**Group size planning of breedings of gene-modified animals**

Philippe Bugnon1, Vlada Milchevskaya2,3, Achim Tresch2,3, Frank Brand4, Thorsten Buch1

1Institute of Laboratory Animal Science, University of Zurich, Zurich, Switzerland

*2* Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50931 Cologne, Germany

3Institute of Medical Statistics and Computational Biology, Faculty of Medicine, University of Cologne, Bachemer Str. 86, 50931, Cologne, Germany.

*4*Berlin School of Economics and Law, Department of Business and Economics,Quantitative Methods, Berlin, Germany

# Abstract

# Introduction

A wide array of different mouse strains and lines are used in biomedical research world-wide. The mouse became the most prominent vertebrate research species because it is a) a mammal and thus closely related to the human, b) easily kept with little demand on food and space and c) has a high reproductive rate with short generation time. While analysis of gene function in the mouse was at first restricted to natural mutants and those generated through untargeted approaches such es chemical or radiation-induced mutagenesis, in the last decades it has undergone an explosive development when more targeted methods of gene manipulation were developed. These started with insertion transgenesis obtained through zygote injection (REF), followed by targeted mutagenesis in embryonic stem cells and blastocyst injection (REF) and has pinnacled recently with the use of the CRISPR/Cas9 system in 1 and 2-cell stage embryos (REF). By now the analysis of gene-modified rodents constitutes a pillar of outstanding importance within the modern biomedical research community. Accordingly, Mouse Genome Informatics (MGI) counts 47´000 curated and 14´000 non-curated entries for mouse lines, excluding so far unpublished strains and those established in industry. The total number of rodents kept in Switzerland per year is around 800 000 – 900 000 (2019), in the EU it is 7 Mio. Today not only single mutants are bred and analyzed: rather combinations of multiple alleles of different genes have become mainstay in research. They facilitate precision research answering scientific question regarding cell and gene function with regard to organ, cell type, time and environment. These complex genotypes are obtained through time and cost intense breeding schemes. Breeding of animals for obtaining the required number and genotypes for a designed experiment is based on Mendel´s laws of inheritance, the laws of segregation and independent assortment. The application of Mendel´s laws makes the assumption that the loci of interest are found on separate chromosomes (also called linkage groups). The mouse normally carries 40 chromosomes, 20 from each parent, including the two sex chromosomes. Before planning any cross of two mouse lines it is thus advisable to first check for the chromosome carrying the genes. This can be easily done by use of gene resources such as MGI, NCBI´s gene database, or the UCSC and ENSEML genome browsers. In cases where loci of interest are located on the same chromosome, rules and mechanisms different from classical Mendelian genetics apply for successful breeding outcome (REF). In daily laboratory life a new breeding of genes located on different chromosomes is planned with help of the Punnett square (Fig. 1A), devised by Reginald C. Punnett at the beginning of the 20th century and based on Mendel´s laws The possible allele combinations found in all possible haploid germ cells, be it sperm cells or oocytes, are written at two orthogonal sides of the Punnett square, with the allele combination of the sperm cells at the top and of the oocytes at the side. The different sperm cells and oocytes defined by their alleles are combined in the center of the square to form the possible allele combinations of the next generations. Further, the Punnett square allows the determination of the frequency of appearance for each new diploid genotype of the next generation. Unfortunately for the biomedical researcher using gene-modified rodents, knowing the frequency of the expected genotypes does not correlate directly to the actual breeding outcomes. With the number of parallel breedings and thus born offspring kept at a minimum to fulfil “reduce” requirements, stochastic events have a large influence on breeding outcomes. Yet, this fact is usually not taken into account when breeding of gene-modified rodents is planned. This omission results in unnecessary breeding delays and scientifically unjustified animal use. In addition to Mendel´s laws further stochastic events are relevant for proficient planning of breeding outcomes. Fertility is to be taken into consideration as not all breeding pairs will produce offspring, in general or in the desired time period. Festing has made a first analysis of this breeding planning component and suggested a binomial distribution of occurrence of fertile females in a breeding cohort depending on known fertility. Also, litter size has an influence on the number of breeding pairs that need to be setup. Again, Festing made predictions for the incorporation of litter size into the planning of experimental breedings, under the assumption of Normal-Distribution of the data and a fixed standard deviation. He provided tables to incorporate both parameters into the proficient planning of breeding of animals. However, the tables regarding fertility and litter size that are now available for breeding planning are neither based on data-verified distribution functions nor are assumptions such as standard deviation verified through analysis of actual breeding outcomes. Further, the tables allow breeding planning only with rather big granularity as predictions are not made in single-animal steps, thus leading to sub-optimal results, with the possible consequence of inflated animal use.

