Adapting to Highly Randomized Experimental Designs in Formulation Development

Steven LaBrenz
Scientific Director,
Janssen R&D

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Contributors

Janssen R&D

Patrick Stahl

Joe Periaccha

August Allen (RIT)

Princeton University

Herschel Rabitz

Genyuan Li

Xi Xing

Rutgers University

Bill Welsh

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Outline

- Why randomize at all?
- Standard DoE development.
- Randomization, the easy way.
- Fully randomized design.
  - Preparation considerations.
  - Making it useable.

Why should I randomize?

- Instrumentation/Analysis bias
  - A common question, but difficult to quantify the impact.
  - Provides a level of confidence in the results.
- Part of a DoE
  - Built-in randomization of samples, allowing for preservation of sample data ordering.
- Neutralizing edge effects while maintaining sample throughput.
  - Enough replicates, and you may average out edge effects in colorimetric assays.

Automation assisting with screening

- We use automation to prepare a multivariate formulation screening design in a 96-well plate.
- The key is to use DoE structure that allows a scientist to evaluate results in a statistically driven manner.
- The regularized structure of the study leads to questions regarding sample prep and analysis bias.

		Excipient 1						Excipient 2					
		c1	c2	c3	c4	c5	c6	c1	c2	c3	c4	c5	c6
Buffer 1	pH1												
	pH2												
	pH3												
	pH4												
Buffer 2	pH1												
	pH2												
	pH3												
	pH4												

A simple DoE layout prepared using our workflow.

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Converting to Random

- Regular Order
  - Easy to prepare
  - Rapid, reliable
  - Data analysis is easier
- Random Order
  - Difficult to Prepare
  - Slow
  - Data analysis needs a key to de-code.

Be mindful of the automation preparation order (column- or row-wise) and the analysis order (column- or row-wise).

De-coding Example

Sequence
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
…

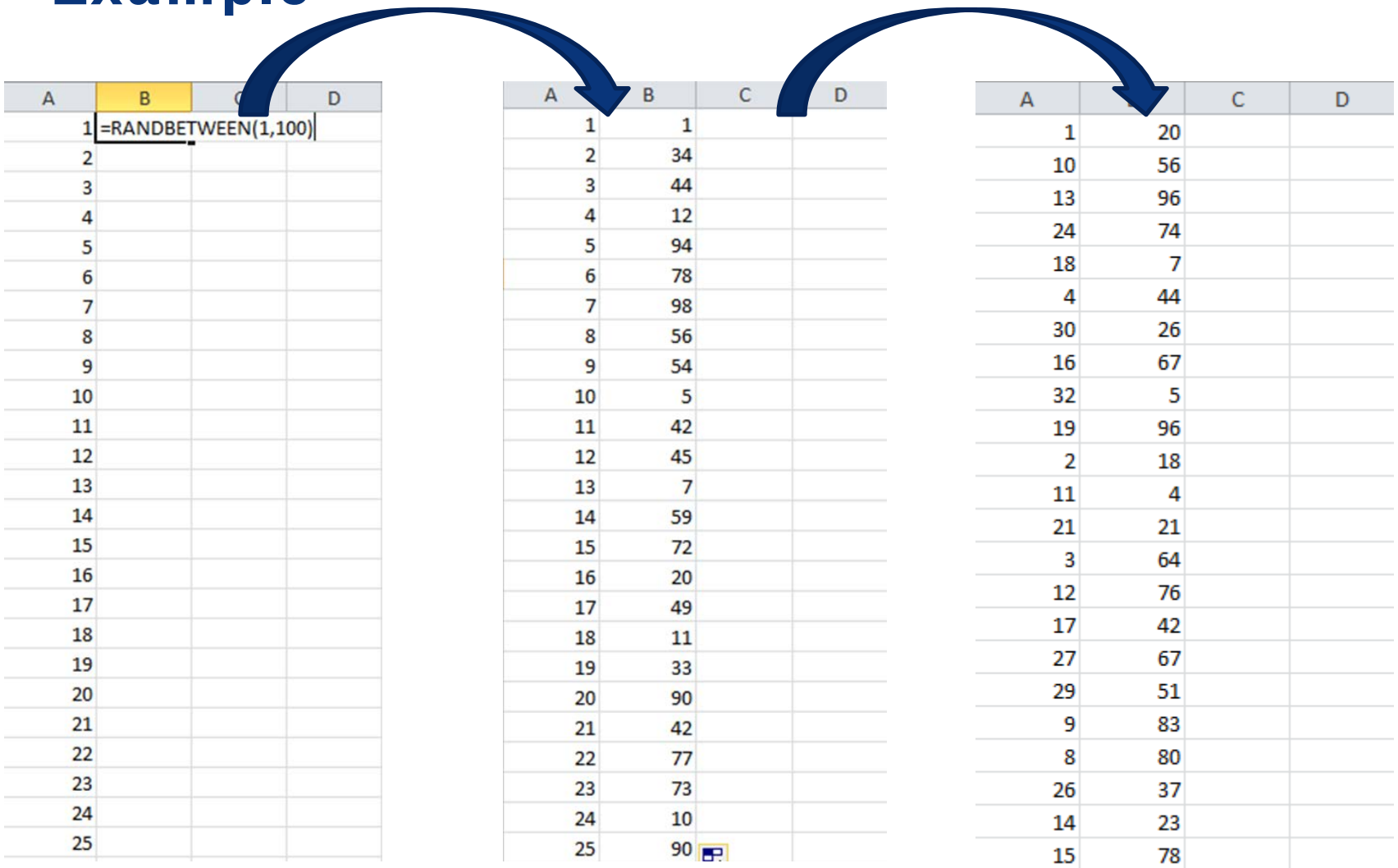
Plate (column)			
1	9	17	…
2	10	18	…
3	11	19	…
4	12	20	…
5	13	21	…
6	14	22	…
7	15	23	…
8	16	24	…

