

Veening-Griffioen et al.:

Tradition, Not Science, Is the Basis of Animal Model Selection in Translational and Applied Research

Supplementary Data

R scripts used for analysis

```
# Install packages
install.packages(tidyr)
install.packages(dplyr)
install.packages(reshape2)
install.packages(ggplot2)

# Load libraries
library(tidyr)
library(dplyr)
library(reshape2)
library(ggplot2)

# Set working directory
setwd("C:/.....")

# Load dataset (.csv) to be analyzed and assign variable to the dataset
AM_SPC <- readr::read_csv2("C:/.....csv")
AM_DIS <- readr::read_csv2("C:/.....csv")
AM_SPCDIS <- readr::read_csv2("C:/.....csv")

# Count frequencies – 2 variables; e.g. number of models per species
# Create a dataset with only the 2 variables of interest; e.g. modelID and species and assign a variable to the dataset
SPC <- select(AM_SPC, modelID, species)
View(SPC)

# Create a dataset with only the 2 variables of interest; e.g. modelID and disease area and assign a variable to the dataset
DIS <- select(AM_DIS, modelID, species)
View(DIS)

# Remove duplicates
SPC <- distinct(SPC)
DIS <- distinct(DIS)

# Occurance (frequency) of distinct species in the data
Count_SPC <- count(SPC, species)
View(Count_SPC)

# Occurance (frequency) of distinct disease areas in the data
Count_DIS <- count(DIS, dis_area)
View(Count_DIS)

# Save new dataset Count_SPC as ";" delimited csv-file (combined dataset)
write.csv2(Count_SPC, "C:/.....csv")

# Save new dataset Count_DIS as ";" delimited csv-file (combined dataset)
write.csv2(Count_DIS, "C:/.....csv")

# Count frequencies – 3 variables; e.g. number of models per species per disease area
# Create a dataset with only the 3 variables of interest; e.g. modelID, species and disease area and assign a variable to the dataset
SPCDIS <- select(AM_SPCDIS, modelID, species, dis_area)
View(SPCDIS)
```

doi:10.14573/altex.2003301s1

```

# Remove duplicates
SPCDIS <- distinct(SPCDIS)

# Occurance (frequency) of animal models in distinct species per disease area
Count_SPCDIS <- SPCDIS %>% count(species, dis_area) %>% spread(dis_area, n)
View(Count_SPCDIS)

# Save new dataset Count_SPCDIS as ";" delimited csv-file (combined dataset)
write.csv2(Count_SPCDIS,"C:/.....csv")

# Convert matrix Count_SPCDIS to a 3 column table, using species as id variables
bubble <- setNames(melt(Count_SPCDIS), c("species", "dis_area", "freq"))

# Omit lines with "NA"
bubble <- na.omit(bubble)
View(bubble)

# Save new dataset bubble as ";" delimited csv-file (combined dataset)
write.csv2(bubble,"C:/.....csv")

# Visualisation in "Bubble plot"
data <- bubble
View(data)

# Most basic bubble plot with controlled circle size and vertical labels for x-axis
ggplot(data, aes(x=dis_area, y=species, size = freq)) +
  geom_point(alpha=0.7) +
  scale_size(range = c(0.5, 8), name = "Animal Models (n)") +
  xlab("Disease Area") +
  ylab("Species") +
  theme(axis.text.x = element_text(angle=90, hjust=1))

# Saves the last plot
ggsave("C:/.....png")

```

Tab. S1: Distribution of animal models across disease areas and species

Disease area classification according to the classification for human disorders from the Central Authority for Food and Consumer Product Safety (NVWA) in the Netherlands.

Disease area	Animal models used in (number)									
	Total	zebra-fish	mouse	rat	guinea pig	rabbit	goat	sheep	pig	dog
Cancer	22		16	2					3	1
Cardiovascular disorders	20		5	8				1	6	
Endocrine or metabolism disorders	4		4							
Gastrointestinal disorders, incl. liver	4			3					1	
Immune disorders	20		13	6						1
Infectious disorders	4			4						
Musculoskeletal disorders	13		4	1		3		4	1	
Nervous or mental disorders	19	2	11	6						
Respiratory disorders	1		1							
Sensory organ disorders (skin, eyes, ears)	10		5		5					
Urogenital or reproductive disorders	5		2	1			2			
Other disorders	3			2					1	

Tab. S2: Correlation between animal model and species choice

Thematic content analysis of any information given by the applicant on the choice of a specific animal model (rows) and the choice of a specific species (columns), grouped per indicated keywords. Data is the fraction (%) of animal models with a correlated explanation.

Animal model choice (explanation)	Species choice (explanation)												
	availability of the animal model	availability of research tools	similar anatomy or organ system	similar disease pathology or symptoms	prior studies	sensitivity	cost effectiveness	simple accommodation and care	not reported	availability of expertise	close genetic/protein homology	availability of GMOs	reduction of discomfort
availability of the animal model	70	46	29	35	23	7	8	8	3	6	4	3	2
availability of expertise	51	40	26	22	28	9	9	8	10	6	4	3	0
similar disease pathology or symptoms	41	22	22	38	19	6	10	9	2	6	4	3	2
prior studies	10	9	2	6	12	0	0	0	6	2	0	0	2
preferred to other model	13	7	10	2	7	2	4	0	0	0	0	0	0
similar anatomy	12	2	13	0	0	1	2	0	6	0	0	0	0
similar organ system	10	2	10	7	2	1	0	0	0	0	0	0	2
close genetic homology	5	2	2	2	0	0	2	2	2	0	2	2	0
reported as fact	3	2	3	0	3	0	0	0	6	0	0	0	0
similar etiology	3	0	0	6	3	3	0	0	0	0	0	0	2
availability of research tools	3	4	2	2	2	1	2	0	0	2	0	0	0
infrastructure	3	0	3	3	0	0	0	2	2	0	2	2	0
experimental design	4	1	3	0	0	2	0	0	0	0	0	0	0
regulatory requirement EMA or FDA	0	0	2	2	2	0	0	0	0	0	0	0	0
availability of regulatory guidelines	2	0	0	2	0	0	0	0	0	0	0	0	0
similar biodistribution	2	0	0	2	0	0	0	0	0	0	0	0	0
lack of equivalent	0	1	0	0	0	0	0	0	0	0	0	0	0

Tab. S3: Correlation between animal model and outcome choice

Thematic content analysis of any information given by the applicant for the choice of a specific animal model (rows) and the choice of a specific species (columns), grouped per indicated keywords. Data is the fraction (%) of animal models with a correlated explanation.

Animal model choice (explanation)	Outcome choice (explanation)					
	disease pathology related	intervention related	not reported	disease specific guideline	(bio)marker identification	technical feasibility
availability of the animal model	67	66	2	1	1	1
availability of expertise	58	48	1	1	1	1
similar disease pathology or symptoms	46	46	3	1	1	1
prior studies	15	14	1	0	0	0
preferred to other model	12	10	0	0	0	0
similar anatomy	10	11	0	0	0	0
similar organ system	7	10	1	0	0	0
close genetic homology	5	3	0	0	0	0
reported as fact	7	6	0	0	0	0
similar etiology	4	6	1	0	0	0
availability of research tools	5	4	0	0	0	0
infrastructure	4	4	0	0	0	0
experimental design	2	2	0	0	0	0
regulatory requirement EMA or FDA	2	2	0	0	0	0
availability of regulatory guidelines	1	1	0	0	0	0
similar biodistribution	1	1	0	0	0	0
lack of equivalent	1	0	0	0	0	0