Taken together, making predictions about breeding outcomes is astonishingly complex and usually ignored by researchers and breeding managers. Yet, this practice of insufficient planning of breedings of gene-modified animals frequently leads to researchers collecting animals for an experiment over time – generating a cohort of animals with grossly varying ages, a possible confounder in many experiments.

Here weaim at providing the breeder of gene-modified animals with easy to use tools allowing the inclusion of stochastic events in the planning process. We show how inclusion of the probability of occurrence of Mendelian genotypes in small breeding cohorts can be incorporated into everyday breeding planning. We incorporate this aspect of breeding planning with the stochastics of litter size and fertility, providing evidence that litter size is poisson distributed for a given mouse line. To support practicioners we provide an R package facilitating sound planning of breedings leading up to defined animal groups required for a particular experiment.

# Results

## The probability of obtaining a genotype in a certain litter.

When for a scientific experiment a specific number of animals of a particular genotype and age is needed, the required number of breeding pairs is set up, usually based upon the available individual genotypes and by taking into account Mendel´s laws of inheritance for making a prediction about the number of animals of the correct genotype in the next generation. However, while Mendel´s laws of segregation and independent assortment allow the calculation of the theoretical frequencies of genotypes in the next generation of a given breeding, this information alone is insufficient to predict the likelihood of appearance of these genotypes, especially when the smallest possible number of breedings and hence offspring are to be used for fulfilling the reduction requirement within the 3R concept. As shown in Fig. 1A for a simple cross of two animals heterozygous for a gene deficiency (knockout, KO), the Mendelian frequencies of genotype appearance in the next generation are 0.25 (WT/WT) : 0.5 (WT/KO) : 0.25 (KO/KO). When we assume a litter size of 6 pups, we can model such a breeding step by randomly picking 6 individuals from an infinite pool of individuals endowed with the three genotypes in the frequencies 1:2:1 (Fig. 1B, depicted in the colors blue, green, and pink, respectively). Hence, the appearances of 1, 2, 3 or more individuals of the KO/KO genotype is a matter of chance and governed by stochastic rules. For calculation of appearance of genotypes in the next generation with their probability of occurrence it has to be taken into account that when for the planned experiment only 1 KO/KO individual is required, a litter does not have to contain exactly 1 KO/KO individual; rather it can contain also 2, 3, 4 ,5 or even 6 KO/KO individuals to still yield the one animal required. Thus, the probabilities for obtaining 1, 2, 3, 4, 5 and 6 KO/KO individuals can be added up. In our example similar calculations apply to 2, 3, 4 or 5 animals that need to be obtained. Obviously, the probability of obtaining a single KO/KO individual is quite high, when we can take it from any litter containing at least one KO/KO individual. Hence, the probability of a litter containing a certain minimal number of KO/KO (or any other) genotype from a given breeding follows a cumulative binomial distribution (Fig. 1C and D). The probability of no animals of the correct genotype included in the litter has to be kept from the cumulation to yield useful information for the breeding scientist. Yet the probability of occurrence of a genotype in a litter does not help directly for planning of breeding outcomes. The probabilities for obtaining a given number of KO/KO individuals increase when we produce more pups by setting up more breedings, as shown for an example where 3 individuals of the correct genotype are needed (Fig. 1E and F). The red lines indicate the required animals to be born to obtain 3 pups of the respective genotype frequency with a likelihood of 90%.