Transpose
1
9
17
25
33
41
49
57
65
73
81
89
2
10
18
26
34
42
50
58
66
74
82
90
…

Simple Randomization

- Start with the standard experiment
- Add in the randomization using pre-defined sequences stored in a *.txt file.
  - Start with an ascending sequence of integers, starting with 1 and running through the final sample number.
  - The random sequence is prepared in the same table using the RANDBETWEEN() function in Excel.
  - Once you have the integer sequence (Col A) and the random sequence (Col B), sort both based on the Random sequence.
    - This randomizes the *integer sequence*.

Example



A	B	C	D
1	=RANDBETWEEN(1,100)		
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			

A	B	C	D
1	1		
2	34		
3	44		
4	12		
5	94		
6	78		
7	98		
8	56		
9	54		
10	5		
11	42		
12	45		
13	7		
14	59		
15	72		
16	20		
17	49		
18	11		
19	33		
20	90		
21	42		
22	77		
23	73		
24	10		
25	90		

A	B	C	D
1	20		
10	56		
13	96		
24	74		
18	7		
4	44		
30	26		
16	67		
32	5		
19	96		
2	18		
11	4		
21	21		
3	64		
12	76		
17	42		
27	67		
29	51		
9	83		
8	80		
26	37		
14	23		
15	78		

Making it work

- Setup a loop that runs the whole sequence of samples.
- Within the loop:
  - Set the aspirate position to match the plate sequence.
  - Set the dispense position to match the randomized daughter plate sequence.

- User Choice – The user is given the choice of three different sequences for any single preparation.

Complex Randomization

- Randomize sample preparation from the start.
- No longer simple aspirate/dispense control.

- Utilize either *.xls or *.txt to control all aspirate functions.
  - Include skipped samples.
  - Exclude tip pickup for zero volumes.
- Follow regular sequence to dispense.
  - Segregate the samples based on based on tip volume.

File Structure

Samples	Stocks	Tips	Sample#	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
384	8	1000	1	795	132	0	0	73	0	0	0
		300	2	283	138	0	0	102	245	232	0
		300	3	726	0	41	0	104	0	129	0
		300	4	689	100	0	75	0	0	129	7
		300	5	704	0	17	0	13	43	177	46
		300	6	585	56	0	109	0	0	178	72
		300	7	825	0	70	0	20	35	34	16
		300	8	729	0	146	0	34	0	0	91
			9	522	0	166	0	79	233	0	0
			10	583	0	138	0	32	247	0	0
			11	713	11	0	0	19	73	184	0
			12	480	36	0	194	0	0	226	64
			13	648	128	0	97	0	30	97	0
			14	857	0	97	0	28	0	0	18
			15	757	0	71	0	69	0	103	0
			16	540	0	101	0	29	117	173	40
			17	709	27	0	0	13	162	0	89
			18	753	132	0	0	73	0	42	0
			19	536	0	52	0	103	0	237	72
			20	815	0	104	0	81	0	0	0
			21	696	52	0	63	0	189	0	0
			22	744	47	0	0	188	21	0	0
			23	586	0	115	0	70	139	0	90
			24	658	68	0	82	0	192	0	0
			25	588	0	121	0	119	0	172	0

The Code to run this experiment…

…took me a very long time to write and perfect.

Each cell in the columns that start with “Col” are the aspirate/dispense volumes.

Considerations

- You have a large quantity of *random* data!
  - "How am I going to analyze this?"
  - Need a strategy to de-code the data.
- How long do you have to complete the method?
- Are you storing the data in a dBase?
  - How do your collaborators need the data?