

When power calculations won't do: Fermi approximation of animal numbers

All animal protocols need quantitative justification of animal numbers, but not all will need power calculations.

Penny S. Reynolds

The number of animals requested in an animal use protocol should be “sufficient but not excessive” for the purposes of a study¹, and be governed by what is justifiable, appropriate, reasonable, and ethical. Statistical power analysis is considered to be the gold standard for determining animal numbers. For example, the *Guide for the Care and Use of Laboratory Animals* states that “whenever possible, the number of animals and experimental group sizes should be statistically justified (e.g., provision of a power analysis)”². Unfortunately, it is not unusual for investigators to request approval for numbers that defy common sense, even when ‘justified’ by a power calculation. Investigators might defend requests for tens or even hundreds of thousands of animals for a 3-year proposal because “unforeseen problems might happen” or “it is unknown how many animals we will require because this is an exploratory study,” for example. Conversely, many investigators simply choose a sample size that “worked” in previous studies “to obtain significance”, and use a power calculation as an *ad hoc* justification³.

The goal for any method used to estimate animal numbers is to obtain reasonable approximations of the number of animals necessary to meet study objectives. However, power analyses cannot, and should not, be the only method used. First, power calculations are only appropriate for studies for which the primary focus is null hypothesis significance testing (NHST) and statistical estimation (concerned with sampling distributions, and point and interval estimation). For those studies, the goal is to make inferences, compare groups, or assess the probability that study results are true. By contrast, exploratory studies, demand–supply protocols, feasibility assessments, and pilot studies are not testing hypotheses; therefore numbers based on NHST and *P*-values are meaningless in these contexts. Second, all projected animal use numbers (for NHST or not) require cross-validation to ensure that they are operationally feasible. Supporting evidence

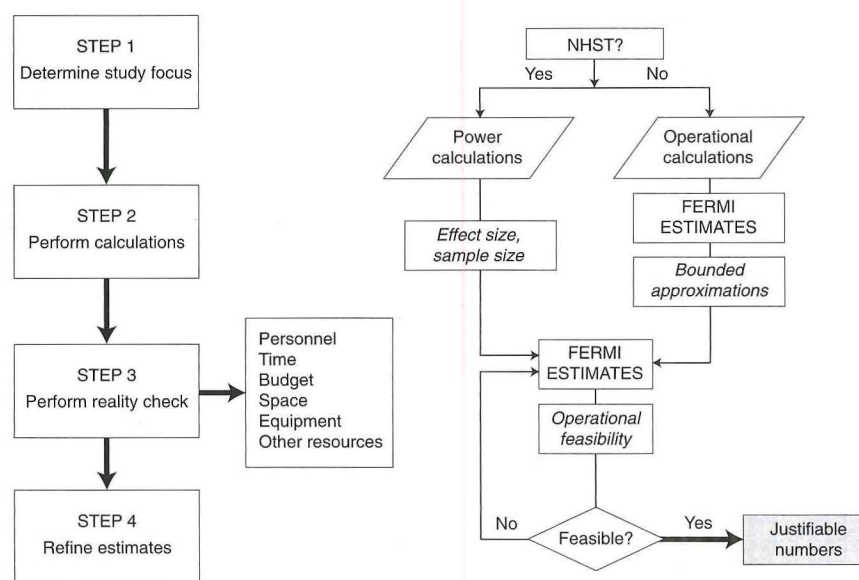


Fig. 1 | Animal numbers decision tree. NHST; null hypothesis significance testing.

should be provided to confirm that the proposed work practices, procedures, and personnel are sufficient to support the project, and that the project can be completed in the allotted time and within budget. It is clearly unethical to propose or approve studies that require numbers of animals that far exceed processing capacity, and therefore do not have a realistic chance of completion. In all cases, reasonable approximations of numbers are required to solve practical problems determined by study objectives.

Quantitative verifiable approximations of animal numbers for a given project can be obtained by structured back-of-the-envelope calculations⁴. The process, known as Fermi estimation⁵, consists of a logical formulation of the approximation problem, systematically broken down into smaller subproblems or components, each of which is solved by basic arithmetic. Demonstration of calculations enables reviewers to understand and verify animal numbers. The method was named after the physicist Enrico Fermi, who had an extraordinary

aptitude for finding quick and accurate answers to practical complex problems using simple calculations^{5,6}.

