The test is relatively complex and involves many individual stages. The labour intensive procedure involves careful preparation of the test apparatus prior to aerosolisation of the dose from the inhaler into the NGI. The drug deposited on each stage of the NGI is quantitatively recovered and solutions analysed by HPLC. Due to the complexity of the NGI test, it is recognised that many sources exist which may contribute towards the variability observed in the data generated. These sources of variability can include factors such as analyst, device, lab, NGI apparatus and environmental conditions.

HPLC analysis is thought to provide minimal contribution towards variability.

## STANDARDISATION OF TEST

In order to standardise the NGI test, each testing workstation should be set up with components in compliance with USP criteria for test Apparatus 5 for dry powder inhalers. This also includes the type and dimensions of tubing used to connect the pump to the critical flow controller, and from the critical flow controller to the NGI. A schematic of the NGI test apparatus is shown below in Figure 1.

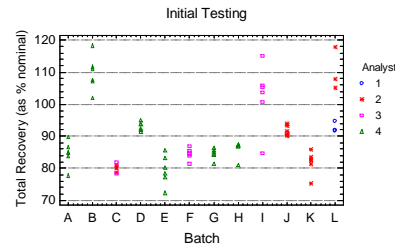
Figure 1. NGI test apparatus



## MANAGING VARIABILITY

An early formulation study was performed and 12 blends were filled into devices for NGI testing. For each of the batches, 6 replicate, one-shot-firing tests were performed (72 NGI tests in total) as shown in Figure 2. Due to logistical ease, typically, all six replicates for a single batch were performed by the same analyst on the same day and with the same NGI instrument. An observation was made for batch L, which happened to be tested by two different analysts on different days, that there was a difference between the sets of results for the two analysts. This prompted us to question whether any apparent differences between the batches could be down to intermediate precision variability (i.e. analyst/instrument/set-up conditions etc.), or alternatively, could any important differences between the batches have been masked by intermediate precision variability.

Figure 2. Initial test data from DOE

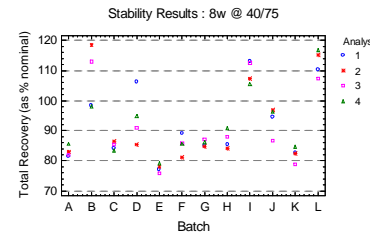


In an ideal world, in order to get the most precise comparisons of APSD between the batches, the testing would all be done by the same analyst on the same day, same instrument etc. Due to the length of time taken to perform each NGI test, this scenario is not possible for the vast majority of studies. Consequently, for stability testing with this study and for other studies that followed, the approach we have taken at Pfizer is to structure NGI testing in such a way so as to reduce potential biases in the comparisons we make between different formulation batches. One such test plan and the resulting data are shown below.

We see in the test plan below, for each of the 12 formulation batches, four replicate NGI tests were done, each one being performed by a different analyst on a different day. Also note that within each batch, two of the replicates were done using NGI instrument 1 and the other two replicates were done using instrument 2. Similarly two tests per batch were done in the morning and two were done in the afternoon. As a result, the mean result for each batch will average out any biases that potentially may exist between analysts or instruments for example, rather than comparisons of batches being completely confounded with differences between analysts/days/instruments as was the case previously.

Day	am/pm	NGI Instrument	Analyst	Batch
1	am	1	A	A, B, C
		2	B	D, E, F
	pm	1	C	G, H, I
		2	D	J, K, L
2	am	1	D	E, F, D
		2	C	B, C, A
	pm	1	B	K, L, J
		2	A	H, I, G
3	am	1	B	I, G, H
		2	A	L, J, K
	pm	1	D	C, A, B
		2	C	E, D, E
4	am	1	C	J, K, L
		2	D	G, H, I
	pm	1	A	D, E, F
		2	B	A, B, C

Figure 3. 8 week test data from DOE



## Advantages of 'Blocked' Test Plans

### Advantages of 'Blocked' Test Plans

- Any systematic biases are applied equally to all batches and so comparisons between batches are unaffected.
- More robust estimates of the batch mean and more realistic estimate of the variability of that batch (includes intermediate precision variability not just repeatability)

### Disadvantages

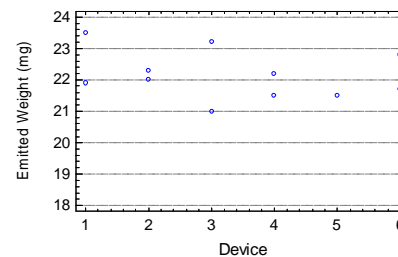
- Logistically more complex and so increased chance of mistakes being made

## HOW MANY DEVICES TO TEST?

Questions of sample size are more complex for inhaled products than with oral product pharmaceutical testing since we need to decide not only how many units to test but how should these units be sampled from different devices? Should we take multiple tests from a single device or should we take fewer tests from a larger number of devices?

If we are interested in estimating the mean of a batch say, the answer to these questions relies on having good estimates of the within and between device variability. This is usually done by having a reasonable amount of past data, or by making some assumptions around the components of variation.

Figure 4. Case A



Case A) Very small between device variability. Here we see that results from different devices are no more different than results from the same device. In the extreme case where no additional variability exists between devices, it is simply the total number of tests that is important. Consequently, the following test schemes give equivalent precision:

6 devices, duplicate tests

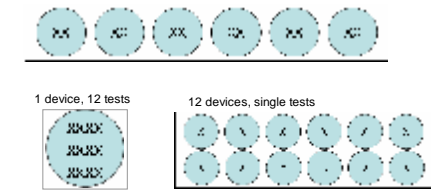
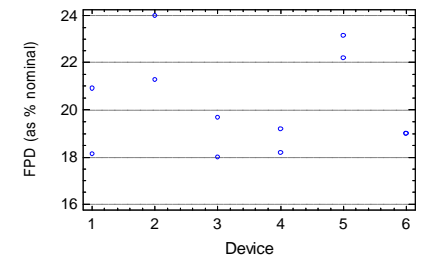
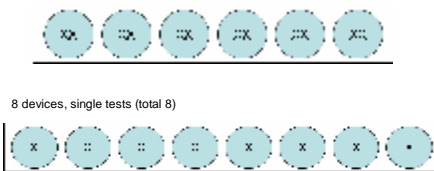


Figure 5. Case B



Case B) Both between and within-device variability present. In this situation, by increasing the number of devices tested, we can reduce the overall number of tests to achieve similar precision as shown in the following two test schemes:

6 devices, duplicate tests (total 12)



## CONCLUSIONS

By standardizing the NGI testing apparatus and employing the type of testing matrices described we have achieved more robust estimates of the batch mean and a more realistic estimate of the variability of that batch. By being very structured about how APSD testing is conducted we have also managed to increase productivity; typically a team of 6 people can complete 36 NGI tests in a normal working day. We have found the approach described valuable when supporting formulation development activities which use a statistical design of experiments (DoE) approach.