Taken together, the probability of a genotype appearing in a single or several litter(s) can be obtained by combining Mendel´s rules and the cumulative binomial function.

## Predicting Breeding Outcome for One Group Cases

While understanding probabilities regarding a breeding outcome explains why many breedings programs do not yield the required animals of a particular genotype in the target time period even though Mendel´s law were used in the planning process, it would be even more useful to be able to do make group size predictions for appropriate setup of breedings of animals of known genotypes with the aim to obtain defined numbers of animals and genotypes. Hence, if for a respective experiment 5 KO/KO individuals are needed, it would be sensible to plan the breedings in such a way, that these 5 KO/KO animals are obtained with a given success probability (e.g. 90%). Based on the above-described binomial distribution of the probabilities of appearance of genotypes in the offspring, predictions about the number of born animals to meet the requirements of the experiment can be made. These predictions depend on the probability of a genotype appearing in the next generation according to Mendel and the required number of animals of a particular genotype. Fig. 2 shows the comparison of the required number of animals as calculated with the Mendel´s laws alone and including a 90% probability of breeding success. Obviously, the least necessary number of offspring should be generated to have the target number of animals with the correct genotype born; yet meeting the requirement of 90% success probability results in a larger offspring cohort than predicted by Mendel´s laws alone. In Table 1 the required number of offspring of 90% success rate of 4 different Mendelian probabilities are presented. When planning a set of breedings, it should be kept in mind, though, that litters come in quantiles (e.g. 6, depending on strain and genetic modification), requiring the researcher to decide at times for breedings yielding less than the optimal number of target animals at 90% success probability – thus leading to lower success probability - or breedings yielding more than the optimal number of target animals at 90% success probability – thus leading to higher success probability. Animal numbers required for success probabilities other than 90% can be easily calculated using R (for the script see Material and Methods). Taken together, similar to power and group size calculations in the planning process for an experiment, group size calculations can be made for breeding outcomes. This allows to meet experimental demand with a given success probability.

## Predicting Breeding Outcome for Multiple Group Cases

In many experimental designs multiple groups of animals necessary for an experiment are produced by the same breeding pairs. Again, an example is the use of WT/WT and KO/KO animals from breedings of heterozygous (WT/KO) parents. For the example experiment 3 WT/WT and 3 KO/KO animals are required. While this appears at first glance to be identical to the above-described case, the use of two genotypes changes predictions as outcome for one genotype depends on outcome of the respective other genotype. This becomes clear in the most extreme (yet possible) outcome cases. When all born pups are WT/WT, then obviously none will be KO/KO (and vice versa). The influence that use of two genotypes from the same set of breedings has on probability of success for obtaining the desired number of animals is shown in Fig. 3A. We plotted the number of born animals versus the success probability of obtaining 3 individuals of each genotype (A and B, both occurring at a 0.25 Mendelian frequency). Below 10 animals success is close to zero, beyond 30 animals no reasonable gains in terms of success can be achieved. The red line indicates the number of pups required to obtain 3 animals of both required genotypes with a 90% probability. Again, this information can be used to “power” the breedings: In Fig. 3B we show the required number of animals born for obtaining the target group size (identical group size for both genotypes) for 4 different Mendelian outcomes with a success probability of 90% in comparison to the direct requirements from a Mendelian calculation. The fraction of additional animals that are required is considerably higher than for the single genocotype case with 90% success probability (Fig. 2). The exact numbers of required animals for the 2-genotypes case are shown in table 2. This, first, assessment was performed with both groups of identical size and both required genotypes appearing with identical Mendelian frequencies. For groups of identical or different sizes another graphical thapproach has to be used: Fig. 3C shows the correlation between the experimentally required number of offspring of two genotypes appearing with the Mendelian frequency of 0.25 (x- and y-axis) and the number of animals necessary to be born for a success probability of 0.9 (the isolines). Here also the required number of offspring can be obtained for situations in which the group sizes are unequal. The same predictions can be generated when the required genotypes have different Mendelian frequencies, as shown in Fig. 3D for Mendelian frequencies of 0.5 and 0.25.