The Fermi estimation method

Fermi estimation is a four-step cyclical process consisting of the identification of the study focus, structured calculations, a reality check, and a refinement step (Fig. 1). Study focus provides the initial framework for subsequent calculations. If the study requires NHST, the objective is to obtain a sample size likely to detect a biologically or clinically meaningful effect size if it exists, and avoid false positives⁷. In this situation, the investigator specifies the hypothesis and all relevant statistical components (such as primary outcome measures, clinically relevant differences to be detected, measures of variability, settings for α and β), and then performs the power calculations appropriate to the statistical hypothesis and type of analyses required⁸. On the other hand, if the problem type is operational—that is, numbers are to be determined by the demand and/or supply needs of the

Table 1 | Estimating numbers of animals required (N) for tissue harvest studies

Subproblems	Items	Calculations
Model		
<ul style="list-style-type: none"> • Quantity available per animal • Quantity required per assay 	<ul style="list-style-type: none"> • Tissue mass or cell number per tissue type • Tissue mass or cell number per well • Number of plating wells • Number of treatments • Number of replicates • Number of sampling time points 	<ul style="list-style-type: none"> • Total N = (total amount of tissue required per replicate) x (number of animals per unit tissue yield) x (number of replicates) x (number of time points)
Example 1: RNA or protein analysis.		
<ul style="list-style-type: none"> • Quantity available per animal • Quantity required per assay 	<ul style="list-style-type: none"> • 1–2.5 x 10⁵ cells can be isolated per mouse • 1 x 10⁶ cells per well are needed to obtain sufficient RNA or protein for analysis. 	<ul style="list-style-type: none"> • Number of mice (upper bound) = (1 x 10⁶ cells per well) / (1 x 10⁵ cells per mouse) = 10 mice per well • Number of mice (lower bound) = (1 x 10⁶ cells per well) / (2.5 x 10⁵ cells per mouse) = 4 mice per well • Total N = 4–10 mice per well
Example 2: ex vivo model of organ injury		
<ul style="list-style-type: none"> • Quantity available per animal • Quantity required per assay 	<ul style="list-style-type: none"> • One organ per animal • One organ per well • Experiment includes two treatments (intact organ control versus injured organ) • Experiment to be run in duplicate (two mice per treatment) • Four time points • Five mouse strains to be tested 	<ul style="list-style-type: none"> • Number of mice per strain = (2 mice per treatment) x (2 treatments per time point) x (4 time points) = 16 mice • Total N = (16 mice per strain) x 5 strains = 80 mice

study—NHST and power calculations are not appropriate. Unlike power calculations, which depend on mathematical relationships between statistical components, operational approximations are derived from pragmatic information specific to the problem to be solved. Simple arithmetic is then used to compute numbers for each component, and the total is obtained by adding all component parts together. Because precise values for each component are usually unknown, or different background information sources might present very different values, approximations should be bracketed by setting biologically reasonable upper and lower bounds.

Reality checks, or feasibility assessments, using Fermi calculations should be performed for both NHST and operational studies. These checks consist of comparing the proposed numbers with approximations based on laboratory-specific processing capacity and operational constraints, such as number of personnel and personnel-hours available, time required to perform important procedures, resource availability, estimated costs, total available budget, deadlines and milestones. If the proposed number of animals is too large to be feasible, the approximation problem will need to be reformulated or refined to meet operational capability.

Examples of Fermi estimation

Tissue-based studies. Power-based sample size calculation algorithms are available for

high dimensionality exploratory studies (such as multiple DNA/RNA microarrays, biochemistry assays, biomarker studies, proteomics, inflammasome profiles). However, sample sizes based on these algorithms are determined, not by considering the animal as the replicate, but by array-specific characteristics, such as the number of arrays, investigator-specified sensitivity, or proportion of differentially expressed genes detected, number of expected false positives, correlation between expression levels of different genes, and number of sampling time points⁹. Therefore, a separate set of calculations is needed to approximate the expected number of animals required for tissue harvest. The number required will depend on the amount of tissue, cells, or antibodies that can be harvested from a single animal, and the total amount of tissue required for the microarrays (Table 1).

