G-TWYST
GM PLANTS TWO YEAR SAFETY TESTING

Stakeholder Workshop on
2-Year Animal Feeding Studies,
16-17 Dec 2014

Statistical aspects of study design

Hilko van der Voet, Biometris, Wageningen UR

SEVENTH FRAMEWORK PROGRAMME

Study designs

- NK603
  90-day trial

Group	% of daily dietary intake (w/w)			No. of animals	
	Isogenic non-GM (NK603)	NK603	NK603 + Roundup	Males	Females
1	33	0	0	16	16
2	22	11	0	16	16
3	0	33	0	16	16
4	22	0	11	16	16
5	0	0	33	16	16
Total				80	80

- NK603
  1- and 2-yr trial

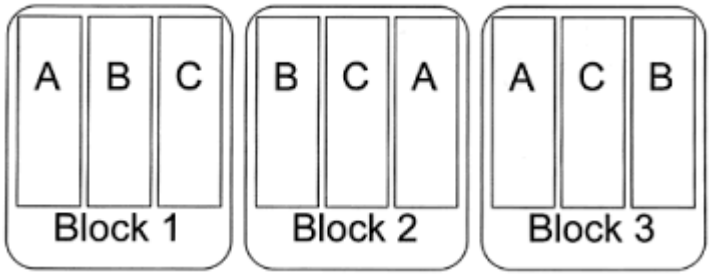
Group	% of daily dietary intake (w/w)			No. of animals			
				Chronic toxicity		Carcinogenicity	
	Isogenic non-GM (NK603)	NK603	NK603 + Roundup	Males	Females	Males	Females
1	33	0	0	20	20	50	50
2	22	11	0	20	20	50	50
3	0	33	0	20	20	50	50
4	22	0	11	20	20	50	50
5	0	0	33	20	20	50	50
Total				100	100	250	250

- MON810
  2-yr trial

Group	% of daily dietary intake (w/w)		No. of animals	
	Isogenic non-GM (MON810)	MON810	Males	Females
1	33	0	50	50
2	0	33	50	50
Total			100	100

Points for attention: Blocking, Randomisation, Blinding, Replication

Blocking



- G-TwYST uses randomised block designs
- Blocking is recommended by OECD-116 (2012) and EFSA (2011, 2013)
- Not all animals can be handled at the same time
- Not all analyses can be performed at the same time
- Principle:
  - combine one cage (2 animals) of each dose group into a block
  - Perform all work block by block (as far as practical)
    - Starting the experiment
    - Feeding and cleaning
    - Observations during the experiment: feed consumption, weighing, blood and urine sampling
    - Necropsy, weighing of organs
    - Analysis of samples
- Blocking minimises the variation between dose groups within blocks
- The statistical analysis can be based on a model that corrects for accidental differences between blocks

Randomisation
Example 90-day study

- Random assignment of:
  - Rats to cages
  - Dose groups to codes 1-5
  - Coded dose groups to cages in each block
- Randomisation is important to prevent confounding of treatment effects with other variation
  - For example, an effect on the cages in the leftmost position of each row is expected to be the same for all treatments

StartWeek	Sexe	BlockNr	Row	StartDay	Rack 1 Male				
Week 1	Male	Block 1	Row 1	Monday	3	2	4	5	1
Week 1	Male	Block 2	Row 2	Monday	1	5	3	4	2
Week 1	Male	Block 3	Row 3	Tuesday	1	3	2	5	4
Week 1	Male	Block 4	Row 4	Tuesday	5	3	2	1	4

StartWeek	Sexe	BlockNr	Row	StartDay	Rack 2 Male				
Week 1	Male	Block 5	Row 1	Wednesday	1	2	5	4	3
Week 1	Male	Block 6	Row 2	Wednesday	5	3	1	2	4
Week 1	Male	Block 7	Row 3	Thursday	4	3	1	5	2
Week 1	Male	Block 8	Row 4	Thursday	2	3	5	1	4

StartWeek	Sexe	BlockNr	Row	StartDay	Rack 3 Female				
Week 3	Female	Block 9	Row 1	Monday	2	1	3	4	5
Week 3	Female	Block 10	Row 2	Monday	3	5	4	2	1
Week 3	Female	Block 11	Row 3	Tuesday	1	3	4	2	5
Week 3	Female	Block 12	Row 4	Tuesday	5	4	1	3	2

StartWeek	Sexe	BlockNr	Row	StartDay	Rack 4 Female				
Week 3	Female	Block 13	Row 1	Wednesday	3	1	4	2	5
Week 3	Female	Block 14	Row 2	Wednesday	5	2	3	1	4
Week 3	Female	Block 15	Row 3	Thursday	3	1	4	2	5
Week 3	Female	Block 16	Row 4	Thursday	4	3	5	1	2

Blinding

Feed group	Code
Iso non-GM 33%	2
NK603 11% + Iso non-GM 22%	4
NK603 33%	1
NK603 11% + Iso non-GM 22% + Roundup	5
NK603 33% + Roundup	3