## Considering fertility and litter size for breeding size planning

Average litter size and fertility of a particular strain or line are additional necessary parameters for calculating the required number of breedings to fulfil the demand of a planned experiment. Fertility reflects the litter size parameter with the value zero for a particular strain and line. We analysed fertility and litter size for x mouse strains at the Laboratory Animal Service Center of the University of Zurich (Table X). The range of fertility was found to extend from X to y Using this data set we determined that litter size (excluding the value zero) can be accurately modelled by a Poisson distribution. Mean litter size (lambda in the poisson distribution) was observed to range from A to B, with a standard deviation (CAN ONE DO THIS?) ranging from O to P.Knowing the distribution allows the making of predictions about the number of breedings necessary to meet the requirements of a respective experiment for minimal number of offspring. Using C57BL/6 as an example we plotted the required number of breeding pairs for defined numbers of offspring in demand with a success probability of 95% in comparison to the use of average litter size without taking into account success probability.

## Optimizing breeding planning by use of Mendelian genetics, fertility and litter size

Finally we aimed at integrating all three breeding parameters (Mendelian probabilities, fertility and litter size) into one R package for ease of use.

Graph of total surplus to Mendel assumption?

Comparison with stepwise approach of FESTING, Graphical?

# Discussion

Application of the 3R has become standard, frequently also a legal requirement, in experimental sciences using animals. In this context also practices in general husbandry including breeding practices are to be scientifically optimized. Further, animals that are produced in husbandries but are not used for experiments are a new focal point of political attention. Many of these surplus animals are result of the very basic biology of mammalian genetics. We show here that for adequate planning of breeding outcomes even more surplus animals will have to be generated than when calculations are based on simple probabilities according to Mendel´s rules. This is due to the stochastic nature of breeding outcomes. Here we show the extent of these additional animals required for breeding success probabilities of 90%. We also show how the common practice of using different genotypes derived from the same breedings influences breeding success and thus requires even further additional animals. We provide tables that facilitate easy breeding planning for the practitioner. The R scripts can be easily adapted for other use cases, such as higher or lower success probabilities. Through simple modification of the respective frequency of occurance, they can also be applied for the planning of breedings where the outcomes are not following strict Mendelian frequencies. An example would be breedings yielding in some allele combinations embryonically lethal genotypes of partial penetrance (e.g. Li 2008 Dev. Cell, Kallapur 1999 Mol Reprod Dev ).



While our calculations take into account the stochastics of obtaining successful breeding outcomes in terms of required genotypes, the scientist who plans his breedings should still take into account the general breeding efficiency of a particular strain. This can be done through the production index (PI, Ayadi 2011, Festing and Peters 1999). A table of production indices of various strains can be found in White 2007.; it is recommended however to establish the index for each mouse line and facility independently. An example how to perform the required calculations can be found in table 3. We introduce an adjustement factor for use of the PI in a timed breeding. Assuming that in a continuous breeding setting females are distributed over the full 3 weeks of gestation, we made the assumption that in a timed mating all these pregnancies occurred together (hence the correction factor of 3). Another method to meet the requirement of obtaining a defined number of animals within a specific period of time was presented by Festing (Festing UFAW Handbook Laboratory Animals Breeding and Genetics 6th edition). It involves the littering percentages at days after start of breeding, the average effective litter size, and the percentage of effective fertility.