Wildlife studies. Wildlife studies that include capture, handling, marking, non-invasive and invasive procedures, or lethal take, or that do not meet the regulatory definition of a field study, are subject to IACUC review and as such, a rationale for animal numbers is required. Projected take numbers must be subject to oversight because the number of captured animals cannot exceed the number on issued permits, and trap-removal procedures must be evaluated for their potential effect on wild populations. Although many field

studies involve NHST, it has been argued that NHST in wildlife science applications is often misapplied and is seldom useful^{10,11}. Regardless, NHST rationale might not be appropriate for numbers proposed for permit applications, species surveys, and inventories. In these cases, projected animal numbers are determined by ecological and investigator-sourced information. This information includes prior census information, expected species density, proposed census and/or capture methods, census coverage and intensity, and anticipated capture ratios. Potential adverse events, estimated trap attrition or mortality, methods to reduce loss, and humane endpoints should also be documented. This information can also be used to assess the feasibility of hypothesis-testing studies (Table 2).

Mouse breeding colony. Breeding protocols are required when requesting production of research animals that cannot be obtained commercially. Examples include the creation of new transgenic, knockout, or other genetically modified animals; back-crossing of genetically modified lines; or production of prenatal or early neonate subjects¹². For some applications such as genome-wide association studies, total numbers might be approximated by simple rules of thumb¹². However, protocols requesting production of a target genotype will require an estimate of the total numbers of adults and pups (or embryos if necessary). Because it is

Table 2 | Estimating numbers of animals captured (N) for wildlife species inventories or surveys

Subproblem	Items	Calculations
Model		
<ul style="list-style-type: none"> Anticipated trap (census) events Trap (census) effort 	<ul style="list-style-type: none"> Number of trapping or survey locations Trap density Number of survey periods Trapping duration per survey period Expected number of species per location Estimated number of specimens per species per locality Trap 'fails' = percentage of traps sprung without capture 	<ul style="list-style-type: none"> Trap effort = (number of traps per site x number of sites) x (number of days per year x number of years) Trap events = (number of captures per trap site per species) x (number of sites) Effort ratio = expected trap events/trap effort Total N = trap effort x effort ratio
Example: collection of five different species of rodents over ten localities for 3 years		
<ul style="list-style-type: none"> Anticipated trap (census) events Trap (census) effort 	<ul style="list-style-type: none"> At each locality, a trap grid consisting of 100 standard Sherman live traps arranged in a 10x10 matrix, will be established Traps will be checked twice daily for ten consecutive days in April-May and November-December. 	<ul style="list-style-type: none"> Trap effort = (100 traps per site x 10 sites) x (2 trap periods per year x 10 days per trap period) x 3 years = 60,000 trap-days Data for similar habitats suggests that the expected number of trap events is ~30,000 animals with a trap effort of 302,400-405,200 trap-days. Effort ratio is [30,000/405,200 to 30,000/302,400], or [0.074-0.099] Total N = (60,000 x 0.074-60,000 x 0.099) = 4200-6000 animals in 3 years Based on previous data, the estimated proportion of captures by species is: 63% for species A, 20% for species B, 5% for species C, 10% for species D, and < 2% for species E. Therefore, the anticipated total N by species is: 882-1260 for A, 280-400 for B, 70-100 for C, 140-200 for D, and 30-40 for E

difficult to predict pup production¹², approximations can be based on data from projections from past demand, past protocols, breeding colony records, current breeding stock numbers, and facility and personnel capability. Other information to be reported includes methods for genotype confirmation or other quality assurance methods, as well as means to dispose of unused animals, together with contingency plans to reduce the number of animals

with the unwanted genotype that will be euthanized (Table 3).

Feasibility assessments. Feasibility assessments are performed to confirm that the study is possible given the available resources¹³. Calculations should be used to support projected numbers of animals ("Does the estimated number of animals make sense?") in the context of operations ("Are current work practices, procedures,

resources, and trained personnel sufficient to support the project with the estimated numbers of animals?" and time-related or budget-related criteria ("Can the project be completed with the estimated numbers of animals in the allotted time and with the allotted amount of funding?"; Table 4)^{14,15}.