- Random assignment of codes to feed groups by feed supplier (Harlan)
- Numbers and colours used for coded groups
- The code will only be given to Ralf Wilhelm and Josefine Engel (JKI)
- Why? “Use of nonblinded outcome assessors of subjective outcomes may cause considerable observer bias in animal model experiments: nonblinded assessors exaggerated odds ratios by approximately 59% in 10 animal model experiments including 2,450 animals” (Bello et al. 2014)
- The dose groups will be unblinded after the analyses have been performed
  - Exception: Histopathology needs to know the identity of the codes because the within-group ‘normal variation’ has to be assessed. Therefore codes will be unblinded for the histopathologist after necropsy and the weighing of organs

Replication

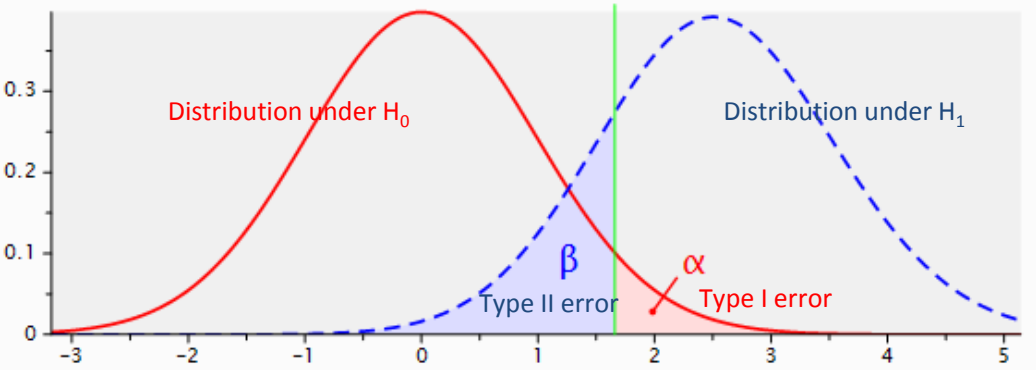
- Replication is the basis to estimate variation and therefore allow statistical testing
- More replication → higher precision

$$SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

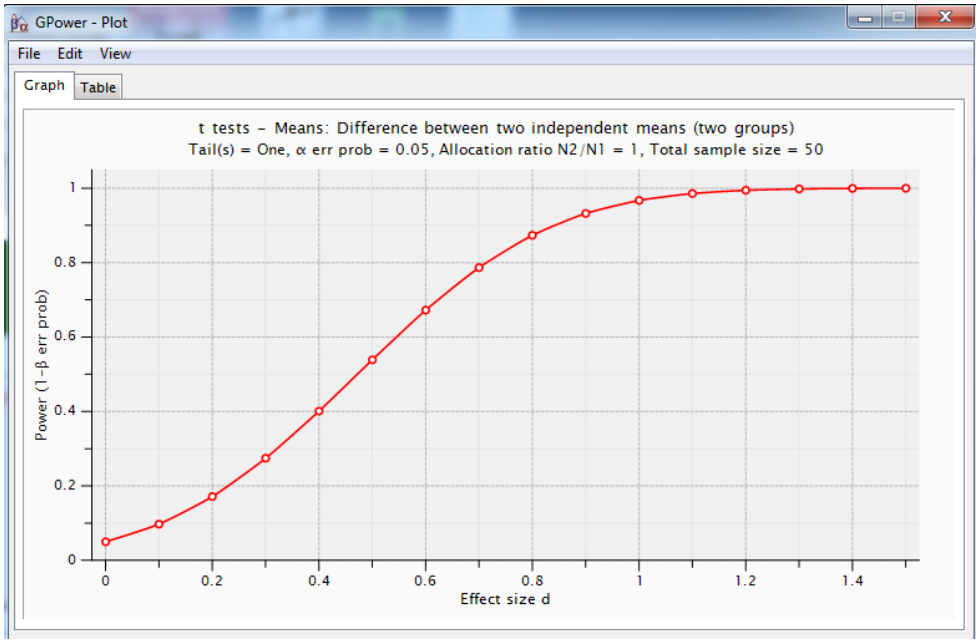
- "A power analysis to estimate a sample size capable of detecting a pre-specified biologically relevant effect size with a specified power and significance level should be used "(COMMISSION IMPLEMENTING REGULATION (EU) No 503/2013**)**
- "The use of a power analysis to estimate a sample size capable of detecting a pre-specified biologically relevant effect size with a specified power and significance level should be done to determine an appropriate sample size" (EFSA 2011)
- "The OECD Test Guidelines indicate the appropriate sample sizes for each group. In the carcinogenicity study, the sample size is usually at least 50 animals of each sex at each dose level. This group size reflects a trade-off between the statistical power of the design and economic practicalities of the design. In practice, the carcinogenicity study has low power in the sense that treatment effects that might be considered biologically important cannot be detected routinely as statistically significant." (OECD-116, 2012)

Power analysis

- In principle easy, at least for simple tests
- Four quantities:
  - Significance level α
  - Power = $1 - \beta$
  - Effect size
  - Sample size N
- Three options:
  - N = f(α, effect size, power)
  - Power = f(α, effect size, N)
  - Effect size = f (α, power, N)
- What is a relevant effect size?



Difference between GMO and Control



Equivalence limits

- Relevant effect sizes to be set as equivalence limits for the difference between GM and control groups
- EFSA (2011) and EC (2013) prescribe both difference testing and equivalence testing
- How to set equivalence limits?
  - Study-internal data
  - External data
  - Expert knowledge
  - Combinations thereof