# Material and Methods

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **90% Breeding Success** | | | | |
|  | **Genotype probability according to Mendel** | | | |
|  | **0.5** | **0.25** | **0.125** | **0.0625** |
| **Number of animals required for experiment** | **Number of animals required to be born** | | | |
| **1** | 4 | 9 | 18 | 36 |
| **2** | 7 | 15 | 30 | 61 |
| **3** | 9 | 20 | 41 | 84 |
| **4** | 12 | 25 | 52 | 106 |
| **5** | 14 | 30 | 62 | 126 |
| **6** | 17 | 35 | 73 | 147 |
| **7** | 19 | 40 | 82 | 167 |
| **8** | 21 | 45 | 92 | 186 |
| **9** | 24 | 50 | 102 | 206 |
| **10** | 26 | 55 | 111 | 225 |
| **11** | 28 | 59 | 121 | 244 |
| **12** | 31 | 64 | 130 | 263 |
| **13** | 33 | 69 | 140 | 282 |
| **14** | 35 | 73 | 149 | 301 |
| **15** | 37 | 78 | 158 | 319 |

Table 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **90% Breeding Success** | | | | |
|  | **Genotype probabilities according to Mendel** | | | |
|  | **0.5** | **0.25** | **0.125** | **0.0625** |
| **Number of animals required for experiment** | **Number of animals required to be born** | | | |
| **1** | 5 | 11 | 23 | 47 |
| **2** | 8 | 17 | 36 | 74 |
| **3** | 11 | 23 | 49 | 99 |
| **4** | 13 | 29 | 60 | 122 |
| **5** | 16 | 34 | 71 | 144 |
| **6** | 18 | 40 | 82 | 165 |
| **7** | 21 | 45 | 92 | 186 |
| **8** | 23 | 50 | 102 | 207 |
| **9** | 26 | 55 | 112 | 228 |
| **10** | 28 | 60 | 122 | 248 |
| **11** | 30 | 65 | 132 | 268 |
| **12** | 33 | 69 | 142 | 287 |
| **13** | 35 | 74 | 152 | 307 |
| **14** | 37 | 79 | 162 | 327 |
| **15** | 40 | 84 | 171 | 346 |

~~Table 3~~

|  |  |  |
| --- | --- | --- |
|  |  | ~~Example~~ |
| ~~Targeted group size~~ | ~~A~~ | ~~10~~ |
| ~~Probability according to Mendel~~ | ~~B~~ | ~~0.125~~ |
| ~~Animals required to be born according to Mendel~~ | ~~C = A / B~~ | ~~80~~ |
| ~~Animals required to be born according to table 1~~ | ~~D~~ | ~~111~~ |
| ~~PI (production index)~~ | ~~E~~ | ~~0.5~~ |
| ~~Adjustment factor for timed breeding (versus continuous)~~ | ~~F~~ | ~~3~~ |
| ~~Adjusted PI~~ | ~~G = F\*E~~ | ~~1.5~~ |
| ~~Age range for experiment~~ | ~~H~~ | ~~2 weeks~~ |
| ~~Number of females required for obtaining experimental animals~~ | ~~(D/G)/H~~ | ~~37~~ |

|  |  |  |
| --- | --- | --- |
|  |  | ~~Example~~ |
| ~~Targeted group size~~ | ~~A~~ | ~~10~~ |
| ~~Probability according to Mendel~~ | ~~B~~ | ~~0.125~~ |
| ~~Animals required to be born according to Mendel~~ | ~~C = A / B~~ | ~~80~~ |
| ~~Animals required to be born according to table 1~~ | ~~D~~ | ~~111~~ |
| ~~average effective litter size~~ | ~~F~~ | ~~7~~ |
| ~~percentage of effective fertility~~ | ~~G~~ | ~~84% [Green and Witham, 1991](http://www.informatics.jax.org/silver/references.shtml" \l "Green6" \t "_blank)~~ |
| ~~No of litters needed~~ | ~~According to Table 3.12 Festing~~ | ~~21~~ |
| ~~Number of females required for obtaining experimental animals~~ |  | ~~ca. 30 females~~ |

# Supplementary Material