Pilot studies: a special case. Preclinical pilot or feasibility studies are essentially mini-rehearsals of the proposed definitive

Table 3 | Estimating pup production or target genotype numbers in breeding colony protocols

Subproblem	Items	Calculations
Model		
<ul style="list-style-type: none"> Target pup production Number of breeding adults 	<ul style="list-style-type: none"> Number of breeder females per strain Sex ratio (number of females per breeding group) Litter size (number of pups per female) Estimated total number of pups produced Proportion of pup attrition Expected genotype distribution Proportion of desired genotype required Time frame (pups produced per unit time) Number of strains 	<ul style="list-style-type: none"> Number of breeder females = (number of breeding groups) x (sex ratio) Number of pups = (number of females) x (number of pups per female per time unit) x (time unit) x (1 - pup attrition) Proportion of pups used = (proportion genotype) x (number of pups) Total N = (number of breeders) + (total number of pups)
Example: Estimating pup production for a colony established to produce heterozygote (HET +/-) and homozygous knockout (KO -/-) mice		
<ul style="list-style-type: none"> Target pup production Number of breeding adults 	<ul style="list-style-type: none"> The investigators have 10 breeding trios (1 M, 2 F) = 30 HET mice From past breeding records, average litter size is ~5 pups Anticipated perinatal losses of KO are approximately 15-20% (3-5 offspring out of 25) 	<ul style="list-style-type: none"> Estimated number of pups over 3 months = 10 trios x 2 litters per trio x 5 pups per litter = 100 pups ~50% of pups will be HET, 25% will be KO, and 25% will be homozygous wild type (WT; +/-). Therefore 25 KO pups are expected over 3 months. Number of viable KO young per quarter = [25 - 5; 25 - 3] = [20-22]

Table 4 | Feasibility calculations (proposed animal numbers versus processing capacity)

Sub-problems	Items	Calculations
Model		
<ul style="list-style-type: none"> Processing capacity Costs 	<ul style="list-style-type: none"> Number of trained personnel Total study duration Number of working hours Duration of procedure Cost per procedure Study budget 	<ul style="list-style-type: none"> Processing capacity $N = (\text{number of animals processed by one person in 1 h}) \times (\text{number of working hours per work day}) \times (\text{number of work days per year}) \times (\text{number of trained personnel})$ Processing rate (N per unit time) = (total number requested) / ((number of trained persons) \times (number of working hours per day) \times (number of working days per year)) Budget assessment = Total budget - ((cost per animal) \times (number of animals requested))
Example 1: An investigator requested 500,000 mice for a 3-year project. Is this project feasible?		
<ul style="list-style-type: none"> Processing capacity Costs 	<ul style="list-style-type: none"> Two trained personnel were identified in the protocol Assumed working days and hours: 350 days per year; 8-10 hours per day Procedure time for each animal: 30-60 min 	<ul style="list-style-type: none"> Number of animals processed per person (lower bound): = 350 working days per year \times 3 years \times 8 h \times 1 processed animal per hour (procedure time 60 min) = 8,400 Number of animals processed per person (upper bound) = 350 days per year \times 3 years \times 10 h \times 2 processed animal per hour (procedure time 30 min) = 21,000 Processing capacity (two people) = 16,800-42,000 animals Processing rate (N/min) = 500,000 animals / (2 persons \times 10 h \times 60 min/h \times 350 days per year \times 3 years) = 0.39 = 1 animal every 3 min per person. Conclusions: projected numbers were unrealistic, and the protocol could not be approved without modification.
Example 2: An investigator requested 120 swines for a 3-year project. Is this study feasible?		
<ul style="list-style-type: none"> Processing capacity Costs 	<ul style="list-style-type: none"> Four trained personnel A full day was required for four persons to perform a single experiment on one animal An estimated additional 8-16 h per week were required for experimental setup and takedown, inventory, data management, and related tasks The investigator considered two experiments per week was a reasonable performance goal The investigator anticipated that \geq 4-6 months would be required for protocol oversight, paperwork to be filed and approved, and to obtain necessary supplies and equipment. 	<ul style="list-style-type: none"> Total project time = 3 years \times 12 months per year = 36 months Available time = 36 - 6 = 30 months Processing rate = 2 animals/ week Time required = 120 animals / (2 animals per week \times 4 weeks per month) = 15 months Conclusions: the investigator has allowed adequate time to complete the project without compromising the well-being of staff or study quality, and to provide a time cushion in case of unforeseen problems or breaks in the workflow.

experiment. They are conducted to assess practical logistics and inform the design and conduct of the subsequent large trial¹⁶⁻¹⁹. A well-designed pilot can thus improve translation and reduce animal use by allowing early identification of problems that contribute to study futility and subsequent waste of animals¹². However, criteria for determining and justifying the best size for a pilot study are problematic because they might not be directly quantifiable. Because pilot trials do not test hypotheses about the effects of an intervention, power-based sample size estimations are generally not applicable (although sample sizes for human clinical pilots can be justified statistically under certain conditions)¹⁸⁻²⁰. Preclinical pilot studies are usually deliberately kept small (\leq 10 subjects), with numbers derived from personal subjective criteria, such as the investigator's experience and guesswork²¹.

Unfortunately, pilot studies can be poorly thought-out and without discernible purpose¹⁸. Without clear-cut objectives, studies cannot fulfill their intended purpose of improving the development and design of later experiments, and data will be insufficient to either obtain power-based sample size calculations for later trials or provide preliminary assessments of efficacy and safety of the treatment intervention^{14,22}.

Pilot studies should have one or more specific objectives with quantifiable success metrics (that is, how will the investigator know the pilot trial works?). Pilot objectives can be utilitarian ("Can it work?"), empirical ("Does it work?"), or translational ("Will it work?")¹⁵. Utilitarian pilots are designed to assess study logistics and enable standardization, correction, and refinement of experimental procedures. Simple assessment metrics can include

percent success of subject recruitment and enrollment procedures¹⁹ (for example for veterinary clinical trials), procedure times, target performance measures, and human performance errors²³. Empirical pilots determine if study endpoints, especially novel ones, are related to study objectives, can be measured or are sufficiently expressed to be useful, and have a clear relationship to the independent or treatment variable; the study also determines if data can be collected correctly and completely, and if results are accurate and reliable¹⁹. Other useful information might be obtained, such as response rates of controls and occurrence of adverse events. Translational pilots are designed to assess if the experimental intervention can be made more broadly applicable to other animal strains, species, or operating conditions. In these cases, standardization of, and

strict adherence to, protocols will enable comparisons (“Does the process work the same way in B as it did in A?”). Depending on pilot objectives, sample sizes can be approximated with information from the literature and from previously described studies, or by using simple rules of thumb if estimates of study endpoints are required²⁴.

Conclusions

Fermi estimation is a formalized method of approximating animal numbers, using a few simple calculations and basic arithmetic, and supported by operationally sensible assumptions about the items and resources involved. Such approximations are appropriate for exploratory and demand-based studies, and can be used as a feasibility assessment for any proposed study, whether or not hypothesis testing is required. At a minimum, adequate justification of animal numbers should be based on four main reporting requirements: study focus, a description of one or more assessment variables, a clear description of any calculations performed (“show your work”), and a feasibility assessment. For complex studies and/or multiple experiments, a flow chart showing group sizes, approximate numbers per group, timeframes, key interventions, primary outcomes, experimental and/or humane endpoints, and other information can be helpful for explaining how the total number of animals was determined and the animals were disposed. Investigators should be expected to practically implement the 3Rs, especially for protocols involving unrelieved pain and distress. Submission of protocols

with unfeasible numbers of animals should always be discouraged.

Justification of animal numbers is related explicitly to considerations of animal use and welfare, and is thus an ethical obligation, as well as a requirement of federal funding agencies²⁵. However, due diligence must be exercised by investigators, oversight committees, and reviewers to confirm that numbers proposed are scientifically and operationally valid. Approximation and justification of animal numbers should be based on more objective and verifiable criteria than unsubstantiated assertions based on experience and guesswork. By providing an evidence and fact-based structure for subsequent calculations, Fermi estimation methods can enhance rigor and reproducibility of the proposed research. □

Penny S. Reynolds

Department of Anesthesiology, University of Florida,
Gainesville, FL, USA.

e-mail: PREynolds@anest.ufl.edu

Published online: 21 August 2019
<https://doi.org/10.1038/s41684-019-0370-2>

References

1. *The IACUC Handbook* 3rd edn (ed. Silverman, J., Suckow, M. & Murthy, S.) (CRC Press, Boca Raton, 2014).
2. Institute for Laboratory Animal Research. *Guide for the Care and Use of Laboratory Animals* 8th edn (National Academies Press, Washington, DC, 2011).
3. Fitzpatrick, B. G., Koustova, E. & Wang, Y. *Lab Animal (NY)* **47**, 175–177 (2018).
4. Weinstein, L. & Adam, J.A. *Guessimation: Solving the World's Problems on the Back of a Cocktail Napkin* (Princeton University Press, Princeton, NJ, 2008).
5. von Baeyer, H.C. The Fermi solution in *The Fermi Solution: Essays on Science* pp 3–12 (Dover Publications, Inc., New York, 2001).
6. Fermi, E. *My observations during the explosion at Trinity on July 16, 1945 - E. Fermi*. Record Group 227, OSRD-S1 Committee, Box 82 Folder 6, “Trinity” U.S. National Archives (1945), http://nuclearfiles.org/menu/key-issues/nuclear-weapons/history/pre-cold-war/manhattan-project/trinity/eyewitness-enrico-fermi_1945-07-16.htm
7. Kimmelman, J., Mogil, J. S. & Dirnagl, U. *PLoS Biol.* **12**, e1001863 (2014).
8. Gosselin, R.-D. *Lab. Anim.* **53**, 28–42 (2018).
9. Jung, S. H. & Young, S. S. *J. Biopharm. Stat.* **22**, 30–42 (2012).
10. Johnson, D. H. *J. Wildlife Management* **66**, 272–276 (2002).
11. Anderson, D. R., Burnham, K. P. & Thompson, W. L. *J. Wildlife Management* **64**, 912–923 (2000).
12. Pitts, M. (ed.) *Institutional Animal Care and Use Committee Guidebook*, 210 pp. (Office of Laboratory Animal Welfare, National Institutes of Health, Bethesda, MD, 2002).
13. Eldridge, S. M. et al. *PLoS One* **11**, e0150205 (2016).
14. Arain, M., Campbell, M. J., Cooper, C. L. & Lancaster, G. A. *BMC Med. Res. Methodol.* **10**, 67 (2010).
15. Bowen, D. J. et al. *Am. J. Prev. Med.* **36**, 452–457 (2009).
16. Festing, M. F. W. & Altman, D. G. *ILAR J.* **43**, 244–258 (2002).
17. Kistin, C. & Silverstein, M. *314*, 1561–1562 (2015).
18. Moore, C. G., Carter, R. E., Nietert, P. J. & Stewart, P. W. *Clin. Transl. Sci.* **4**, 332–337 (2011).
19. Lancaster, G. A., Dodd, S. & Williamson, P. R. *J. Eval. Clin. Pract.* **10**, 307–312 (2004).
20. Whitehead, A. L., Julious, S. A., Cooper, C. L. & Campbell, M. J. *Stat. Methods Med. Res.* **25**, 1057–1063 (2016).
21. Dell, R. B., Holleran, S. & Ramakrishnan, R. *ILAR J.* **43**, 207–213 (2002).
22. Altman, D.G., et al. Why do a pilot study? *NC3Rs: National Centre for the Replacement, Refinement and Reduction of Animals in Research*, 2 pp. (2006).
23. Reynolds, P.S., et al. *Lab. Anim.* <https://doi.org/10.1177/0023677218802999> (2018).
24. van Belle, G. *Statistical Rules of Thumb* (John Wiley & Sons Inc, Hoboken NJ, 2008).
25. Landis, S. C. et al. *Nature* **490**, 187–191 (2012).

Acknowledgements

The author thanks C. Garvan and T. Vasilopoulos (Statistics in Anesthesiology Research (STAR) Core, University of Florida, USA) for their helpful comments on an earlier version of this manuscript.

Competing interests

The author declares no competing interests. The author discloses membership with the University of Florida IACUC committee.